Alibaba Cloud
Log Service
Data Collection

Issue: 20200323
Legal disclaimer

Alibaba Cloud reminds you to carefully read and fully understand the terms and conditions of this legal disclaimer before you read or use this document. If you have read or used this document, it shall be deemed as your total acceptance of this legal disclaimer.

1. You shall download and obtain this document from the Alibaba Cloud website or other Alibaba Cloud-authorized channels, and use this document for your own legal business activities only. The content of this document is considered confidential information of Alibaba Cloud. You shall strictly abide by the confidentiality obligations. No part of this document shall be disclosed or provided to any third party for use without the prior written consent of Alibaba Cloud.

2. No part of this document shall be excerpted, translated, reproduced, transmitted, or disseminated by any organization, company, or individual in any form or by any means without the prior written consent of Alibaba Cloud.

3. The content of this document may be changed due to product version upgrades, adjustments, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and the updated versions of this document will be occasionally released through Alibaba Cloud-authorized channels. You shall pay attention to the version changes of this document as they occur and download and obtain the most up-to-date version of this document from Alibaba Cloud-authorized channels.

4. This document serves only as a reference guide for your use of Alibaba Cloud products and services. Alibaba Cloud provides the document in the context that Alibaba Cloud products and services are provided on an "as is", "with all faults" and "as available" basis. Alibaba Cloud makes every effort to provide relevant operational guidance based on existing technologies. However, Alibaba Cloud hereby makes a clear statement that it in no way guarantees the accuracy, integrity, applicability, and reliability of the content of this document, either explicitly or implicitly. Alibaba Cloud shall not bear any liability for any errors or financial losses incurred by any organizations, companies, or individuals arising from their download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, bear responsibility for any indirect, consequent
ial, exemplary, incidental, special, or punitive damages, including lost profits arising from the use or trust in this document, even if Alibaba Cloud has been notified of the possibility of such a loss.

5. By law, all the contents in Alibaba Cloud documents, including but not limited to pictures, architecture design, page layout, and text description, are intellectual property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade secrets. No part of this document shall be used, modified, reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates. The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion, or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos, marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates.

6. Please contact Alibaba Cloud directly if you discover any errors in this document.
# Document conventions

<table>
<thead>
<tr>
<th>Style</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>🚨</td>
<td>A danger notice indicates a situation that will cause major system changes, faults, physical injuries, and other adverse results.</td>
<td>Danger: Resetting will result in the loss of user configuration data.</td>
</tr>
<tr>
<td>⚠️</td>
<td>A warning notice indicates a situation that may cause major system changes, faults, physical injuries, and other adverse results.</td>
<td>Warning: Restarting will cause business interruption. About 10 minutes are required to restart an instance.</td>
</tr>
<tr>
<td>🚦</td>
<td>A caution notice indicates warning information, supplementary instructions, and other content that the user must understand.</td>
<td>Notice: If the weight is set to 0, the server no longer receives new requests.</td>
</tr>
<tr>
<td>📄</td>
<td>A note indicates supplemental instructions, best practices, tips, and other content.</td>
<td>Note: You can use Ctrl + A to select all files.</td>
</tr>
<tr>
<td>&gt;</td>
<td>Closing angle brackets are used to indicate a multi-level menu cascade.</td>
<td>Click Settings &gt; Network &gt; Set network type.</td>
</tr>
<tr>
<td><strong>Bold</strong></td>
<td>Bold formatting is used for buttons, menus, page names, and other UI elements.</td>
<td>Click OK.</td>
</tr>
<tr>
<td><code>Courier font</code></td>
<td>Courier font is used for commands.</td>
<td>Run the <code>cd /d C:/window</code> command to enter the Windows system folder.</td>
</tr>
<tr>
<td><em>Italic</em></td>
<td>Italic formatting is used for parameters and variables.</td>
<td><code>bae log list --instanceid Instance_ID</code></td>
</tr>
<tr>
<td>[ ] or [a</td>
<td>b]</td>
<td>This format is used for an optional value, where only one item can be selected.</td>
</tr>
<tr>
<td>Style</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>-----------</td>
<td>------------------------------------------------------------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>{} or {a</td>
<td>b}</td>
<td>This format is used for a required value, where only one item can be selected.</td>
</tr>
</tbody>
</table>
Contents

Legal disclaimer.................................................................................................................... I
Document conventions........................................................................................................ I
1 Log collection methods..................................................................................................... 1
2 Collection acceleration...................................................................................................... 5
   2.1 Overview.................................................................................................................. 5
   2.2 Enable Global Acceleration...................................................................................... 9
   2.3 Configure Logtail collection acceleration................................................................ 13
   2.4 Disable Global Acceleration..................................................................................... 15
3 Logtail collection............................................................................................................. 17
   3.1 Overview.................................................................................................................. 17
   3.1.1 Overview............................................................................................................ 17
   3.1.2 Logtail collection process.................................................................................... 23
   3.1.3 Logtail configuration and recording files............................................................ 27
   3.2 Select a network type.............................................................................................. 37
   3.3 Install....................................................................................................................... 42
   3.3.1 Install Logtail in Linux......................................................................................... 42
   3.3.2 Install Logtail in Windows.................................................................................... 52
   3.3.3 Set startup parameters......................................................................................... 59
   3.4 Machine Group...................................................................................................... 66
   3.4.1 Overview............................................................................................................ 66
   3.4.2 Create an IP address-based server group............................................................ 68
   3.4.4 Configure AliUids for ECS servers under other Alibaba Cloud accounts or on-premises IDCs................................................................................................................................. 72
   3.4.5 Manage collection configurations....................................................................... 74
   3.4.6 Manage server groups....................................................................................... 76
   3.5 Text logs................................................................................................................ 81
   3.5.1 Configuration procedure for text log collection.................................................. 81
   3.5.2 Collect logs by line............................................................................................. 91
   3.5.3 Use regular expressions to collect logs............................................................. 97
   3.5.4 DSV formatted logs.......................................................................................... 104
   3.5.5 JSON logs......................................................................................................... 114
   3.5.6 NGINX logs...................................................................................................... 121
   3.5.7 IIS logs.............................................................................................................. 128
   3.5.8 Apache logs...................................................................................................... 136
   3.5.9 Configure and parse text logs.......................................................................... 145
   3.5.10 Configure the time format............................................................................... 148
   3.5.11 Import history logs........................................................................................... 152
   3.5.12 Set a log topic.................................................................................................. 155
3.6 Container log collection............................................................................................. 159
   3.6.1 Kubernetes log collection process........................................................................ 159
4 Cloud product collection

4.1 Cloud service logs
4.2 API Gateway access logs
4.3 MNS logs
4.4 OSS access logs
4.5 SLB Layer 7 access log
4.6 DDoS log collection
4.7 Logs of BGP-line Anti-DDoS Pro
4.8 TDS logs
4.9 WAF logs
4.10 Anti-Bot logs
4.11 ActionTrail access logs................................................................. 395
  4.11.1 Overview............................................................................. 395
  4.11.2 Procedure........................................................................ 399
4.12 Inner-ActionTrail........................................................................ 405
  4.12.1 Overview............................................................................. 405
  4.12.2 Activate Inner-ActionTrail.................................................... 407
  4.12.3 Inner-ActionTrail fields....................................................... 410
4.13 RDS SQL execution logs........................................................... 412
  4.13.1 Overview............................................................................. 412
  4.13.2 Procedure........................................................................ 415
  4.13.3 Log fields........................................................................ 418
4.14 NAS access logs.......................................................................... 419
  4.14.1 Overview............................................................................. 419
  4.14.2 Procedure........................................................................ 421
4.15 Cloud Firewall logs.................................................................... 425
  4.15.1 Overview............................................................................. 425
  4.15.2 Log analysis billing method............................................... 427
  4.15.3 Enable the log analysis service.......................................... 429
  4.15.4 Collect the log................................................................... 430
  4.15.5 Log analysis....................................................................... 434
  4.15.6 Fields in the log entry....................................................... 441
  4.15.7 Advanced Settings............................................................. 444
  4.15.8 Export log entries............................................................... 445
  4.15.9 Authorize RAM user accounts with Log Analysis function..... 446
  4.15.10 Manage log storage......................................................... 450
4.16 SMS logs..................................................................................... 451
5 Other collection methods............................................................. 453
  5.1 Web Tracking........................................................................... 453
  5.2 Use DataWorks to export MaxCompute data to Log Service....... 457
  5.3 Kafka protocol......................................................................... 465
  5.4 Collect data by using Syslog.................................................... 476
  5.5 Logstash.................................................................................. 483
    5.5.1 Install Logstash................................................................. 483
    5.5.2 Create Logstash collection configurations.......................... 485
    5.5.3 Set Logstash as a Windows service.................................... 488
    5.5.4 Advanced functions......................................................... 489
    5.5.5 Logstash error processing............................................... 490
  5.6 SDK collection.......................................................................... 490
    5.6.1 Producer Library............................................................... 490
    5.6.2 Log4j Appender.................................................................. 490
    5.6.3 Logback Appender............................................................ 491
    5.6.4 C Producer Library............................................................ 491
    5.6.5 Go Producer Library.......................................................... 492
  5.7 Common log formats.................................................................. 492
    5.7.1 Log4j logs........................................................................ 492
5.7.2 Python logs.................................................................494
5.7.3 Node.js logs................................................................. 500
5.7.4 WordPress logs............................................................ 501
5.7.5 ThinkPHP logs..............................................................502
5.7.6 Use Logstash to collect IIS logs........................................504
5.7.7 Use Logstash to collect CSV logs.......................................505
5.7.8 Use Logstash to collect other logs.....................................508
5.7.9 Unity3D logs................................................................. 509

6 Best practices.................................................................512
6.1 Collect IoT or embedded development logs...............................512
6.2 Collect logs through Web Tracking........................................522
6.3 Collect Container Service logs..............................................530
6.4 Build a service to read logs from mobile apps directly.................535
6.5 Collect data over the public network....................................541
6.6 Collect data from multiple channels.....................................544
6.7 Manage logs......................................................................551
6.8 Use logging handlers to automatically upload Python logs...........555

7 FAQ..................................................................................560
7.1 Troubleshoot collection errors..............................................560
7.2 What can I do if the Logtail client has no heartbeat?..................561
7.3 Do I need to update Logtail settings after the network type is changed?569
7.4 Diagnose collection errors..................................................570
7.5 Log collection error types..................................................572
7.6 Query local collection status...............................................589
7.7 How do I use the Logtail automatic diagnostic tool?................565
7.8 What are the differences between LogHub and Kafka?............561
7.9 What are the differences among log collection agents?.............615
7.10 Collect logs in complete regular mode................................621
7.10.1 How do I modify a regular expression?..........................621
7.10.2 How do I optimize regular expressions?............................624
7.10.3 How do I collect various formats of logs in complete regular
    mode?...........................................................................624
7.11 Why am I unable to collect SLB access logs?.........................625
7.12 How do I set the time format?............................................627
7.13 Troubleshoot log collection exceptions in containers..............628
1 Log collection methods

LogHub supports multiple methods to collect logs, such as by using clients, Web pages, protocols, SDKs, and APIs. All collection methods are implemented based on Restful APIs. You can also implement new collection methods by using APIs and SDKs.

Data sources

The following table describes the data sources from which Log Service can collect logs.

<table>
<thead>
<tr>
<th>Category</th>
<th>Source</th>
<th>Collection method</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program output</td>
<td></td>
<td>Logtail</td>
<td></td>
</tr>
<tr>
<td>Access log</td>
<td></td>
<td>Logtail</td>
<td>#unique_5</td>
</tr>
<tr>
<td>Link tracking</td>
<td>Jaeger Collector</td>
<td>and Logtail</td>
<td></td>
</tr>
<tr>
<td>Java</td>
<td></td>
<td>Log Service Java SDK</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>and Java Producer Library</td>
<td></td>
</tr>
<tr>
<td>Log4j appender</td>
<td>1.x and 2.x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logback appender</td>
<td>Logback</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>Log Service C SDK</td>
<td></td>
</tr>
<tr>
<td>Python</td>
<td></td>
<td>Log Service Python SDK</td>
<td></td>
</tr>
<tr>
<td>Python logging</td>
<td></td>
<td>Python logging handler</td>
<td></td>
</tr>
<tr>
<td>PHP</td>
<td></td>
<td>Log Service PHP SDK</td>
<td></td>
</tr>
<tr>
<td>C#</td>
<td></td>
<td>Log Service C SDK</td>
<td></td>
</tr>
<tr>
<td>C++</td>
<td></td>
<td>Log Service C++ SDK</td>
<td></td>
</tr>
<tr>
<td>Go</td>
<td></td>
<td>Log Service Go SDK</td>
<td></td>
</tr>
<tr>
<td>Node.js</td>
<td></td>
<td>Node.js</td>
<td></td>
</tr>
<tr>
<td>JavaScript</td>
<td></td>
<td>JavaScript/Web tracking</td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td>Source</td>
<td>Collection method</td>
<td>Reference</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-------------------------</td>
<td>--------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Operating system (OS)</td>
<td>Linux</td>
<td>Logtail</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Windows</td>
<td>Logtail</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Mac OS or Unix</td>
<td>Native C</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Docker files</td>
<td>Use Logtail to collect Docker files</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Docker output</td>
<td>Use Logtail to collect container logs</td>
<td>-</td>
</tr>
<tr>
<td>Mobile client</td>
<td>iOS/Android</td>
<td>Log Service Android SDK, Log Service iOS SDK</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Web page</td>
<td>JavaScript/Web tracking</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Intelligent IoT</td>
<td>C Producer Library</td>
<td>-</td>
</tr>
<tr>
<td>Alibaba Cloud services</td>
<td>ECS, OSS, and other Alibaba Cloud services. For more information, see Cloud service logs.</td>
<td>Activate Log Service in the Alibaba Cloud console</td>
<td>Cloud service logs</td>
</tr>
<tr>
<td></td>
<td>MaxCompute import</td>
<td>Use DataWorks to export MaxCompute data</td>
<td>Use DataWorks to export MaxCompute data to Log Service</td>
</tr>
<tr>
<td>Third-party software</td>
<td>Logstash</td>
<td>Logstash</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Flume</td>
<td>#unique_15</td>
<td>-</td>
</tr>
</tbody>
</table>

The following table lists the Alibaba Cloud services from which Log Service can collect logs.

<table>
<thead>
<tr>
<th>Type</th>
<th>Cloud service</th>
<th>Activation method</th>
<th>Documentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elastic computing</td>
<td>ECS</td>
<td>Install Logtail.</td>
<td>Overview</td>
</tr>
<tr>
<td></td>
<td>Container Service or Alibaba Cloud Container Service for Kubernetes</td>
<td>Activate the service in the Container Service or Container Service for Kubernetes console.</td>
<td>Container text logs and Container standard output</td>
</tr>
<tr>
<td>Type</td>
<td>Cloud service</td>
<td>Activation method</td>
<td>Documentation</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------------</td>
<td>------------------------------------------------------------------------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>Storage</td>
<td>OSS</td>
<td>Activate the service in the OSS console.</td>
<td>Overview</td>
</tr>
<tr>
<td>Networking</td>
<td>SLB</td>
<td>Activate the service in the SLB console.</td>
<td>SLB Layer-7 Access logs</td>
</tr>
<tr>
<td></td>
<td>Virtual Private Cloud (VPC)</td>
<td>Activate the service in the VPC console.</td>
<td>#unique_18</td>
</tr>
<tr>
<td></td>
<td>API Gateway</td>
<td>Activate the service in the API Gateway console.</td>
<td>API Gateway access logs</td>
</tr>
<tr>
<td>Security</td>
<td>ActionTrail</td>
<td>Activate the service in the ActionTrail console.</td>
<td>Overview</td>
</tr>
<tr>
<td></td>
<td>Security Center</td>
<td>Purchase Security Center Enterprise Edition and activate the service in the Security Center console.</td>
<td>TDS logs</td>
</tr>
<tr>
<td></td>
<td>Anti-Bot Service</td>
<td>Activate the service in the Anti-Bot Service console.</td>
<td>Anti-Bot Service logs</td>
</tr>
<tr>
<td>Application</td>
<td>Log Service</td>
<td>Activate the service in the Log Service console.</td>
<td>#unique_24</td>
</tr>
<tr>
<td></td>
<td>Short Message Service (SMS)</td>
<td>Activate the service in the Short Message Service console.</td>
<td>SMS logs</td>
</tr>
</tbody>
</table>

Select a network

Log Service provides service endpoints for different Alibaba Cloud regions. For more information, see #unique_26. Each region allows access from the following networks:
• Internal network (classic network) or private network (VPC): Log Service can access other Alibaba Cloud services in the same region, offering optimal link bandwidth. We recommend that you select this option.

• Public network (classic network): accessible without any limits. The transmission speed depends on the link quality. We recommend that you use HTTPS to ensure secure transmission of data.

FAQ

• Q: Which network do I select for private line access?
  
  A: Select the internal network or private network.

• Q: Can I collect public IP addresses when collecting public network data?
  
  A: You need to enable Log Service to record public IP addresses. For more information, see #unique_27.

• Q: Which network do I select if I want to collect ECS logs from Region A and write these logs into the Log Service project in Region B?
  
  A: Select the public network. You can install Logtail on the ECS instance in Region A for Internet transmission and specify the service endpoint that is associated with Region B. For more information about how to select a network, see Select a network type.

• Q: How can I determine whether a service endpoint is accessible?
  
  A: You can run the following command. The service endpoint is accessible if any information is returned.

```bash
curl $myproject.cn-hangzhou.log.aliuncs.com
```

$myproject specifies the project name and cn-hangzhou.log.aliuncs.com specifies the service endpoint.
2 Collection acceleration

2.1 Overview

In addition to Virtual Private Cloud (VPC) and the Internet, Log Service adds a network type of Internet-based Global Acceleration. Compared with the ordinary Internet access, Internet-based Global Acceleration has significant advantages in terms of latency and stability. It is suitable for scenarios with high requirements for data collection, low consumption latency, and reliability. Global Acceleration for Log Service depends on the acceleration environment provided by Alibaba Cloud Dynamic Route for CDN. Cross-carrier access, network instability, burst traffic, and network congestion used to cause problems such as slow response, packet loss, and unstable services. In this acceleration environment, Alibaba Cloud has solved such problems to improve overall performance and user experience.

Global Acceleration for Log Service is based on Alibaba Cloud Content Delivery Network (CDN) hardware resources. Alibaba Cloud has optimized the stability of log collection and data transmission from various forms of data sources, such as mobile phones, Internet of Things (IoT) devices, smart devices, on-premises Internet Data Centers (IDCs), and other cloud servers.

Technical principles

Global Acceleration for Log Service is based on Alibaba Cloud CDN hardware resources. Your terminals (such as mobile phones, IoT devices, smart devices
on-premises IDCs, and other cloud servers) can access the nearest edge node of Alibaba Cloud CDN all over the world and be routed to Log Service through the inner high-speed channels of CDN. Compared with data transmission on the Internet, this method can greatly reduce the network latency and jitter.

The preceding figure shows the flowchart for processing Global Acceleration requests for Log Service. The overall process is described as follows:

1. Before sending requests for log upload or log download to the accelerating domain name `your-project.log-global.aliyuncs.com` of Log Service, the client needs to send a domain name resolution request to the public DNS.

2. The public DNS resolves the domain name `your-project.log-global.aliyuncs.com` into the CNAME `your-project.log-global.aliyuncs.com.w.kunlungr.com`. The domain name resolution request is then forwarded to the CNAME in Alibaba Cloud CDN.

3. Based on the intelligent scheduling system, Alibaba Cloud CDN returns the IP address of the optimal edge node to the public DNS.

4. The public DNS returns the resolved IP address to the client.

5. The client sends a request to the server based on the obtained IP address.

6. After receiving the request, the CDN edge node uses dynamic routing and a private transport protocol to route the request to the node nearest to the Log Service server. The request is then forwarded to Log Service.

7. After receiving the request from the CDN edge node, the Log Service server returns the result of the request to the CDN edge node.

8. CDN transparently transmits the result or data returned by Log Service to the client.
Billing methods

**Global Acceleration costs for Log Service include:**

- **Cost for accessing Log Service**
  
  Log Service charges the access in pay-as-you-go mode, which is the same as Internet access. Log Service also provides certain FreeTier quota. For more information, see #unique_31.

- **Service cost for Dynamic Route for CDN**
  
  For more information, see the pricing of Dynamic Route for CDN.

**Scenarios**

- **Advertising**
  
  Log data about ad views and clicks is crucial to the billing of ads. In addition, advertising carriers are distributed all over the world, including mobile terminals, HTML 5 pages, and PCs. In some remote areas, data transmission is less stable on the Internet and logs may be lost during transmission. In this scenario, Global Acceleration for Log Service can provide a more stable and reliable channel for you to upload logs.

- **Online game**
  
  The online game industry raises high requirements for the performance and stability of data collection from various sources, such as the official website, logon service, sales service, and game service. In scenarios where data is collected from mobile games or transmitted from globalized games, timely and...
stable data collection is hard to be guaranteed. We recommend that you use Global Acceleration for Log Service to resolve the preceding issues.

- **Finance**

Financial applications require a highly available and secured network. Audit logs of each transaction and each user operation must be collected securely and reliably on the server. At present, mobile transactions are popular, such as online banking, credit card malls, and mobile securities. HTTPS Global Acceleration for Log Service can provide a secure, fast, and stable channel for you to collect logs for such transactions.

- **IoT**

IoT devices and smart devices (such as smart speakers and smart watches) send collected sensor data, operations logs, critical system logs, and other data to the server for data analysis. These devices are usually distributed all over the world. The surrounding network is not always reliable. To collect logs stably and reliably, we recommend that you use Global Acceleration for Log Service.

**Acceleration effects**

<table>
<thead>
<tr>
<th>Region</th>
<th>Latency in ms (Internet)</th>
<th>Latency in ms (Global Acceleration)</th>
<th>Percentage of timed-out requests (Internet)</th>
<th>Percentage of timed-out requests (Global Acceleration)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hangzhou</td>
<td>152.881</td>
<td>128.501</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Europe</td>
<td>1750.738</td>
<td>614.227</td>
<td>0.5908</td>
<td>0.0</td>
</tr>
<tr>
<td>United States</td>
<td>736.614</td>
<td>458.340</td>
<td>0.0010</td>
<td>0.0</td>
</tr>
<tr>
<td>Singapore</td>
<td>567.287</td>
<td>277.897</td>
<td>0.0024</td>
<td>0.0</td>
</tr>
<tr>
<td>Middle East</td>
<td>2849.070</td>
<td>444.523</td>
<td>1.0168</td>
<td>0.0</td>
</tr>
<tr>
<td>Australia</td>
<td>1491.864</td>
<td>538.403</td>
<td>0.014</td>
<td>0.0</td>
</tr>
</tbody>
</table>

The test environment is as follows:

- Region of Log Service: China (Hohhot)
- Average upload packet size: 10 KB
- Test time range: one day (average)
- Request method: HTTPS
Log Service

2. Collection acceleration

- Request server: Alibaba Cloud Elastic Compute Service (ECS) (instance type: 1 vCPU 1 GiB)

Note:
The acceleration effects are for reference only.

2.2 Enable Global Acceleration

This topic describes how to enable the Global Acceleration function of Log Service.

Prerequisites

- Log Service is activated. A project and a Logstore are created.
- Dynamic Route for CDN (DCDN) is activated.
- HTTP acceleration is enabled before you enable HTTPS acceleration.

Procedure

After you enable HTTP acceleration for the target project, you can also configure the Global Acceleration function for Logtail- and SDK-based log collection.

1. Enable HTTP acceleration.

2. Optional: Enable HTTPS acceleration.

If you use HTTPS to access Log Service, ensure that HTTPS acceleration has been enabled. For more information about how to configure HTTPS acceleration, see Enable HTTPS acceleration.

3. Collect logs in Global Acceleration mode.

- Logtail

- To install Logtail after Global Acceleration is enabled, you can follow the installation procedure for Global Acceleration in Install Logtail in Linux. Global
Acceleration automatically takes effect when logs are collected through Logtail.

- If Logtail has been installed before you enable Global Acceleration, you can manually switch the log collection mode of Logtail to Global Acceleration. For more information, see Configure Logtail collection acceleration.

- **SDK, Producer, or Consumer**

  If you use other methods such as the SDK, Producer, and Consumer to access Log Service, you can replace the configured endpoint with `log-global.aliyuncs.com` to achieve Global Acceleration.

Enable HTTP acceleration

1. **Log on to the DCDN console.**
2. In the left-side navigation pane, click Domain Names to go to the Domain Names page.
3. Click Add Domain Name in the upper-left corner to go to the Add Domain Name page.
4. Set DCDN Domain Name, specify other parameters as required, and then click Next.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCDN Domain Name</td>
<td>Enter <code>project_name.log-global.aliyuncs.com</code>. The <code>project_name</code> field must be replaced with your actual project name.</td>
</tr>
<tr>
<td>Type</td>
<td>Select Origin Domain.</td>
</tr>
<tr>
<td>Domain Name</td>
<td>Enter the public endpoint for the region of your project. For more information about endpoints, see <code>#unique_26</code>.</td>
</tr>
<tr>
<td>Port</td>
<td>Select Port 80. If you need HTTPS acceleration, you can configure HTTPS separately. For more information, see Enable HTTPS acceleration in this topic.</td>
</tr>
<tr>
<td>Parameter</td>
<td>Description</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Acceleration Region</td>
<td>The default value is Mainland China. If you need to use Global Acceleration, submit a ticket to apply for a whitelist from DCDN. After your application is approved, you can select acceleration regions based on your needs.</td>
</tr>
</tbody>
</table>

For more information about how to add domain names, see Add an accelerated domain name.

5. Return to the Domain Names page.

On the Domain Names page, you can view the CNAME of the added domain name.
   a) Log on to the Log Service console, and click the name of the target project.
   b) On the Overview page, find Global Acceleration, move your pointer over it, and then click Modify next to it.
   c) In the dialog box that appears, enter the CNAME of the accelerated domain name and click Enable Acceleration.

   ![Global Acceleration dialog box](image)

   Status: Unopened

   * Project Name: datav-k8s-log4j

   * Accelerated Domain: datav-k8s-log4j.log-global.aliyuncs.com

   * CNAME: test-project.log-global.aliyuncs.com.w.kunluncan.com

   More about Global Acceleration: Global Acceleration Introduction
   How to Enable: Enable Global Acceleration

   Enable Acceleration   Cancel

   After you complete the preceding steps, Global Acceleration is enabled for Log Service.

Enable HTTPS acceleration

If you need to use HTTPS to access Log Service after HTTP acceleration is enabled, you can enable HTTPS acceleration by performing the following steps:

1. Log on to the DCDN console. In the left-side navigation pane, click Domain Names to go to the Domain Names page.
2. Click Configure in the Actions column corresponding to the target domain name.
3. In the left-side navigation pane, click HTTPS Settings. On the page that appears, click Modify in the SSL Certificate section. The HTTPS Settings dialog box appears.
4. **Configure** SSL Acceleration **and** Certificate Type.
   
   a) **Enable** SSL Acceleration.
   
   b) **Set** Certificate Type to **Free Certificate**.
   
   c) Select **Agree to grant Alibaba Cloud permission to apply for a free certificate.**, and click OK.

### FAQ

- **How do I check whether the acceleration configuration takes effect?**

  After the configuration is completed, you can access the accelerated domain name configured in the DCDN console to check whether the acceleration configuration takes effect.

  For example, if Global Acceleration is enabled for the `test-project` project, you can run the `curl` command to send a request to the accelerated domain name.

  The following response indicates that the acceleration configuration takes effect:

  ```bash
  $curl test-project.log-global.aliyuncs.com
  {
  "Error":{
  "Code":"OLSInvalidMethod","Message":"The script name is invalid : /","RequestId":"5B55386A2CE41D1F4FBCF7E7"}
  }
  ```

- **What can I do if the project not exist error message appears when I access an accelerated domain name configured in the DCDN console?**

  This error is typically caused by an invalid origin domain name. You need to log on to the DCDN console and change the origin domain name to the public endpoint for the region of your project. For more information, see [Service endpoint](#).

  **Note:**
  
  The changes for the origin domain name take effect after several minutes.

## 2.3 Configure Logtail collection acceleration

After global acceleration is enabled, the Logtail that is installed in global acceleration mode automatically collects logs in global acceleration mode. For the Logtail that is installed before global acceleration is enabled, you need to manually
switch the acceleration mode to global acceleration by performing the steps in this topic.

Prerequisites

1. Enable HTTP acceleration.

2. (Optional) Enable HTTP acceleration.
   
   If you use HTTPS to access Log Service, make sure that HTTPS acceleration has been enabled and that you have configured HTTPS acceleration by following the instructions provided in Enable HTTP acceleration.

3. Make sure that acceleration functions normally by following the instructions provided in Enable Global Acceleration.

Before you begin

Before you configure Logtail collection acceleration, note that:

- If the Logtail is installed after global acceleration enabling, you must set the installation mode to global acceleration by following the instructions provided in Install Logtail in Linux. Then, the Logtail collects logs using global acceleration mode methods.

- If the Logtail is installed before global acceleration is enabled, you must switch the Logtail collection mode to global acceleration by performing the steps in this topic.

Switch the Logtail collection mode to global acceleration.

1. Stop the Logtail.

   - In Linux, run /etc/init.d/ilogtaild stop as the admin user.
   - In Windows:
     b. Open the Services program and locate the LogtailWorker file.
     c. Right-click the file and click Stop in the shortcut menu.

2. Modify the Logtail startup configuration file ilogtail_config.json.

   Change the endpoint in data_server_list to log-global.aliyuncs.com by following the instructions provided in Startup configuration file (ilogtail_config.json).
3. Start the Logtail.
   - In Linux, run `/etc/init.d/ilogtaild start` as the admin user.
   - In Windows:
     b. Open the Services program and locate the LogtailWorker file.
     c. Right-click the file and click Start in the shortcut menu.

2.4 Disable Global Acceleration

   This topic describes how to disable the DCDN-based Global Acceleration function of Log Service.

Context

   After the Global Acceleration function is disabled, the configured accelerated domain name cannot be used. Before you disable the Global Acceleration function, make sure that your users no longer use this domain name to upload or request data.

Procedure

1. Log on to the DCDN console.

2. In the left-side navigation pane, click Domain Names to go to the Domain Names page.

3. Query the CNAME corresponding to the domain name that will be disabled.
4. Disable the Global Acceleration function in Log Service.
   
a) Log on to the Log Service console, and click the name of the target project.

b) On the Overview page, find Global Acceleration, move your pointer over it, and then click Modify next to it.

c) In the dialog box that appears, enter the CNAME corresponding to the accelerated domain name and click Disable Acceleration.
3 Logtail collection

3.1 Overview

3.1.1 Overview

The Logtail access service is a log collection agent provided by Log Service. You can use Logtail to collect logs from servers such as Alibaba Cloud Elastic Compute Service (ECS) instances in real time in the Log Service console.

Figure 3-1: Function advantages

Benefits

- Non-invasive log collection based on log files. You do not have to modify codes of any application, and log collection does not affect the operating logic of your applications.
- In addition to text log collection, more collection methods are supported, such as binlog, http, and container stdout.
- Containers are well supported. This service supports data collection in standard containers, swarm clusters, and Kubernetes clusters.
- Logtail handles exceptions occurred in the log collection process. When problems (such as the network or Log Service is abnormal, and the user data
temporarily exceeds the reserved bandwidth writing limit) occur, Logtail actively retries and caches data locally to guarantee the data security.

- Centralized management capability based on Log Service. After installing Logtail, you can configure settings such as the machines from which logs are to be collected and the collection method in Log Service in a centralized way, without logging on to the servers and configuring settings separately. For how to install Logtail, see Install Logtail in Windows and Install Logtail in Linux.

- Comprehensive self-protection mechanism. To make sure that the collection agent running on your machine does not significantly affect the performance of your services, the Logtail client strictly protects and limits the usage of CPU, memory, and network resources.

Processing capabilities and limits

See Logtail limits.
Follow these steps to use Logtail to collect logs from servers:

1. **Install Logtail.** Install Logtail on the servers from which logs are to be collected. For more information, see Install Logtail in Windows and Install Logtail in Linux.

2. **Skip this step if you are about to collect logs from Alibaba Cloud ECS instances.**

3. **Create an IP address-based server group.** Log Service manages all the servers from which logs are to be collected by using the Logtail client in the form of machine groups. Log Service allows you to define machine groups by using IP addresses or custom identifiers. You can create a machine group as instructed when applying Logtail configurations to machine groups.

4. Create a Logtail collection configuration and apply it to the machine group. You can collect data such as Configuration procedure for text log collection and #unique_47 by
creating a Logtail configuration in the data import wizard. Then, you can apply the Logtail configuration to the machine group.

After completing the preceding steps, incremental logs on servers from which logs are to be collected are actively collected and sent to the corresponding Logstore. Historical logs are not collected. You can query these logs in the console or by using APIs/SDKs. You can also query the Logtail log collection status in the console, such as check whether the collection is normal or if any error occurs.

For the complete procedure for Logtail access service in the Log Service console, see Collect text logs.

Container

- Alibaba Cloud Container Service Swarm cluster: see #unique_48.
- Alibaba Cloud Container Service Kubernetes cluster: see Kubernetes log collection process.
- Self-built Kubernetes: see Self-built Kubernetes installation
- Other self-built Docker clusters: see Collect standard Docker logs

Major concepts

- Machine group: A machine group contains one or more machines from which a type of logs is to be collected. By applying a Logtail configuration to a machine group, Log Service collects logs from all the machines in the machine group according to the same Logtail configuration. You can also manage a machine group in the Log Service console, such as creating/deleting a machine group, and adding/removing a machine to/from a machine group. You must note that a single machine group cannot contain a mix of Windows and Linux machines, but may have machines with different versions of Windows Server or different release versions of Linux.
- Logtail client: Logtail is the agent that collects logs and runs on servers from which logs are to be collected. For how to install Logtail, see Install Logtail in Windows and Install Logtail in Linux. After installing Logtail on the server, create a Logtail configuration and then apply it to a machine group.
  - In Linux, Logtail is installed in the /usr/local/ilogtail directory and starts two independent processes (a collection process and a daemon process) whose
names start with ilogtail. The program running log is `/usr/local/ilogtail/ilogtail.LOG`.

- In Windows, Logtail is installed in the `C:\Program Files\Alibaba\Logtail` directory (for 32-bit system) or the `C:\Program Files (x86)\Alibaba\Logtail` directory (for 64-bit system). Navigate to Windows Administrative Tools > Services, you can view two Windows services: LogtailWorker and LogtailDaemon. LogtailWorker is used to collect logs and LogtailDaemon works as a daemon. The program running log is `logtail_*.log` in the installation directory.

- Logtail configuration: Logtail configuration is a collection of policies to collect logs by using Logtail. By configuring Logtail parameters such as data source and collection mode, you can customize the log collection policy for all the machines in the machine group. A Logtail configuration is used to collect a type of logs from machines, parse the collected logs, and send them to a specified Logstore of Log Service. You can add a Logtail configuration for each Logstore in the console to enable the Logstore to receive logs collected by using this Logtail configuration.

Basic functions

The Logtail access service provides the following functions:

- Real-time log collection: Logtail dynamically monitors log files, and reads and parses incremental logs in real time. Generally, a delay of less than three seconds exists between the time when a log is generated and the time when a log is sent to Log Service.

  [Note:

  Logtail does not support collection of historical data. Logs with an interval of more than 12 hours between the time when a log is read and the time when a log is generated are discarded.

- Automatic log rotation processing: Many applications rotate log files according to the file size or date. During the rotation process, the original log file is renamed and a new blank log file is created for log writing. For example, the monitored `app.LOG` is rotated to generate `app.LOG.1` and `app.LOG.2`. You can specify the file to which collected logs are written, for example, `app.LOG`. Logtail
automatically detects the log rotation process and guarantees that no log data is lost during this process.

- Multiple collection input sources: Besides text logs, Logtail supports the input sources such as syslog, HTTP, MySQL, and binlog. For more information, see Data Source in Log Service user guide.

- Compatible with open-source collection agent: Input source of Logtail can be data collected by open-source softwares, such as Logstash and Beats. For more information, see Data Source in Log Service user guide.

- Automatic handling of collection exceptions: When data transmission fails because of exceptions such as Log Service errors, network measures, and quota exceeding the limit, Logtail actively retries based on specific scenario. If the retry fails, Logtail writes the data to the local cache and then automatically resends the data later.

- Flexible collection policy configuration: You can use Logtail configuration to flexibly specify how logs are collected from a server. Specifically, you can select log directories and files, which support exact match or fuzzy match with wildcards, based on actual scenarios. You can customize the extraction method for log collection and the names of extracted fields. Log Service supports extracting logs by using regular expressions. The log data models of Log Service require that each log must have a precise timestamp. Therefore, Logtail provides custom log time formats, allowing you to extract the required timestamp information from log data of different formats.

- Automatic synchronization of collection configuration: Generally, after you create or update a configuration in the Log Service console, Logtail automatically accepts and brings the configuration into effect within three minutes. No collected data is lost when configuration is being updated.

- Automatic upgrade of client: After you manually install Logtail on a server, Log Service automatically performs the Operation & Maintenance (O&M) and upgrade of Logtail. No log data is lost when Logtail is being upgraded.

- Status monitoring: To prevent the Logtail client from consuming too many resources and thus affecting your services, the Logtail client monitors its consumption of CPU and memory in real time. The Logtail client is automatically restarted when its resource usage exceeds the limit to avoid affecting other operations on the machine. The Logtail client actively limits network traffic to avoid excessive bandwidth consumption.
- Data transmission with a signature: To prevent data tampering during the transmission process, the Logtail client obtains your Alibaba Cloud AccessKey (AK) and provides a signature to all log data packets to be sent.

**Note:**
To maintain security of your Alibaba Cloud AK, the Logtail client uses the HTTPS tunnel to obtain your AK.

Data collection reliability

During data collection, Logtail stores the collected checkpoint information to your local directory on a regular basis. If any exception occurs (such as the server unexpectedly shuts down or a process crashes), Logtail will collect data from the last recorded checkpoint after being restarted to prevent data loss. Then, Logtail functions according to the resource limits listed in the configuration file. However, if a resource is occupied for five minutes more than the preset time period, Logtail will be forcibly restarted. After the force restart, some of the existing data may be regenerated.

Although Logtail uses multiple methods to guarantee data collection reliability, absolute data integrity cannot be guaranteed. Specifically, data loss may occur due to the following reasons:

- Logtail is not functioning, but logs have been rotated several times.
- The log rotation speed is exceedingly high, for example, one time per second.
- The log collection speed is slower than the log generation speed for a long period of time.

3.1.2 Logtail collection process

The Logtail client performs the following six steps to collect logs from your server: monitor files, read files, process logs, filter files, aggregate logs, and send logs.

After you install the Logtail client on your server and configure a Logtail Config, Logtail starts collecting logs to Log Service. The log collection process involves the following steps:

1. **Monitor files**
2. **Read files**
3. **Process logs**
4. **Filter logs**

5. **Aggregate logs**

6. **Send logs**

---

**Note:**
After a Logtail Config is configured for a machine group, unmodified logs on a server in the machine group will be regarded as historical files. However, Logtail does not collect historical files. If you want to collect historical logs, see [Import history logs](#).

---

**Monitor files**

After you install the Logtail client on your server and configure a Logtail Config based on data sources, the Logtail Config sends logs to Logtail in real time. Then, Logtail uses the Logtail Config to monitor files.

1. Specifically, Logtail scans the log directories and files that conform to the specified file naming conventions layer by layer according to the configured log path and maximum monitoring directory depth.

   To ensure the efficiency and stability of log collection, Logtail registers event monitoring for the collection directory (namely, the `Inotify` directory on Linux or the `ReadDirectoryChangesW` directory on Windows) and performs periodic polling.

2. If the monitoring results show that unmodified log files that conform to the file naming conventions exist in the specified directory, Logtail will not collect the files. If there are modified log files, a collection process will be triggered and Logtail will read the files.

---

**Read files**

Logtail starts to read the modified files.

1. Logtail checks the size of a file when reading the file for the first time.
   - If the file size is smaller than 1 MB, Logtail reads the file from the beginning.
   - If the file size is larger than 1 MB, Logtail reads the last 1-MB content of the file.

2. If Logtail has read the file before, Logtail reads the file from the last checkpoint.

3. Logtail can read up to 512 KB at a time. Therefore, you need to limit the log size to 512 KB.
Log Service

Data Collection / 3 Logtail collection

Note:
If you have modified the time on your server, you need to manually restart Logtail. Otherwise, the log generation time will be incorrect and some logs may be mistakenly discarded.

Process logs

Logtail splits a log into lines, parses the log, and confirms the correctness of the time field settings.

1. Line splitting:
   If a line start regular expression has been specified in the Logtail Config, Logtail will split the log into lines according to the line start settings. In this case, Logtail processes the lines as multiple logs. If no line start regular expression has been specified, Logtail regards a data block as a log and processes it.

2. Parsing:
   Logtail uses the Logtail Config to parse the log content based on specified rules, such as regular expressions, delimiters, and JSON arrays.

   Note:
   An excessively complex regular expression may lead to an abnormally high CPU usage. Therefore, we recommend that you use an efficient regular expression.

3. Parsing failure handling:
   Depending on whether the discarding logs with parsing failure function is enabled in the Logtail Config, you can handle logs with parsing failure as follows:
   
   • If the function is enabled, Logtail discards the log and reports a corresponding error.
   • If the function is disabled, you need to upload the original log with its key of raw_log and Value of the log content.
4. Time field settings:
   - If the time field is not set, the log generation time is the current parsing time.
   - If the time field is set and the log generation time is:
     - Less than 12 hours from the current time, Logtail extracts the time from the parsed time field.
     - More than 12 hours from the current time, Logtail discards the log and reports a corresponding error.

Filter logs

Logtail filters logs according to the filter settings in the Logtail Config.
   - If the filter is not set, Logtail will not filter logs but directly aggregates logs.
   - If the filter is set, Logtail will traverse and verify all fields in each log.
     - Logtail collects logs that conform to filter settings, that is, all fields in filter settings can be found in the log and all the fields conform to the setting requirements.
     - Logtail does not collect logs that do not conform to filter settings.

Aggregate logs

Logtail sends log data to Log Service. To reduce the number of network requests, Logtail caches the logs for some time. Then, Logtail aggregates and packages the logs to send them to Log Service.

During caching, Logtail will immediately package logs and send them if any of the following conditions is met:
   - Log aggregation lasts more than 3s.
   - There are more than 4.096 logs to be aggregated.
   - The target log size exceeds 512 KB.

Send logs

Logtail sends the aggregated log to Log Service. You can set the startup parameters max_bytes_per_sec and send_request_concurrency by following the instructions provided in Set startup parameters to adjust the log sending rate and the maximum number of logs that can be concurrently sent. In this case, Logtail ensures that the preset values are not exceeded.
If the log sending fails, Logtail automatically retries or quits the task according to the corresponding error message.

<table>
<thead>
<tr>
<th>Error message</th>
<th>Description</th>
<th>Handling method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error code: 401</td>
<td>The Logtail client does not have the permission to collect data.</td>
<td>Logtail discards the log package</td>
</tr>
<tr>
<td>Error code: 404</td>
<td>The project or Logstore specified in the Logtail Config does not exist.</td>
<td>Logtail discards the log package</td>
</tr>
<tr>
<td>Error code: 403</td>
<td>The Shard quota exceeds the upper limit.</td>
<td>Wait for 3s and try again.</td>
</tr>
<tr>
<td>Error code: 500</td>
<td>An error occurs on the server.</td>
<td>Wait for 3s and try again.</td>
</tr>
<tr>
<td>Network expiration</td>
<td>A network connection error occurs.</td>
<td>Wait for 3s and try again.</td>
</tr>
</tbody>
</table>

3.1.3 Logtail configuration and recording files

The running of Logtail depends on a series of configuration files, which generates specific information recording files. This topic describes the basic information and paths of commonly generated files.

**Configuration files:**

- Startup configuration file (ilogtail_config.json)
- AliUid configuration file
- User-defined identity file (user_defined_id)
- Logtail Config file (user_log_config.json)

**Recording files:**

- AppInfo recording file (app_info.json)
- Logtail operational log file (ilogtail.LOG)
- Logtail plug-in log file (logtail_plugin.LOG)
- Container path mapping file (docker_path_config.json)

Startup configuration file (ilogtail_config.json)

The file is used to view or set Logtail running parameters. The file is in JSON format.

After installing Logtail, you can use the file to:
· Modify Logtail running parameters.

You can modify common settings, such as the CPU usage threshold and resident memory usage threshold by modifying the file.

· Check whether installation commands are correct.

In the file, config_server_address and data_server_list are determined by parameters and commands used during installation. If the region specified by the parameters is different from the region where Log Service resides or the address is inaccessible, incorrect parameters or commands are used during installation. In this case, Logtail cannot collect logs, and you need to reinstall it.

Note:

· The file must be valid JSON arrays. Otherwise, Logtail cannot be started.
· The modified file can take effect only after Logtail is restarted.

The following table lists default configuration items. For details about other configuration items, see Set startup parameters.

Table 3-1: Default configuration items in the startup configuration file

<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>config_server_address</td>
<td>Address of the configuration file Logtail obtains from your server. The address is determined by the parameters and commands you use during installation. The address must be accessible, and the region specified by the parameters must be the same as the region where Log Service resides.</td>
</tr>
<tr>
<td>data_server_list</td>
<td>Address of the data server, which is determined by the parameters and commands you use during installation. The address must be accessible, and the region specified by the parameters must be the same as the region where Log Service resides.</td>
</tr>
<tr>
<td>cluster</td>
<td>Region name</td>
</tr>
<tr>
<td>endpoint</td>
<td>Service endpoint</td>
</tr>
<tr>
<td>Configuration item</td>
<td>Description</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>cpu_usage_limit</td>
<td>CPU usage threshold, which is calculated by core</td>
</tr>
<tr>
<td>mem_usage_limit</td>
<td>Resident memory usage threshold</td>
</tr>
<tr>
<td>max_bytes_per_sec</td>
<td>Maximum amount of raw data Logtail can send. The amount will not be limited if the data sending rate exceeds 20 Mbit/s.</td>
</tr>
<tr>
<td>process_thread_count</td>
<td>Number of threads Logtail uses to write data to log files</td>
</tr>
<tr>
<td>send_request_concurrency</td>
<td>Number of data packets Logtail can send concurrently and asynchronously. By default, Logtail sends data packets asynchronously. You can set the configuration item to a larger value if the write TPS is excessively high.</td>
</tr>
</tbody>
</table>

**File address:**

- **Linux:** `/usr/local/ilogtail/ilogtail_config.json`
- **Container:** The file is stored in the Logtail container, and the file address is configured through the environment variable `ALIYUN_LOGTAIL_CONFIG`. You can view the address through `Docker inspect $ {logtail_container_name} | grep ALIYUN_LOGTAIL_CONFIG`, for example, `/Etc/ilogtail/CONF/CN-Hangzhou/FIG`.
- **Windows:**
  - x64: `C:\Program Files (x86)\Alibaba\Logtail\ilogtail_config.json`
  - x32: `C:\Program Files\Alibaba\Logtail\ilogtail_config.json`

**File example:**

```
$ cat /usr/local/ilogtail/ilogtail_config.json
{
  "config_server_address" : "http://logtail.cn-hangzhou-intranet.log.aliyuncs.com",
  "data_server_list" : [
    {
      "cluster" : "ap-southeast-2",
      "endpoint" : "cn-hangzhou-intranet.log.aliyuncs.com"
    }
  ],
  "cpu_usage_limit" : 0.4,
  "mem_usage_limit" : 100,
  "max_bytes_per_sec" : 2097152,
  "process_thread_count" : 1,
  "send_request_concurrency" : 4,
  "streamlog_open" : false
}
```
AliUid configuration file

The file contains the AliUid of your Alibaba Cloud account. AliUid is used to indicate that your Alibaba Cloud account has the permissions to access your server and collect logs. You need to manually create the AliUid configuration file when collecting logs from an ECS instance that does belong to your Alibaba Cloud account or from on-premises IDCs. For more information, see Configure AliUids for ECS servers under other Alibaba Cloud accounts or on-premises IDCs.

Note:

- This file is optional and is used only when you collect logs from an ECS instance that does belong to your Alibaba Cloud account or from on-premises IDCs.
- The file can only contain the AliUid of your Alibaba Cloud account. It cannot contain the AliUid of any RAM user account under your Alibaba Cloud account.
- The file name cannot contain any suffix.
- Logtail can be configured with multiple AliUid configuration files, but a Logtail container can be configured with only one AliUid configuration file.

File address

- **Linux**: `/etc/ilogtail/users/`
- **Container**: The file is directly configured through the environment variable `ALIYUN_LOGTAIL_USER_ID` in the Logtail container. You can view the file through `docker inspect ${logtail_container_name} | grep ALIYUN_LOGTAIL_USER_ID`
- **Windows**: `C:\LogtailData\users\`

File example

```bash
$ls /etc/ilogtail/users/
155912253502**** 132923253502****
```

User-defined identity file (user_defined_id)

The file is used to configure machine groups with custom identifiers. For more information, see #unique_44.

Note:
• This file is optional and is used only when configuring machine groups with custom identifiers.
• If multiple custom identifiers are configured for a machine group, they must be separated by delimiters.

File address

• **Linux:** `/etc/ilogtail/user_defined_id`
• **Container:** The file is directly configured through the environment variable `ALIYUN_LOGTAIL_USER_DEFINED_ID` in the Logtail container. You can view the file through `docker inspect ${logtail_container_name} | grep ALIYUN_LOGTAIL_USER_DEFINED_ID`.
• **Windows:** `C:\LogtailData\user_defined_id`

File example

```
$cat /etc/ilogtail/user_defined_id
aliyun-ecs-rs1e16355
```

Logtail Config file (user_log_config.json)

The file contains Logtail Config information Logtail obtains from your server. The file is in JSON format and is updated with Logtail Config updates. The file is used to check whether Logtail Config sends logs to your server. If the file exists and the file content is up-to-date, the Logtail Config has sent logs.

**Note:**

• We recommend that you do not modify the file unless you need to manually configure keys and modify database passwords.
• The file must be uploaded when you open a ticket.

File address

• **Linux:** `/usr/local/ilogtail/user_log_config.json`
• **Container:** `/usr/local/ilogtail/user_log_config.json`
• **Windows**
  - x64: `C:\Program Files (x86)\Alibaba\Logtail\user_log_config.json`
  - x32: `C:\Program Files\Alibaba\Logtail\user_log_config.json`
$cat /usr/local/ilogtail/user_log_config.json
{
   "metrics": {
      "##1.0##k8s-log-c12ba2028c*****ac1286939f0b$app-java": {
         "aliuid": "16542189*****50",
         "category": "app-java",
         "create_time": 1534739165,
         "defaultEndpoint": "cn-hangzhou-intranet.log.aliyuncs.com",
         "delay_alarm_bytes": 0,
         "enable": true,
         "enable_tag": true,
         "filter_keys": [],
         "filter_regs": [],
         "group_topic": "",
         "local_storage": true,
         "log_type": "plugin",
         "log_tz": ",",
         "max_send_rate": -1,
         "merge_type": "topic",
         "plugin": {
            "inputs": [
               {
                  "detail": {
                     "IncludeEnv": {
                        "aliyun_logs_app-java": "stdout"
                     },
                     "IncludeLabel": {
                        "io.kubernetes.container.name": "java-log-demo-2",
                        "io.kubernetes.pod.namespace": "default"
                     },
                     "Stderr": true,
                     "Stdout": true
                  }
               }
            ],
            "priority": 0,
            "project_name": "k8s-log-c12ba2028c*****ac1286939f0b",
            "raw_log": false,
            "region": "cn-hangzhou",
            "send_rate_expire": 0,
            "sensitive_keys": [],
            "tz_adjust": false,
            "version": 1
         }
      }
   }
}
In normal cases, Logtail obtains the server IP address according to the following rules:

- Logtail automatically obtains the IP address if the IP address has been attached to your host through the server file `/etc/hosts`.
- Logtail automatically obtains the IP address of the first NIC on your host if no IP address is attached to your host.

**Note:**
- The file only contains internal information about Logtail. Manual modifications to the file content do not change basic Logtail information.
- If you have modified network configurations of your server, for example, host name, you need to restart Logtail to obtain the new IP address.

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>UUID</td>
<td>Server serial number</td>
</tr>
<tr>
<td>hostname</td>
<td>Host name</td>
</tr>
<tr>
<td>instance_id</td>
<td>Randomly generated identifier for uniquely indicating Logtail</td>
</tr>
<tr>
<td>ip</td>
<td>IP address obtained by Logtail. An empty field indicates that Logtail does not obtain the IP address and cannot function normally. In this case, you need to set an IP address for your server and restart Logtail.</td>
</tr>
<tr>
<td>logtail_version</td>
<td>Version of the Logtail client</td>
</tr>
<tr>
<td>os</td>
<td>OS version</td>
</tr>
</tbody>
</table>
Log Service

Data Collection / 3 Logtail collection

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>update_time</td>
<td>Time when Logtail is last started</td>
</tr>
</tbody>
</table>

**Field address**

- **Linux:** /usr/local/iLogtail/app_info.json
- **Container:** /usr/local/iLogtail/app_info.json
- **Windows**
  - **x64:** C:\Program Files (x86)\Alibaba\Logtail\app_info.json
  - **x32:** C:\Program Files\Alibaba\Logtail\app_info.json

**File example**

```
$cat /usr/local/iLogtail/app_info.json
{
  "UUID" : "",
  "hostname" : "Logtail-ds-slpn8",
  "instance_id" : "E5F93BC6-B024-11E8-8831-0A58AC14039E_172.20.3.158_1536053315",
  "ip" : "172.20.3.158",
  "logtail_version" : "0.16.13",
  "os" : "Linux; 3.10.0-693.2.2.el7.x86_64; #1 SMP Tue Sep 12 22:26:13 UTC 2017; x86_64",
  "update_time" : "2018-09-04 09:28:36"
}
```

Logtail operational log file (iLogtail.LOG)

The file contains running information about the Logtail client. Log levels are ranked as follows in ascending order: INFO, WARN, ERROR. INFO-type logs can be ignored.

**Note:**

- First, you need to diagnose collection exceptions and troubleshoot errors according to specific error types and Logtail operational logs.
- The file must be uploaded when you open a ticket due to Logtail collection exceptions.

**File address**

- **Linux:** /usr/local/iLogtail/iLogtail.LOG
- **Container:** /usr/local/iLogtail/iLogtail.LOG
Windows

- x64: C:\Program Files (x86)\Alibaba\Logtail\logtail_*.log
- x32: C:\Program Files\Alibaba\Logtail\logtail_*.log

File example

$tail /usr/local/ilogtail/ilogtail.LOG
[2018-09-13 01:13:59.024679] [INFO] [3155] [build/release64/sls/ilogtail/elogtail.cpp:123] change working dir:/usr/local/ilogtail/
[2018-09-13 01:13:59.025443] [INFO] [3155] [build/release64/sls/ilogtail/AppConfig.cpp:175] load logtail config file, path:/etc/ilogtail/conf/ap-southeast-2/ilogtail_config.json
[2018-09-13 01:13:59.025460] [INFO] [3155] [build/release64/sls/ilogtail/AppConfig.cpp:176] load logtail config file, detail:
  "config_server_address" : "http://logtail.ap-southeast-2-intranet.log.aliyuncs.com",
  "data_server_list" : [
    {
      "cluster" : "ap-southeast-2",
      "endpoint" : "ap-southeast-2-intranet.log.aliyuncs.com"
    }
  ]

Logtail plug-in log file (logtail_plugin.LOG)

The file contains running information about the container stdout, binlogs, http plug-in, and other plug-ins. Log levels are ranked as follows in ascending order: INFO, WARN, ERROR. INFO-type logs can be ignored.

If there is any plug-in error, for example, CANAL_RUNTIME_ALARM, when you diagnose collection exceptions, you can troubleshoot the error according to Logtail plug-in logs.

Note:
The file must be uploaded when you open a ticket due to plug-in exceptions.

File address

- Linux: /usr/local/ilogtail/logtail_plugin.LOG
- Container: /usr/local/ilogtail/logtail_plugin.LOG
- Windows: plug-in logs are not supported.

File example

$tail /usr/local/ilogtail/logtail_plugin.LOG
2018-09-13 02:55:30 [INF] [docker_center.go:525] [func1] docker fetch all:start
Container path mapping file (docker_path_config.json)

The file is automatically created only when container files are collected. The file is used to record the mapping between the path of container files and the actual file path. The file is in JSON format.

When you diagnose collection exceptions, if an error indicating DOCKER_FILE_MAPPING_ALARM is reported, Logtail fails to add Docker file mapping. In this case, you can use the file to troubleshoot the error.

Note:

- The file only contains information. Any modification to the file does not take effect. The file will be automatically recreated once it is deleted. This does not impact services.
- The file must be uploaded when you open a ticket due to container log collection exceptions.
File address

/usr/local/ilogtail/docker_path_config.json

File example

```bash
$cat /usr/local/ilogtail/docker_path_config.json
{
  "detail": [
    {
      "config_name": "##1.0##k8s-log-c12ba2028c9f238cd
9ac1286939f0b$nginx",
      "container_id": "df19c06e854a072e7fca7e0378b0450f7bd312
450f7bd3122f94fe3e754d8483fd330d10",
      "params": "{"ID": "df19c06e854a072e7fca7e0378b0
450f7bd3122f94fe3e754d8483fd330d10",
      "Path": "/logtail_ho
st/var/lib/docker/overlay2/947db34695a1f65e63e582ecfd10ae1f57019a
b99260b683d00fcd1892874/diff/var/log",
      "Tags": ["nginx-type",
                 "access-log",
                 "registry.cn-hangzhou.aliyuncs.com/log-service/docker-log-test",
                 "nginx-log-demo",
                 "_container_name_",
                 "registry.cn-hangzhou.aliyuncs.com/log-service/docker-log-test",
                 "nginx-log-demo",
                 "_namespace_",
                 "87e56ac3-b65b-11e8-b172-00163f086885",
                 "_container_ip_",
                 "172.20.4.224",
                 "purpose",
                 "test"]
    }
  ]
  "version": "0.1.0"
}
```

3.2 Select a network type

The collected log data can be sent to Log Service through the Alibaba Cloud intranet, the Internet, or through Global Acceleration.

Network types

- **Internet**: Sending log data through the Internet can be limited by the network bandwidth. Additionally, network issues such as jitters, latency, and packet loss may affect the speed and stability of data transmission.

- **Alibaba Cloud intranet**: The Alibaba Cloud intranet supports shared bandwidth at the gigabit-level and can transmit log data more quickly and stably than the Internet. The intranet includes the Virtual Private Cloud (VPC) environment and the classic network environment.

- **Global Acceleration**: This network service accelerates log collection by using the edge nodes of Alibaba Cloud Content Delivery Network (CDN). Compared with the Internet, Global Acceleration provides lower transmission delay and greater stability.
Select a network type

- **intranet:**

  Whether your log data is transmitted through the Alibaba Cloud intranet depends on your server type and if the server and the Log Service Project are in the same region. The Alibaba Cloud intranet can transmit log data in only the following two scenarios:
  
  - The ECS instances of your account and the Log Service Project are in the same region.
  - The ECS instances of other accounts and the Log Service Project are in the same region.

  Therefore, we recommend that you create a Log Service Project in the region where your ECS instances reside, and collect logs to this Project. Then the log data of the ECS instances is written to Log Service through the Alibaba Cloud intranet, without consuming the Internet bandwidth.

  **Note:**

  When you install a Logtail client on a server, you must select the region in which the Log Service Project resides. Otherwise, the log data cannot be collected.

- **Global Acceleration:**

  If your servers are located in your self-built IDCs overseas, or your servers are hosted by overseas cloud vendors, using the Internet to transmit data may cause problems such as high latency and unstable transmission. In this case, you can use *Global Acceleration* instead. *Global Acceleration* accelerates log collection by using the edge nodes of Alibaba Cloud CDN. Compared with data transmission through the Internet, Global Acceleration offers a more stable network with minimal transmission delays.

- **Internet:**

  We recommend that you select the Internet for the following two scenarios:
  
  - The server is an ECS instance, but it does not reside in the same region as the Log Service Project.
  - The server is located in your own IDC or provided by a vendors.
### Log Service

#### Data Collection / 3 Logtail collection

<table>
<thead>
<tr>
<th>Server type</th>
<th>Reside in the same region as the Project</th>
<th>Configure an AliUid</th>
<th>Network type</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECS instances under your account</td>
<td>Yes</td>
<td>Not required</td>
<td>Alibaba Cloud intranet</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>Not required</td>
<td>Internet or Global Acceleration</td>
</tr>
<tr>
<td>ECS instances of other accounts</td>
<td>Yes</td>
<td>Required</td>
<td>Alibaba Cloud intranet</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>Required</td>
<td>Internet or Global Acceleration</td>
</tr>
<tr>
<td>Cloud vendor servers or your own IDC servers</td>
<td>-</td>
<td>Required</td>
<td>Internet or Global Acceleration</td>
</tr>
</tbody>
</table>

**Note:**

Log Service cannot obtain owner information of the ECS instances that are under other accounts or servers. Therefore, you need to configure an AliUid for each server after you complete the Logtail client installation. Otherwise, the server heartbeat is abnormal and the server logs cannot be collected. For more information, see *Configure AliUids for ECS servers under other Alibaba Cloud accounts or on-premises IDCs*.

#### Examples of selecting a network type

The following examples describe how to select an appropriate network in several common scenarios.

**Note:**

In the Global Acceleration scenario, the speed and reliability of data collection are important factors because the Log Service Project is created in the Hong Kong region but the servers are from the IDCs located worldwide. Therefore, we recommend that you select the Global Acceleration network type in the Hong Kong region when installing a Logtail client in similar scenarios. Compared with the Internet, Global Acceleration transmits log data with higher stability and performance.
<table>
<thead>
<tr>
<th>Scenario</th>
<th>Region of the Log Service Project</th>
<th>Server type</th>
<th>Region of the ECS instance</th>
<th>Selected region for installing a Logtail client</th>
<th>Network type</th>
<th>Configure an AliUid</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECS and the Project are in the same region.</td>
<td>China East 1 (Hangzhou)</td>
<td>ECS of your current account</td>
<td>China East 1 (Hangzhou)</td>
<td>China East 1 (Hangzhou)</td>
<td>intranet</td>
<td>Not required</td>
</tr>
<tr>
<td>ECS and the Project are in different regions.</td>
<td>China East 2 (Shanghai)</td>
<td>ECS of your current account</td>
<td>China North 1 (Beijing)</td>
<td>China North 1 (Beijing)</td>
<td>Internet</td>
<td>Not required</td>
</tr>
<tr>
<td>Other accounts</td>
<td>China East 2 (Shanghai)</td>
<td>ECS belongs to other accounts.</td>
<td>China North 1 (Beijing)</td>
<td>China North 1 (Beijing)</td>
<td>Internet</td>
<td>Required</td>
</tr>
<tr>
<td>Server is in the local IDC.</td>
<td>China East 5 (Shenzhen)</td>
<td>Self-built IDC</td>
<td>-</td>
<td>China East 5 (Shenzhen)</td>
<td>Internet</td>
<td>Required</td>
</tr>
</tbody>
</table>
Update configurations after a classic network is switched to a VPC

After a Logtail client is installed, you must update the network configurations if your ECS instance is switched from a classic network to a VPC. To do so, follow these steps:

1. Restart the Logtail client as the administrator.
   • Linux:

   ```bash
   sudo /etc/init.d/ilogtaild stop
   ```
sudo /etc/init.d/logtaild start

- Windows:
  Open Management Tool in Control Panel, open Service, right-click LogtailWorker, and then select Restart.

2. Update machine group configurations.

- Custom ID
  If a custom ID is set to define the machine group, you can directly use the VPC network without updating machine group configurations.

- IP address
  If the ECS instance IP address is used when you define the machine group, you must replace the original IP address with the new IP address obtained by the restarted Logtail client. That is, the IP address field in the app_info.json file.

  The file path of app_info.json:

  - Linux: /usr/local/logtail/app_info.json
  - Windows x64: C:\Program Files (x86)\Alibaba\Logtail\app_info.json
  - Windows x32: C:\Program Files\Alibaba\Logtail\app_info.json

3.3 Install

3.3.1 Install Logtail in Linux

  The Logtail client is a log collection agent provided by Log Service. This topic describes how to install the Logtail client on a Linux server.

  Supported systems

  The Logtail client for Linux supports the following x86-64 (64-bit) Linux systems:

  - Aliyun Linux
  - Ubuntu
  - Debian
  - CentOS
  - OpenSUSE
  - Red Hat
Prerequisites

1. One or more servers are available.
2. The network type for log collection is determined based on the type and region of the server. For more information, see Select a network type.

Figure 3-4: Select a network type

Precautions

- Logtail is installed in overwrite mode. If you have installed Logtail before, the installer will uninstall your current version of Logtail, delete the `/usr/local/iLogtail` directory, and reinstall Logtail. By default, Logtail is started after the installation and at startup.
- The `${your_region_name}` parameter is one of the installation parameters used for the installation of Docker and Kubernetes. Copy the value of the parameter from the region name table.
- If the installation fails, click here to open a ticket.

Select an installation method

Select one of the following installation methods according to the network type you selected.

- Install Logtail through the Alibaba Cloud internal network
• **Install Logtail through the Internet**

• **Install Logtail with Global Acceleration enabled**

Before running the installation command, replace `$\{your\_region\_name\}` with the actual region name. The following table lists the names of different regions. You can also copy and run the installation commands for the corresponding region and network type.

Table 3-3: Region names for Logtail installation

<table>
<thead>
<tr>
<th>Region</th>
<th>Region name</th>
<th>Region</th>
<th>Region name</th>
</tr>
</thead>
<tbody>
<tr>
<td>China (Hangzhou)</td>
<td>cn-hangzhou</td>
<td>Australia (Sydney)</td>
<td>ap-southeast-2</td>
</tr>
<tr>
<td>China (Shanghai)</td>
<td>cn-shanghai</td>
<td>Malaysia (Kuala Lumpur)</td>
<td>ap-southeast-3</td>
</tr>
<tr>
<td>China (Qingdao)</td>
<td>cn-qingdao</td>
<td>Indonesia (Jakarta)</td>
<td>ap-southeast-5</td>
</tr>
<tr>
<td>China (Beijing)</td>
<td>cn-beijing</td>
<td>India (Mumbai)</td>
<td>ap-south-1</td>
</tr>
<tr>
<td>China (Zhangjiakou)</td>
<td>cn-zhangjiakou</td>
<td>Japan (Tokyo)</td>
<td>ap-northeast-1</td>
</tr>
<tr>
<td>China (Hohhot)</td>
<td>cn-huhehaote</td>
<td>Germany (Frankfurt)</td>
<td>eu-central-1</td>
</tr>
<tr>
<td>China (Shenzhen)</td>
<td>cn-shenzhen</td>
<td>UAE (Dubai)</td>
<td>me-east-1</td>
</tr>
<tr>
<td>China (Chengdu)</td>
<td>cn-chengdu</td>
<td>UK (London)</td>
<td>eu-west-1</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>cn-hongkong</td>
<td></td>
<td></td>
</tr>
<tr>
<td>US (Silicon Valley)</td>
<td>us-west-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>US (Virginia)</td>
<td>us-east-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Singapore</td>
<td>ap-southeast-1</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Install Logtail through the Alibaba Cloud internal network

The Alibaba Cloud internal network is a shared gigabit network, which provides faster and more stable data transfer than the Internet and does not consume Internet bandwidth.

You can install Logtail through the Alibaba Cloud internal network when the following conditions are met:

• Alibaba Cloud ECS instances are deployed.

• The ECS instances and the Log Service project are located in the same region.
When running the installation command, you need to specify the region. You can use the auto parameter or manually specify the region.

- **Use the auto parameter**

  If you are not sure about the region of the ECS instance, you can use the auto parameter of the installer to install Logtail. The Logtail installer obtains the metadata from the server and automatically determines the region of the ECS instance.

  1. Download the Logtail installer through the Internet. This operation requires access to the Internet and consumes about 10 KB of Internet traffic.

     ```
     ```

  2. Use the auto parameter for installation. This operation does not consume Internet traffic. The installation program of the corresponding region will be automatically downloaded.

     ```
     ./logtail.sh install auto
     ```

- **Manually specify the region**

  You can also manually install Logtail. Downloading the Logtail installer through the internal network does not consume Internet traffic.

  1. Obtain the name of the region where the Log Service project is located.

     In the installation command, `${your_region_name}` indicates the name of the region where the Log Service project is located. Select the region name according to the region name table. For example, the name of the China (Hangzhou) region is `cn-hangzhou`.

  2. Run the installation command after replacing `${your_region_name}` with the actual region name.

     Replace `${your_region_name}` with the actual region name, and then run the installation command.

     ```
     wget http://logtail-release-${your_region_name}.oss-${your_region_name}-
The following table lists the installation commands for different regions. You can also install Logtail by running the command corresponding to the region where your Log Service project is located.

<table>
<thead>
<tr>
<th>Region</th>
<th>Installation command</th>
</tr>
</thead>
<tbody>
<tr>
<td>China (Hangzhou)</td>
<td><code>wget http://logtail-release-cn-hangzhou.oss-cn-hangzhou-internal.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install cn-hangzhou</code></td>
</tr>
<tr>
<td>China (Shanghai)</td>
<td><code>wget http://logtail-release-cn-shanghai.oss-cn-shanghai-internal.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install cn-shanghai</code></td>
</tr>
<tr>
<td>China (Qingdao)</td>
<td><code>wget http://logtail-release-cn-qingdao.oss-cn-qingdao-internal.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install cn-qingdao</code></td>
</tr>
<tr>
<td>China (Beijing)</td>
<td><code>wget http://logtail-release-cn-beijing.oss-cn-beijing-internal.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install cn-beijing</code></td>
</tr>
<tr>
<td>China (Zhangjiakou)</td>
<td><code>wget http://logtail-release-cn-zhangjiakou.oss-cn-zhangjiakou-internal.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install cn-zhangjiakou</code></td>
</tr>
<tr>
<td>China (Hohhot)</td>
<td><code>wget http://logtail-release-cn-huhehaote.oss-cn-huhehaote-internal.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install cn-huhehaote</code></td>
</tr>
<tr>
<td>China (Shenzhen)</td>
<td><code>wget http://logtail-release-cn-shenzhen.oss-cn-shenzhen-internal.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install cn-shenzhen</code></td>
</tr>
<tr>
<td>China (Chengdu)</td>
<td><code>wget http://logtail-release-cn-chengdu.oss-cn-chengdu-internal.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install cn-chengdu</code></td>
</tr>
<tr>
<td>Hong Kong</td>
<td><code>wget http://logtail-release-cn-hongkong.oss-cn-hongkong-internal.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install cn-hongkong</code></td>
</tr>
<tr>
<td>US (Silicon Valley)</td>
<td><code>wget http://logtail-release-us-west-1.oss-us-west-1-internal.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install us-west-1</code></td>
</tr>
<tr>
<td>US (Virginia)</td>
<td><code>wget http://logtail-release-us-east-1.oss-us-east-1-internal.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install us-east-1</code></td>
</tr>
<tr>
<td>Singapore</td>
<td><code>wget http://logtail-release-ap-southeast-1.oss-ap-southeast-1-internal.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install ap-southeast-1</code></td>
</tr>
<tr>
<td>Japan (Tokyo)</td>
<td><code>wget http://logtail-release-ap-northeast-1.oss-ap-northeast-1-internal.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install ap-northeast-1</code></td>
</tr>
<tr>
<td>India (Mumbai)</td>
<td><code>wget http://logtail-release-ap-south-1.oss-ap-south-1-internal.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install ap-south-1</code></td>
</tr>
<tr>
<td>Germany (Frankfurt)</td>
<td><code>wget http://logtail-release-eu-central-1.oss-eu-central-1-internal.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install eu-central-1</code></td>
</tr>
<tr>
<td>UAE (Dubai)</td>
<td><code>wget http://logtail-release-me-east-1.oss-me-east-1-internal.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install me-east-1</code></td>
</tr>
<tr>
<td>UK (London)</td>
<td><code>wget http://logtail-release-eu-west-1.oss-eu-west-1-internal.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install eu-west-1</code></td>
</tr>
</tbody>
</table>
Note:
Log Service cannot obtain the owner information about other types of servers. In this case, you must manually configure AliUids after installing Logtail. Otherwise, Logtail has abnormal heartbeats and cannot collect logs. For more information about AliUids, see Configure AliUids for ECS servers under other Alibaba Cloud accounts or on-premises IDCs.

1. Obtain the name of the region where the Log Service project is located.

   In the installation command, `{your_region_name}` indicates the name of the region where the Log Service project is located. Select the region name according to the region name table. For example, the name of the China (Hangzhou) region is `cn-hangzhou`.

2. Run the installation command after replacing `{your_region_name}` with the actual region name.

   Replace `{your_region_name}` with the actual region name, and then run the installation command.

   ```bash
   wget http://logtail-release-{your_region_name}.oss-{your_region_name}.aliyuncs.com/
   ```
The following table lists the installation commands for different regions. You can also install Logtail by running the command corresponding to the region where your Log Service project is located.

<table>
<thead>
<tr>
<th>Region</th>
<th>Installation command</th>
</tr>
</thead>
<tbody>
<tr>
<td>China (Hangzhou)</td>
<td><code>wget http://logtail-release-cn-hangzhou.oss-cn-hangzhou.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install cn-hangzhou-internet</code></td>
</tr>
<tr>
<td>China (Shanghai)</td>
<td><code>wget http://logtail-release-cn-shanghai.oss-cn-shanghai.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install cn-shanghai-internet</code></td>
</tr>
<tr>
<td>China (Qingdao)</td>
<td><code>wget http://logtail-release-cn-qingdao.oss-cn-qingdao.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install cn-qingdao-internet</code></td>
</tr>
<tr>
<td>China (Beijing)</td>
<td><code>wget http://logtail-release-cn-beijing.oss-cn-beijing.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install cn-beijing-internet</code></td>
</tr>
<tr>
<td>China (Hohhot)</td>
<td><code>wget http://logtail-release-cn-huhehaote.oss-cn-huhehaote.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install cn-huhehaote-internet</code></td>
</tr>
<tr>
<td>China (Shenzhen)</td>
<td><code>wget http://logtail-release-cn-shenzhen.oss-cn-shenzhen.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install cn-shenzhen-internet</code></td>
</tr>
<tr>
<td>China (Chengdu)</td>
<td><code>wget http://logtail-release-cn-chengdu.oss-cn-chengdu.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install cn-chengdu-internet</code></td>
</tr>
<tr>
<td>US (Silicon Valley)</td>
<td><code>wget http://logtail-release-us-west-1.oss-us-west-1.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install us-west-1-internet</code></td>
</tr>
<tr>
<td>US (Virginia)</td>
<td><code>wget http://logtail-release-us-east-1.oss-us-east-1.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install us-east-1-internet</code></td>
</tr>
<tr>
<td>UAE (Dubai)</td>
<td><code>wget http://logtail-release-me-east-1.oss-me-east-1.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install me-east-1-internet</code></td>
</tr>
<tr>
<td>India (Mumbai)</td>
<td><code>wget http://logtail-release-ap-south-1.oss-ap-south-1.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install ap-south-1-internet</code></td>
</tr>
</tbody>
</table>
transmit data may cause problems such as high latency and unstable transmission. In this case, you can enable *Global Acceleration*. *Global Acceleration* accelerates log collection by using the edge nodes of Alibaba Cloud CDN. Compared with data transmission through the Internet, Global Acceleration offers a more stable network with minimal transmission latency.

1. Obtain the name of the region where the Log Service project is located.

   *In the installation command, ${your_region_name} indicates the name of the region where the Log Service project is located. Select the region name according to the *region name table*. For example, the name of the China (Hangzhou) region is `cn-hangzhou`.*

2. Run the installation command after replacing ${your_region_name} with the actual region name.

   *Replace ${your_region_name} with the actual region name, and then run the installation command.*

   ```
   wget http://logtail-release-${your_region_name}.oss-${your_region_name}.aliyuncs.com/
   ```
The following table lists the installation commands for different regions. You can also install Logtail by running the command corresponding to the region where your Log Service project is located.

<table>
<thead>
<tr>
<th>Region</th>
<th>Command</th>
</tr>
</thead>
<tbody>
<tr>
<td>China (Beijing)</td>
<td><code>wget http://logtail-release-cn-beijing.oss-cn-beijing.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install cn-beijing-acceleration</code></td>
</tr>
<tr>
<td>China (Qingdao)</td>
<td><code>wget http://logtail-release-cn-qingdao.oss-cn-qingdao.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install cn-qingdao-acceleration</code></td>
</tr>
<tr>
<td>China (Hangzhou)</td>
<td><code>wget http://logtail-release-cn-hangzhou.oss-cn-hangzhou.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install cn-hangzhou-acceleration</code></td>
</tr>
<tr>
<td>China (Shanghai)</td>
<td><code>wget http://logtail-release-cn-shanghai.oss-cn-shanghai.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install cn-shanghai-acceleration</code></td>
</tr>
<tr>
<td>China (Shenzhen)</td>
<td><code>wget http://logtail-release-cn-shenzhen.oss-cn-shenzhen.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install cn-shenzhen-acceleration</code></td>
</tr>
<tr>
<td>China (Zhangjiakou)</td>
<td><code>wget http://logtail-release-cn-zhangjiakou.oss-cn-zhangjiakou.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install cn-zhangjiakou-acceleration</code></td>
</tr>
<tr>
<td>China (Hohhot)</td>
<td><code>wget http://logtail-release-cn-huhehaote.oss-cn-huhehaote.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install cn-huhehaote-acceleration</code></td>
</tr>
<tr>
<td>China (Chengdu)</td>
<td><code>wget http://logtail-release-cn-chengdu.oss-cn-chengdu.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install cn-chengdu-acceleration</code></td>
</tr>
<tr>
<td>Hong Kong</td>
<td><code>wget http://logtail-release-cn-hongkong.oss-cn-hongkong.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install cn-hongkong-acceleration</code></td>
</tr>
<tr>
<td>US (Silicon Valley)</td>
<td><code>wget http://logtail-release-us-west-1.oss-us-west-1.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install us-west-1-acceleration</code></td>
</tr>
<tr>
<td>US (Virginia)</td>
<td><code>wget http://logtail-release-us-east-1.oss-us-east-1.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install us-east-1-acceleration</code></td>
</tr>
<tr>
<td>Indonesia (Jakarta)</td>
<td><code>wget http://logtail-release-ap-southeast-5.oss-ap-southeast-5.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install ap-southeast-5-acceleration</code></td>
</tr>
<tr>
<td>Japan (Tokyo)</td>
<td><code>wget http://logtail-release-ap-northeast-1.oss-ap-northeast-1.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install ap-northeast-1-acceleration</code></td>
</tr>
<tr>
<td>Germany (Frankfurt)</td>
<td><code>wget http://logtail-release-eu-central-1.oss-eu-central-1.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install eu-central-1-acceleration</code></td>
</tr>
<tr>
<td>UAE (Dubai)</td>
<td><code>wget http://logtail-release-me-east-1.oss-me-east-1.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install me-east-1-acceleration</code></td>
</tr>
<tr>
<td>India (Mumbai)</td>
<td><code>wget http://logtail-release-ap-south-1.oss-ap-south-1.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install ap-south-1-acceleration</code></td>
</tr>
<tr>
<td>UK (London)</td>
<td><code>wget http://logtail-release-eu-west-1.oss-eu-west-1.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install eu-west-1-acceleration</code></td>
</tr>
</tbody>
</table>
Upgrade Logtail

You can use the Logtail installer (logtail.sh) to upgrade Logtail. The installer automatically selects an appropriate upgrade method based on the configuration information of the installed Logtail.

**Note:**
During the upgrade, Logtail will be temporarily stopped. Only necessary files are overwritten. The configuration file, checkpoint file, and logs are retained.

Run the following commands to upgrade Logtail:

```
# Download the Logtail installer.
# Upgrade Logtail.
sudo ./logtail.sh upgrade
```

**Response:**

```
# The upgrade is successful.
Stop logtail successfully.
ilogtail is running
Uptrain logtail success
{}
```

```
# The upgrade fails because the current version is the latest version.
```
[Error]: Already up to date.

Manually start or stop Logtail

- **Start Logtail**
  
  Run the following command as an administrator to start Logtail:

  ```bash
  /etc/init.d/ilogtaild start
  ```

- **Stop Logtail**
  
  Run the following command as an administrator to stop Logtail:

  ```bash
  /etc/init.d/ilogtaild stop
  ```

Uninstall Logtail

Download the Logtail installer logtail.sh, and then run the following commands to uninstall Logtail:

```bash
chmod 755 logtail.sh; ./logtail.sh uninstall
```

### 3.3.2 Install Logtail in Windows

The Logtail client is a log collection agent provided by Log Service. This topic describes how to install the Logtail client on a Windows server.

**Supported systems**

The Logtail client for Windows supports the following operating systems:

- Windows 7 (Client) 32-bit
- Windows 7 (Client) 64-bit
- Windows Server 2008 32-bit
- Windows Server 2008 64-bit
- Windows Server 2012 64-bit
- Windows Server 2016 64-bit

**Prerequisites**

1. One or more servers are available.
2. The network type for log collection is determined based on the type and region of the server. For more information, see Select a network type.

Figure 3-5: Select a network type

Install Logtail

1. Download the installation package.

   Download links:
   - If you are in Mainland China, click here.
   - If you are outside Mainland China, click here.

2. Decompress the logtail_installer.zip package to the current directory.
3. Select a network type based on the type and region of the server, and then install Logtail based on the region of the Log Service project.

Run PowerShell or CMD as an administrator to go to the logtail_installer directory where you decompress the Logtail installation package. Then, run the installation command based on the region and network type.

The following table lists the installation commands for different network types in different regions.

<table>
<thead>
<tr>
<th>Region</th>
<th>Alibaba Cloud internal network (classic network or VPC)</th>
<th>Internet</th>
<th>Global Acceleration</th>
</tr>
</thead>
<tbody>
<tr>
<td>China (Qingdao)</td>
<td>.\logtail_installer.exe install cn-qingdao</td>
<td>.\logtail_installer.exe install cn-qingdao-internet</td>
<td>.\logtail_installer.exe install cn-qingdao-acceleration</td>
</tr>
<tr>
<td>China (Beijing)</td>
<td>.\logtail_installer.exe install cn-beijing</td>
<td>.\logtail_installer.exe install cn-beijing-internet</td>
<td>.\logtail_installer.exe install cn-beijing-acceleration</td>
</tr>
<tr>
<td>China (Zhangjiakou)</td>
<td>.\logtail_installer.exe install cn-zhangjiakou</td>
<td>.\logtail_installer.exe install cn-zhangjiakou-internet</td>
<td>.\logtail_installer.exe install cn-zhangjiakou-acceleration</td>
</tr>
<tr>
<td>China (Hohhot)</td>
<td>.\logtail_installer.exe install cn-huhehaote</td>
<td>.\logtail_installer.exe install cn-huhehaote-internet</td>
<td>.\logtail_installer.exe install cn-huhehaote-acceleration</td>
</tr>
<tr>
<td>China (Hangzhou)</td>
<td>.\logtail_installer.exe install cn-hangzhou</td>
<td>.\logtail_installer.exe install cn-hangzhou-internet</td>
<td>.\logtail_installer.exe install cn-hangzhou-acceleration</td>
</tr>
<tr>
<td>Region</td>
<td>Alibaba Cloud internal network (classic network or VPC)</td>
<td>Internet</td>
<td>Global Acceleration</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------------------------------------------------</td>
<td>-----------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>China (Shanghai)</td>
<td>.\logtail_in installer.exe install cn-shanghai</td>
<td>.\logtail_in installer.exe install cn-shanghai-internet</td>
<td>.\logtail_in installer.exe install cn-shanghai-acceleration</td>
</tr>
<tr>
<td>China (Shenzhen)</td>
<td>.\logtail_in installer.exe install cn-shenzhen</td>
<td>.\logtail_in installer.exe install cn-shenzhen-internet</td>
<td>.\logtail_in installer.exe install cn-shenzhen-acceleration</td>
</tr>
<tr>
<td>China (Chengdu)</td>
<td>.\logtail_in installer.exe install cn-chengdu</td>
<td>.\logtail_in installer.exe install cn-chengdu-internet</td>
<td>.\logtail_in installer.exe install cn-chengdu-acceleration</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>.\logtail_in installer.exe install cn-hongkong</td>
<td>.\logtail_in installer.exe install cn-hongkong-internet</td>
<td>.\logtail_in installer.exe install cn-hongkong-acceleration</td>
</tr>
<tr>
<td>US (Silicon Valley)</td>
<td>.\logtail_in installer.exe install us-west-1</td>
<td>.\logtail_in installer.exe install us-west-1-internet</td>
<td>.\logtail_in installer.exe install us-west-1-acceleration</td>
</tr>
<tr>
<td>US (Virginia)</td>
<td>.\logtail_in installer.exe install us-east-1</td>
<td>.\logtail_in installer.exe install us-east-1-internet</td>
<td>.\logtail_in installer.exe install us-east-1-acceleration</td>
</tr>
<tr>
<td>Singapore</td>
<td>.\logtail_in installer.exe install ap-southeast-1</td>
<td>.\logtail_in installer.exe install ap-southeast-1-internet</td>
<td>.\logtail_in installer.exe install ap-southeast-1-acceleration</td>
</tr>
<tr>
<td>Region</td>
<td>Alibaba Cloud internal network (classic network or VPC)</td>
<td>Internet</td>
<td>Global Acceleration</td>
</tr>
<tr>
<td>-------------------------</td>
<td>--------------------------------------------------------</td>
<td>----------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Australia (Sydney)</td>
<td>.\logtail_inсталler.exe install ap-southeast-2</td>
<td>.\logtail_inсталler.exe install ap-southeast-2-internet</td>
<td>.\logtail_inсталler.exe install ap-southeast-2-acceleration</td>
</tr>
<tr>
<td>Malaysia (Kuala Lumpur)</td>
<td>.\logtail_inсталler.exe install ap-southeast-3</td>
<td>.\logtail_inсталler.exe install ap-southeast-3-internet</td>
<td>.\logtail_inсталler.exe install ap-southeast-3-acceleration</td>
</tr>
<tr>
<td>Indonesia (Jakarta)</td>
<td>.\logtail_inсталler.exe install ap-southeast-5</td>
<td>.\logtail_inсталler.exe install ap-southeast-5-internet</td>
<td>.\logtail_inсталler.exe install ap-southeast-5-acceleration</td>
</tr>
<tr>
<td>India (Mumbai)</td>
<td>.\logtail_inсталler.exe install ap-south-1</td>
<td>.\logtail_inсталler.exe install ap-south-1-internet</td>
<td>.\logtail_inсталler.exe install ap-south-1-acceleration</td>
</tr>
<tr>
<td>Japan (Tokyo)</td>
<td>.\logtail_inсталler.exe install ap-northeast-1</td>
<td>.\logtail_inсталler.exe install ap-northeast-1-internet</td>
<td>.\logtail_inсталler.exe install ap-northeast-1-acceleration</td>
</tr>
<tr>
<td>Germany (Frankfurt)</td>
<td>.\logtail_inсталler.exe install eu-central-1</td>
<td>.\logtail_inсталler.exe install eu-central-1-internet</td>
<td>.\logtail_inсталler.exe install eu-central-1-acceleration</td>
</tr>
<tr>
<td>UAE (Dubai)</td>
<td>.\logtail_inсталler.exe install me-east-1</td>
<td>.\logtail_inсталler.exe install me-east-1-internet</td>
<td>.\logtail_inсталler.exe install me-east-1-acceleration</td>
</tr>
</tbody>
</table>
### Logtail Collection

<table>
<thead>
<tr>
<th>Region</th>
<th>Alibaba Cloud internal network (classic network or VPC)</th>
<th>Internet</th>
<th>Global Acceleration</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK (London)</td>
<td>.\logtail_installer.exe install eu-west-1</td>
<td>.\logtail_installer.exe install eu-west-1-internet</td>
<td>.\logtail_installer.exe install eu-west-1-acceleration</td>
</tr>
</tbody>
</table>

#### Note:
If you use Logtail on a server deployed in an on-premises IDC or provided by another cloud service vendor, Log Service cannot obtain the owner information about ECS instances under other Alibaba Cloud accounts or other types of servers. In this case, you must manually configure AliUids after installing Logtail. Otherwise, Logtail has abnormal heartbeats and cannot collect logs. For more information, see [Configure AliUids for ECS servers under other Alibaba Cloud accounts or on-premises IDCs](#).

Go to the installation path

After you run the installation command, Logtail is installed in the specified path, which cannot be changed. In this path, you can view the Logtail version in the `app_info.json` file or uninstall Logtail.

The installation path is as follows:

- **32-bit Windows**: `C:\Program Files\Alibaba\Logtail`
- **64-bit Windows**: `C:\Program Files (x86)\Alibaba\Logtail`

#### Note:
You can run a 32-bit or 64-bit application in 64-bit Windows. However, the operating system stores 32-bit applications in an `x86` folder to ensure compatibility.

Logtail for Windows is a 32-bit application. Therefore, it is installed in the `Program Files (x86)` folder in 64-bit Windows. If Logtail for 64-bit Windows becomes available in the future, it will be automatically installed in the `Program Files` folder.
View the Logtail version

Logtail is automatically installed in the default directory. To view the Logtail version, you can go to the directory and use Notepad or another text editor to open the app_info.json file. The logtail_version field indicates the version of the installed Logtail.

In the following example, the Logtail version is 1.0.0.0:

```
{
    "logtail_version": "1.0.0.0"
}
```

Upgrade Logtail

- Automatic upgrade
  In normal cases, Logtail for Windows is automatically upgraded. However, you must manually upgrade Logtail earlier than 1.0.0.0 to Logtail 1.0.0.0 or later.

- Manual upgrade
  You must manually upgrade Logtail earlier than 1.0.0.0 to Logtail 1.0.0.0 or later. The procedure for manually upgrading Logtail is the same as that for installing Logtail. You only need to download and decompress the latest installation package and install Logtail by following the steps.

  **Note:**
  During manual upgrade, Logtail is automatically uninstalled and then reinstalled. In this case, files in the original installation directory are deleted. If necessary, we recommend that you back up the files before manually upgrading Logtail.

Manually start or stop Logtail

In the Control Panel, choose System and Security > Administrative Tools, and then double-click Services.

Find the target service based on your Logtail version.

- Logtail 0.x.x.x: LogtailWorker.
- Logtail 1.0.0.0 and later: LogtailDaemon.

Perform the following operations as required:

- Manually start Logtail: Right-click Logtail and select Start.
Stop Logtail: Right-click Logtail and select Stop.
Restart Logtail: Right-click Logtail and select Restart.

Uninstall Logtail

Run PowerShell or CMD as an administrator to go to the logtail_installer directory where you decompress the Logtail installation package. Then, run the following command to uninstall Logtail:

```bash
./logtail_installer.exe uninstall
```

After Logtail is uninstalled, the installation directory of Logtail will be deleted. However, some residual configuration information is kept in the C:\LogtailData directory. You can manually delete the information as needed. The residual configuration information includes:

- `checkpoint`: contains checkpoint information of all plug-ins, for example, the Windows event log plug-in.
- `logtail_check_point`: contains major checkpoint information of Logtail.

### 3.3.3 Set startup parameters

This topic describes how to set the Logtail startup parameters. You can refer to this topic for parameter setting as needed.

Scenarios

In the following scenarios, you need to set the Logtail startup parameters:

- A large number of log files are to be collected. They may occupy a large amount of memory. The metadata of each file, such as the file signature, collection location, and file name, needs to be maintained in memory.
- A high volume of log data leads to a high CPU usage.
- A high volume of log data leads to heavy traffic sent to Log Service.

Startup configurations

- **File path:**

  `/usr/local/ilogtail/ilogtail_config.json`

- **File format:**

  JSON
File sample (only partial configuration items are shown):

```json
{
  ...
  "cpu_usage_limit" : 0.4,
  "mem_usage_limit" : 100,
  "max_bytes_per_sec" : 2097152,
  "process_thread_count" : 1,
  "send_request_concurrency" : 4,
  "buffer_file_num" : 25,
  "buffer_file_size" : 20971520,
  "buffer_file_path" : "",
  ...
}
```

Common configuration parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>cpu_usage_limit</td>
<td>The CPU usage threshold, which is calculated by core.</td>
<td>The value is of the Double type. Valid values: [0.1, the number of CPU cores of the current machine]. Default value: 2. For example, the value 0.4 indicates that the CPU usage of Logtail is limited to 40% of single-core CPUs. Logtail restarts automatically when the threshold is exceeded.</td>
</tr>
<tr>
<td>Parameter</td>
<td>Description</td>
<td>Value</td>
</tr>
<tr>
<td>--------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>mem_usage_limit</td>
<td>The usage threshold of resident memory.</td>
<td>The value is of the Int type. Unit: MB. Valid values: [128, the valid memory value of the current machine]. Default value: 2048. For example, the value 100 indicates that the memory usage of Logtail is limited to 100 MB. Logtail restarts automatically when the threshold is exceeded.</td>
</tr>
<tr>
<td>max_bytes_per_sec</td>
<td>The traffic limit on the raw data sent by Logtail. Traffic exceeding 20 MB/s is not throttled.</td>
<td>The value is of the Int type. Unit: Byte/s. Valid values: [1024, 52428800]. Default value: 20971520. For example, the value 2097152 indicates that the data transfer rate of Logtail is limited to 2 MB/s.</td>
</tr>
<tr>
<td>process_thread_count</td>
<td>The number of threads with which Logtail processes written data of log files.</td>
<td>The value is of the Int type. Valid values: [1, 64]. Default value: 1.</td>
</tr>
</tbody>
</table>

Generally, Logtail supports a write speed of 24 Mbit/s in simple mode and 12 Mbit/s in full mode. By default, there is no need to modify this value, but you can increase the threshold value when necessary.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>send_request_concurrency</td>
<td>The asynchronous concurrency. By default, Logtail sends data packets asynchronously. You can set a larger asynchronous concurrency value if the write TPS is large. A single concurrency occupies 0.5 to 1 Mbit/s network throughput, depending on the network delay.</td>
<td>The value is of the Int type. Valid values: [1, 1000]. Default value: 20.</td>
</tr>
<tr>
<td>buffer_file_num</td>
<td>The maximum number of cached files. When a network exception occurs and the writing quota is exceeded, Logtail writes the logs that are parsed in real time to local files in the installation directory, and then tries to resend the logs to Log Service after the recovery.</td>
<td>The value is of the Int type. Valid values: [1, 100]. Default value: 25.</td>
</tr>
<tr>
<td>buffer_file_size</td>
<td>The maximum number of bytes that each cached file allows. The product of the value of buffer_file_num and that of buffer_file_size indicates the maximum disk space available for cached files.</td>
<td>The value is of the Int type. Unit: Byte. Valid values: [1048576, 104857600]. Default value: 20971520, which is 20 MB.</td>
</tr>
<tr>
<td>buffer_file_path</td>
<td>The directory that stores cached files. After modifying this parameter, you need to move the files named in the format logtail_buffer_file_* in the old cache directory to the new directory so that Logtail can read the cached files and delete them after sending logs.</td>
<td>By default, the value is an empty string. In this case, the cached files are stored in the Logtail installation directory /usr/local/ilogtail.</td>
</tr>
<tr>
<td>Parameter</td>
<td>Description</td>
<td>Value</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>bind_interface</td>
<td>The name of the NIC bound to the local machine. For example, eth1. This parameter is valid only for Logtail for Linux.</td>
<td>By default, the value is an empty string. The available NIC is bound automatically. If this parameter is set, Logtail uses the specified NIC to upload logs.</td>
</tr>
<tr>
<td>check_point_filename</td>
<td>The full path for storing the checkpoint file of Logtail. We recommend that Docker users modify this file storage path and mount the directory where the checkpoint file resides to the host. Otherwise, duplicate collection occurs when the container is released due to checkpoint information loss. For example, set check_point_filename to /data/logtail/check_point.dat in Docker, and add -v /data/docker1/logtail:/data/logtail to the Docker startup command to mount the /data/docker1/logtail directory of the host to the /data/logtail directory of Docker.</td>
<td>Default value: /tmp/logtail_check_point.</td>
</tr>
<tr>
<td>Parameter</td>
<td>Description</td>
<td>Value</td>
</tr>
<tr>
<td>-------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>user_config_file_path</td>
<td>The full path for storing the collection configuration file of Logtail. We recommend that Docker users modify this file storage path and mount the directory where the collection configuration file resides to the host. Otherwise, duplicate collection occurs when the container is released due to checkpoint information loss. For example, set <code>user_config_file_path</code> to <code>/data/logtail/user_log_config.json</code> in Docker, and add <code>-v /data/docker1/logtail:/data/logtail</code> to the Docker startup command to mount the /data/docker1/logtail directory of the host to the /data/logtail directory of Docker.</td>
<td>By default, the <code>user_log_config.json</code> file is stored in the directory where the binary process is located.</td>
</tr>
<tr>
<td>discard_old_data</td>
<td>Specifies whether to discard historical logs. A value of true indicates that logs generated more than 12 hours ago will be discarded.</td>
<td>The value is of the Boolean type. Default value: true.</td>
</tr>
<tr>
<td>working_ip</td>
<td>The local IP address reported by Logtail. If the value is an empty string, Logtail automatically obtains the IP address of the local machine.</td>
<td>The value is an IP address. By default, the value is an empty string.</td>
</tr>
<tr>
<td>working_hostname</td>
<td>The local hostname reported by Logtail. If the value is an empty string, Logtail automatically obtains the hostname of the local machine.</td>
<td>The value is of the String type. By default, the value is an empty string.</td>
</tr>
<tr>
<td>Parameter</td>
<td>Description</td>
<td>Value</td>
</tr>
<tr>
<td>-------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>max_read_buffer_size</td>
<td>The maximum size of a log, in Bytes. If the size of a single log exceeds 512 KB, you can adjust the parameter value.</td>
<td>The value is of the Long type. Default value: 524288, which is 512 KB.</td>
</tr>
<tr>
<td>oas_connect_timeout</td>
<td>The connection timeout period when Logtail sends a request, for example, to obtain the configuration or AccessKey. This parameter applies to scenarios where the connection takes a long period of time due to poor network conditions.</td>
<td>The value is of the Long type. Unit: second. Default value: 5.</td>
</tr>
<tr>
<td>oas_request_timeout</td>
<td>The total timeout period when Logtail sends a request, for example, to obtain the configuration or AccessKey. This parameter applies to scenarios where the connection takes a long period of time due to poor network conditions.</td>
<td>The value is of the Long type. Unit: second. Default value: 10.</td>
</tr>
</tbody>
</table>

**Note:**

- The preceding table only lists the common startup parameters. If the ilogtail_config.json file contains parameters that are not listed in the table, use the default settings.
- Add parameters or modify the values of existing parameters as needed. Do not add unnecessary parameters to the ilogtail_config.json file.

**Modify configurations**

1. **Modify the ilogtail_config.json file as needed.**

   Ensure that the modified configurations are in the valid JSON format.

2. Restart Logtail for the modified configurations to take effect.

   ```
   /etc/init.d/ilogtaild stop
   /etc/init.d/ilogtaild start
   ```
3.4 Machine Group

3.4.1 Overview

Log Service uses machine groups to manage all the servers whose logs are collected by Logtail clients.

A machine group is a virtual group that contains multiple servers. If you want the logs of multiple servers to be collected by Logtail clients with the same configuration, you can add the servers to a machine group and apply the Logtail configuration to the machine group.

You can define a machine group by using either of the following identification types:

- **IP address**: Add the IP addresses of all the servers to the machine group. Each server in the group can be identified by using its unique IP address.
- **Custom ID**: Customize an ID for the machine group and use this same custom ID for each server of the machine group.

**Note:**

- Before adding a server of other cloud vendors or your local IDC, or adding an ECS instance of other accounts to a machine group, you must set an AliUid for the server or instance. For more information, see [Configure AliUids for ECS servers under other Alibaba Cloud accounts or on-premises IDCs](#).
- You cannot add Windows servers and Linux servers to the same machine group.

**IP address-based machine group**

You can add multiple servers to a machine group by adding their IP addresses to the machine group. Then you can configure the Logtail clients on all the servers at the same time.

- If you use ECS servers that are not bound to hostnames, and the network types of these ECS servers remain unchanged, you can use their private IP addresses to define the machine group.
In other cases, use the server IP address obtained automatically by the Logtail client when you define a machine group. The IP address of each server is recorded in the IP address field of the `app_info.json` server file on the server.

**Note:**

`app_info.json` is the file that records the internal information of the Logtail client. The internal information includes the server IP addresses obtained by the Logtail client automatically. Manually modifying the IP address field of this file does not change the IP addresses obtained by the Logtail client.

A Logtail client automatically obtains a server IP address by using the following methods:

- If the IP address of a server has been bound with its host name in the `/etc/hosts` server file, the Logtail client automatically obtains the IP address.
- If the IP address of a server has not been bound with its host name, the Logtail automatically obtains the IP address of the first Network Interface (NI) on the server.

**Note:**

Whether the Alibaba Cloud intranet is used for data collection does not depend on whether you use a private IP address to define a machine group. Your server log data can be collected to Log Service through the Alibaba Cloud intranet only when you use an ECS instance of Alibaba Cloud and you have selected Alibaba Cloud intranet (Classic Network and VPC) when installing a Logtail on the instance.

For more information, see [Create an IP address-based server group](#).

**Custom ID-based machine group**

In addition to IP addresses, custom IDs can also be used to define machine groups. We recommend that you use a machine group defined by a custom ID in the following scenarios:

- In a custom network, for example a VPC, different servers may have the same IP address. In that case, Log Service cannot manage the Logtail clients on the servers. Using a custom ID to define a machine group can eliminate such a problem.
Multiple servers in a machine group can use one custom ID to implement machine group auto scaling. If you set the same custom ID for a new server, the Log Service identifies the new server automatically and adds it to the machine group.

Typically, the system consists of multiple modules. Each module can be expanded horizontally. That is, multiple servers can be added to each module. By creating a machine group separately for each module, you can collect logs by module. Therefore, you need to create a custom ID for each module, and set the machine group ID for the servers of each module. For example, a common website consists of an HTTP request processing module, a cache module, a logic processing module, and a storage module. The custom IDs can be set as http_module for the HTTP request processing module, cache_module for the cache module, logic_module for the logic processing module, and store_module for the storage module.

For more information, see #unique_44.

3.4.2 Create an IP address-based server group

Log Service allows you to create IP address-based server groups. After adding the IP addresses of servers retrieved by Logtail to an IP address-based server group, you can use the same Logtail configuration to collect logs from these servers.

Prerequisites

- A project and a Logstore are created.
- At least one server is available. If you are using an Alibaba Cloud ECS instance, make sure that the ECS instance and the current project in Log Service are in the same Alibaba Cloud region.
- Logtail is installed on the server. For more information, see Install Logtail in Linux and Install Logtail in Windows.
- For servers of other cloud service providers, servers in user-built IDCs, and ECS instances under other Alibaba Cloud accounts, make sure that AliUids have been configured for them. For more information, see Configure AliUids for ECS servers under other Alibaba Cloud accounts or on-premises IDCs.

Context

Whether the Alibaba Cloud internal network is used for data collection does not depend on whether the IP addresses configured in a server group are private IP...
addresses. The Alibaba Cloud internal network can be used to collect log data to Log Service only when your server is an Alibaba Cloud ECS instance that resides in the same Alibaba Cloud region as the current Log Service project, and you have selected Alibaba Cloud Intranet (Classic Network, VPC) when installing Logtail.

Procedure

1. Check the IP addresses of servers that are automatically retrieved by Logtail.
   The IP addresses automatically retrieved by Logtail are recorded in the IP address field of the `app_info.json` file.

   On the server where Logtail is installed, check the `app_info.json` file. The file path is as follows:
   
   - Linux: `/usr/local/ilogtail/app_info.json`
   - 64-bit Windows: `C:\Program Files (x86)\Alibaba\Logtail\app_info.json`
   - 32-bit Windows: `C:\Program Files\Alibaba\Logtail\app_info.json`

   The following figure shows how to check the retrieved IP addresses in a Linux system.

   ![Linux example]

2. Log on to the Log Service console, and then click the target project name.

3. In the left-side navigation pane, click the Server Groups icon to expand the list of server groups.

4. Click the icon next to Server Groups and choose Create Server Group from the shortcut menu.

   You can also create a server group by using the data import wizard.
5. Create a server group.
   
a) Set name.
   
   The name of the server group must be 3 to 128 characters in length and can contain only lowercase letters, digits, hyphens (-), and underscores (_). It must start and end with a lowercase letter or digit.

   **Note:**

   After the server group is created, you cannot modify its name. Proceed with caution.

b) Set Identifier to IP Addresses.

c) Configure IP Addresses.

   Enter the server IP addresses obtained in 1.

   **Note:**

   - Make sure that the server IP addresses have been obtained based on 1.
   - If the server group contains multiple servers, you must separate their IP addresses with line breaks.
Do not add Windows and Linux servers to the same server group.

   For more information about server group topics, see Set a log topic.

7. Click OK.

Result
You can view the created server group in the list of server groups.
3.4.4 Configure AliUids for ECS servers under other Alibaba Cloud accounts or on-premises IDCs

If Logtail is installed on ECS servers under other Alibaba Cloud accounts, provided by other cloud vendors, or located in on-premises IDCs, you must configure AliUids for the servers so that they can be added into machine groups for log collection.

Context

If the target server for log collection through Logtail is purchased by another Alibaba Cloud account or provided by another cloud vendor, you need to install Logtail on the server and configure an AliUid for it. By doing so, you grant your Alibaba Cloud account the permissions to access and collect logs from the server. Otherwise, the server does not receive heartbeat information and cannot collect logs.

Prerequisites

- The target server for log collection is under another Alibaba Cloud account, provided by another cloud vendor, or located in an on-premises IDC.
- The Logtail client is installed on the server.

For more information, see Install Logtail in Linux and Install Logtail in Windows as needed.

Procedure
1. View the Alibaba Cloud account ID, namely, the AliUid.
   
   a. Log on to the Log Service console.
   
   b. In the upper-right corner, click the icon. For more information, see Cloud Shell.
   
   c. Run the `echo $ALIBABA_CLOUD_ACCOUNT_ID` command to obtain your AliUid.

![Image showing command output]

You can view the AliUid of the account to which the Log Service project belongs on the Account Management page.

Figure 3-6: View your Alibaba Cloud account ID

![Image showing command output]
2. Configure an AliUid for the server.

   - In Linux:

     Create a file named after the AliUid in the `/etc/logtail/users` directory. If the directory does not exist, you need to create one. You can configure multiple AliUids for a single server by running a command similar to the following:

     ```
touch /etc/logtail/users/1**************
touch /etc/logtail/users/1**************
```

     If you do not need Logtail to collect data to your Log Service project, you can delete the AliUid:

     ```
rm /etc/logtail/users/1**************
```

   - In Windows:

     Create a file named after the AliUid in the `C:\LogtailData\users` directory.

     If you want to delete the AliUid, you can simply delete this file. (`C:\LogtailData\users\1**************` is used as an example file.)

     **Note:**

     - After an AliUid is configured for a server, the Alibaba Cloud account has the permission to collect logs from the server by using Logtail. You need to delete unnecessary AliUid files from the server in a timely manner.
     - Addition and deletion of an AliUid take effect within 1 minute.

3.4.5 Manage collection configurations

   The Logtail client can help Log Service users easily collect logs from ECS instances by using the Log Service console. After installing the Logtail client, you must create a log collection configuration for the Logtail client. You can create and modify Logtail configurations for Logstores in the Logstore list.

   Create a Logtail configuration

   For more information about how to create a Logtail configuration in the Log Service console, see *Configuration procedure for text log collection* and #unique_47.

   View the list of Logtail configurations

   1. Log on to the Log Service console.
2. Find the target project in the project list and click the project name.
3. On the Logstores page, click the angle bracket icon next to the target Logstore, and choose Data Import > Logtail Configurations. Each item indicates a different Logtail configuration.

Figure 3-7: Logtail configuration list

Note:
A log file can be collected by only one configuration.

Modify Logtail configurations

1. Log on to the Log Service console.
2. Find the target project in the project list and click the project name.
3. On the Logstores page, click the angle bracket icon before the target Logstore, and choose Data Import > Logtail Configurations.
4. Click the name of the Logtail configuration to be modified. On the Configure Logtail page that appears, click Modify.

You can change the log collection mode and specify the server group to which the log configuration is applied. The process of modifying a Logtail configuration is the same as the process of creating a Logtail configuration.

Delete a Logtail configuration

1. Log on to the Log Service console.
2. Find the target project in the project list and click the project name.
3. On the Logstores page, click the angle bracket before the target Logstore, and choose Data Import > Logtail Configurations.
4. Find the Logtail configuration to be deleted, click the icon next to it, and then select Delete.

After the configuration is deleted, it is unbound from the server group to which it is applied, and Logtail no longer collects logs specified by the configuration.

Note:
Before deleting a Logstore, you must delete all its Logtail configurations.

3.4.6 Manage server groups

Log Service uses server groups to manage all ECS instances whose logs need to be collected by Logtail. Log Service allows you to create, modify, and delete server groups, view the list and status of server groups, manage configurations, and apply server group identifiers.

Create a server group

- **IP**: Define the name of a server group and add internal IP addresses of the servers in the server group. For more information, see Create an IP address-based server group.

- **Identifier**: Customize an identifier for a server group and use the same identifier for servers in the server group. For more information, see #unique_44.

View the server group list

1. Log on to the Log Service console.

2. Find the target project in the project list and click the project name.
3. In the left-side navigation pane on the page that appears, click the Server Groups icon to display the list of server groups.

You can view all server groups under the project.

Figure 3-8: View the server group list

Modify a server group

After creating a server group, you can adjust the list of servers in the server group as needed.

Note:
The name of the server group cannot be changed after the server group is created.

1. Log on to the Log Service console.
2. Find the target project in the project list and click the project name.
3. In the left-side navigation pane on the page that appears, click the Server Groups icon to display the list of server groups.

You can view all server groups under the project.
4. Click the name of the server group to be modified. On the Server Group Settings page that appears, click Modify in the upper-right corner.
5. Modify parameters of the server group, and then click Save.

Note:
The server group name cannot be changed.

Figure 3-9: Modify a server group

View the server group status

You can view the heartbeat status of a server group to check whether Logtail is successfully installed on all servers in the server group.

Note:
After servers are added to a server group, you need to wait approximately 2 minutes before the heartbeat status of the server group is updated.

1. Log on to the Log Service console.
2. Find the target project in the project list and click the project name.
3. In the left-side navigation pane on the page that appears, click the Server Groups icon to display the list of server groups.
4. Click the name of the target server group. On the Server Group Settings page that appears, check the server group status.

- If the heartbeat is **OK**, the Logtail client is installed on all servers and Logtail is connected to Log Service.
- If the heartbeat is **FAIL**, the Logtail connection is abnormal. If the heartbeat is **FAIL** for an extended period of time, perform troubleshooting based on the prompts and the following documentation: *What can I do if the Logtail client has no heartbeat?*. If the issue cannot be resolved, submit a ticket for assistance.

![Figure 3-10: View server group status](image)

Manage configurations

Log Service manages all the servers whose logs need to be collected through server groups. One important management item is the collection configuration of the Logtail client. For more information, see *Configuration procedure for text log collection* and #unique_47. You can add or delete a Logtail configuration for a server group to determine what logs are collected on each server by Logtail, how the logs are parsed, and which Logstore the logs are sent to.

1. Log on to the Log Service console.
2. Find the target project in the project list and click the project name.
3. In the left-side navigation pane on the page that appears, click the Server Groups icon to display the list of server groups.
4. Click the name of the server group to be modified. On the Server Group Settings page that appears, click Modify in the upper-right corner.
5. In the Configurations section, modify the Logtail configuration applied to the server group, and then click Save.

After a Logtail configuration is added, the configuration is pushed to the Logtail client on each server in the server group. After a Logtail configuration is removed, it is removed from the Logtail client.

Figure 3-11: Manage server group configurations

Delete a server group

1. Log on to the Log Service console.
2. Find the target project in the project list and click the project name.
3. In the left-side navigation pane on the page that appears, click the Server Groups icon to display the list of server groups.
4. Find the target server group, click the icon next to the server group, and then select Delete.
5. In the dialog box that appears, click OK.

Figure 3-12: Delete a server group

```
Delete Machine Group

⚠️ The machine group cannot be restored after being deleted. Do you want to delete it?

Confirm  Cancel
```

3.5 Text logs

3.5.1 Configuration procedure for text log collection

The Logtail client can help you collect text logs from Elastic Compute Service (ECS) instances or on-premises servers. You can specify the required settings in the Log Service console.

Prerequisites

- Logtail is installed. Logtail supports the Windows and Linux operating systems. For more information about how to install Logtail, see Install Logtail in Linux and Install Logtail in Windows.
- For the servers from which you want to collect logs, ports 80 and 443 are enabled for remote access.

Limits

- A log file can only be collected by a single configuration set. To collect multiple copies of a file, we recommend that you use symbolic links. For example, to collect two copies of a file in the `/home/log/nginx/log` directory, you can use the original log path for one Logtail configuration set. Then, you can run the `ln -s /home/log/nginx/log /home/log/nginx/link_log` command to create a symbolic link for this directory, and use the symbolic link as the log path for the other Logtail configuration set.
- For more information about the operating systems supported by the Logtail client, see Overview.
If you are using classic networks or VPCs, the ECS instances and the target Log Service project must belong to the same region. If your source data is transmitted over the Internet, you can select a region for the Log Service project based on your actual needs.

Configuration procedure for log collection

You can specify the Logtail settings in the Log Service console. Logtail supports various collection modes, including the Simple Mode, NGINX Configuration Mode, Apache Configuration Mode, IIS Configuration Mode, Delimiter Mode, JSON Mode, and Full Regex Mode.

Collection modes

Logtail supports various collection modes, including the Simple Mode, NGINX Configuration Mode, Apache Configuration Mode, IIS Configuration Mode, Delimiter Mode, JSON Mode, and Full Regex Mode.

- **Simple Mode**
  
  Logtail supports collecting logs in the Simple Mode

- **Full Regex Mode**
  
  Logtail supports collecting logs in the Full Regex Mode

- **Delimiter Mode**
  
  Logtail supports collecting logs in the Delimiter Mode. For more information, see DSV formatted logs.

- **JSON Mode**
  
  Logtail supports collecting logs in the JSON mode. For more information, see JSON logs.

- **NGINX Configuration Mode**
  
  Logtail supports collecting logs in the NGINX Configuration Mode. For more information, see NGINX logs.
• IIS Configuration Mode

Logtail supports collecting logs in the IIS Configuration Mode. For more information, see #unique_91.

• Apache Configuration Mode

Logtail supports collecting logs in the Apache Configuration Mode. For more information, see Apache logs.

Configuration procedure for log collection

1. Log on to the Log Service console.
2. Select a log source.

You can use one of the following three methods to select a log source.

- On the Log Service homepage, select a log source in the Import Data section.

- In the Projects section, click a project name. On the project overview page, click Import Data to go to the log source page.

- On the Logstores tab, find a Logstore and unfold its settings. Click the plus sign (+) next to Data Import to go to the log source page.
You can select a log source based on your business needs. Log Service supports the following log sources of text logs: RegEx-Text Log, Single Row-Text Log, Delimiter Mode-Text Log, JSON-Text Log, Nginx-Text Log, IIS-Text Log, and Apache-Text Log.

Select an existing project and Logstore. You can also click Create Now to create a project and Logstore. For more information, see #unique_93.

If you enter the configuration procedure by clicking the plus sign (+) next to Data Import under a Logstore on the Logstores tab, the system automatically skips this step.

4. Create a server group.

Before creating a server group, you must ensure that Logtail has been installed on the servers.

- Install Logtail on ECS instances. For Linux-based instances, click Install and Logtail is automatically installed. Windows-based instances do not support automatic installation of Logtail. In this case, you need to manually install Logtail. For more information, see Install Logtail in Windows.

- Install Logtail on user-created servers. You can install Logtail by following the prompted instructions. For more information, see Install Logtail in Linux or Install Logtail in Windows.

After you have installed Logtail, click Complete Installation to create a server group. For more information, see Overview. If you have created a server group, click Use Existing Server Groups.
5. **Configure server groups**

   Select a server group and move the group from Source Server Groups to Applied Server Groups.

![May Server Groups](image)

6. **Configure Logtail.**

   Configure Logtail based on the different collection modes. For more information, see configuration items for specific collection modes.

7. **Optional: Specify Advanced Options and click Next.**

   Specify Advanced Options based on your needs. We recommend that you keep the default settings unless otherwise required.

<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upload Raw Log</td>
<td>Specifies whether to upload raw logs. If you enable this feature, raw logs are written into the <strong>raw</strong> field and uploaded with the parsed logs.</td>
</tr>
<tr>
<td>Configuration item</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| **Topic Generation Mode** | • Null - Do not generate topic: This mode is selected by default. In this mode, the topic is set to a null string and you can query logs without entering a topic.  
• Server Group Topic Attributes: This mode is used to differentiate log data that is generated in different frontend servers.  
• File Path RegEx: If you select this mode, you must enter a value in the Custom RegEx field to extract part of the path as the topic. This mode is used to differentiate log data that is generated by users or instances. |
| **Custom RegEx**       | If you select File Path RegEx as the topic generation mode, you must enter a custom regular expression.                                                                                                     |
| **Log File Encoding**  | • utf8: indicates UTF-8 encoding.  
• gbk: indicates GBK encoding.                                                                                                                        |
| **Timezone**           | Specify the time zone where logs are collected.  
• System Timezone: This option is selected by default, which indicates that the time zone where logs are collected is the same as the time zone to which the server belongs.  
• Custom: You can select a time zone.                                                                                                           |
| **Timeout**            | If a log file is not updated within the specified period of time, the system considers that the file has timed out. The following Timeout options are available:  
• Never: All log files are continuously monitored and never time out.  
• 30 Minute Timeout: If a log file is not updated within 30 minutes, the system considers that the log file has timed out and no longer monitors the file. |
Log Service

Data Collection / 3 Logtail collection

<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filter Configuration</td>
<td>Only log entries that meet all filtering conditions are collected.</td>
</tr>
<tr>
<td></td>
<td>Example:</td>
</tr>
<tr>
<td></td>
<td>• Collect log entries that meet a condition: Set a condition Key:level Regex:WARNING</td>
</tr>
<tr>
<td></td>
<td>• Filter log entries that do not meet a condition:</td>
</tr>
<tr>
<td></td>
<td>- Set a condition Key:level Regex:^(?!.*(INFO</td>
</tr>
<tr>
<td></td>
<td>- Set a condition Key:url Regex:.<em>(?!.</em>(healthcheck)).*, which indicates that log entries with &quot;healthcheck&quot; in url are not collected. For example, log entries in which the key is url and the value is /inner/healthcheck/jiankong.html are not collected.</td>
</tr>
<tr>
<td></td>
<td>For more information about examples, see <code>regex-exclude-word</code> and <code>regex-exclude-pattern</code>.</td>
</tr>
</tbody>
</table>

8. Configure query and analytics settings.

The index is configured by default. You can reconfigure the index based on your business needs. For more information, see #unique_94.

After completing the preceding procedure, you can start to collect logs.

Logtail configuration items

The following table lists the configuration items and limits.

<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
</table>
| **Log Path** | Specifies the path of the log file. The specified log file name can be a complete file name or a file name that contains wildcards. Recursive directory matching is adopted in the log file search. If this matching method is applied, all files that match the specified file name under the specified directory and all its sub-directories are monitored.
 |
| **Log File Name** | The name of the log file to be collected, which is case-sensitive and can contain wildcards, for example, *.log. The file name wildcards in Linux include *, ?, and [...].
 |
| **Local Storage** | Specifies whether to enable the local cache to temporarily store log entries that cannot be sent due to short-term network interruptions.
 |
| **First-line Log Header** | The regular expression used to specify the starting header of a multi-line log file. Line breaks cannot be used to separate multi-line log entries, such as log entries that contain stack traces. In this case, you must specify the starting header of a multi-line log entry. When a header is detected, the previous log entry ends and a new log entry starts. The starting header may vary with each log entry, such as the timestamp. Therefore, you must specify a matching rule for the starting header, that is, a regular expression.
 |
| **Log Parsing Expression** | The regular expression that extracts the information of a log entry and converts it into a log format supported by Log Service. You must specify a regular expression to extract the required log fields and then name each extracted field.
 |
3.5.2 Collect logs by line

Log Service allows you to collect logs by line and configure indexes. You can specify the required settings in the Log Service console.

Context

To collect logs by line, you must select the Simple Mode. In this mode, each line of log data is considered as one log entry. Two log entries in a log file are separated by a line break.

Logtail does not extract log fields in this mode. The default regular expression is (. *). Logtail records the system time of the current server as the timestamp of a log entry. You can modify or manage advanced Logtail settings after you have completed the configuration procedure. For more information, see Manage collection configurations.

After you select the Simple Mode, you only need to specify the file directory and file name. Logtail collects logs by line without extracting log fields. Logtail records the system time at which a log entry is collected as the timestamp of the log entry.

This topic describes how to configure Logtail to collect logs by line. For more information about the configuration procedure for log collection, see Configuration procedure for text log collection.

Configuration procedure for log collection

1. Log on to the Log Service console.
2. Select a log source.
   Select Single Row-Text Log.

Select an existing project and Logstore. You can also click Create Now to create a project and Logstore. For more information, see #unique_93.

If you enter the configuration procedure by clicking the plus sign (+) next to Data Import under a Logstore on the Logstores tab, the system automatically skips this step.

4. Create a server group.

Before creating a server group, you must ensure that Logtail has been installed on the servers.

- Install Logtail on ECS instances. For Linux-based instances, click Install and Logtail is automatically installed. Windows-based instances do not support automatic installation of Logtail. In this case, you need to manually install Logtail. For more information, see Install Logtail in Windows.

- Install Logtail on user-created servers. You can install Logtail by following the prompted instructions. For more information, see Install Logtail in Linux or Install Logtail in Windows.

After you have installed Logtail, click Complete Installation to create a server group. For more information, see Overview. If you have created a server group, click Use Existing Server Groups.
5. Configure server groups

Select a server group and move the group from Source Server Groups to Applied Server Groups.

![May Server Groups](image)

6. Configure Logtail.

The following table lists the Logtail configuration items.

<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Config Name</td>
<td>The configuration name can only contain lowercase letters, digits, hyphens (-), and underscores (_). It must start and end with a lowercase letter or digit and must be 3 to 128 characters in length.</td>
</tr>
</tbody>
</table>

**Note:**
The configuration name cannot be modified after it is created.
<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
</table>
| Log Path           | Specifies the directory and name of the log file.  
  - The specified log file name can be a complete file name or a file name that contains wildcards. For more information about file naming conventions, see [Wildcard matching](#).  
  - Recursive directory matching is adopted in the log file search. If this matching method is applied, all files that match the specified file name under the specified directory and all its sub-directories are monitored.  
    - **Example 1**: `/apsara/nuwa/.../*.log` indicates the files whose extension is `.log` under the `/apsara/nuwa` directory and all its sub-directories are monitored.  
    - **Example 2**: `/var/logs/app_*.../*.log*` indicates the files whose file name contains `.log` in the following directories are monitored: the sub-directories of the `/var/logs` directory that match the `app_*` format and all the sub-directories of these matching sub-directories. |
| Docker File        | If the log file to be collected is in a Docker container, you can configure the internal path and container tag. Logtail automatically monitors the creation and destruction of the container, filters logs of the container based on the tag, and collects the filtered logs. For more information, see [unique_97](#). |
| Mode               | If you have specified Single Row-Text Log for the log source, the default mode is Simple Mode. You can manually change the mode. |

**Note:**  
- A log file can only be collected by a single configuration set.  
- Only the asterisk (*) and question mark (?) can be used as wildcards in the log path.
<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drop Failed to Parse</td>
<td>Specifies whether to upload logs that fail to be parsed to Log Service.</td>
</tr>
<tr>
<td></td>
<td>• If you enable this feature, logs that fail to be parsed are not uploaded to Log Service.</td>
</tr>
<tr>
<td></td>
<td>• If you disable this feature, raw logs are uploaded to Log Service when logs fail to be parsed.</td>
</tr>
<tr>
<td>Maximum Monitor Directory Depth</td>
<td>Specifies the maximum number of directory layers that can be recursively monitored when logs are</td>
</tr>
<tr>
<td></td>
<td>collected from the log source. The value ranges from 0 to 1000. The value 0 indicates that only</td>
</tr>
<tr>
<td></td>
<td>the directory that is specified in the log path is monitored.</td>
</tr>
</tbody>
</table>

7. Optional: Specify Advanced Options and click Next.

Specify Advanced Options based on your needs. We recommend that you keep the default settings unless otherwise required.

<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upload Raw Log</td>
<td>Specifies whether to upload raw logs. If you enable this feature, raw logs are written into the <strong>raw</strong> field and uploaded with the parsed logs.</td>
</tr>
<tr>
<td>Topic Generation Mode</td>
<td>• Null - Do not generate topic: This mode is selected by default. In this mode, the topic is set to a null string and you can query logs without entering a topic.</td>
</tr>
<tr>
<td></td>
<td>• Server Group Topic Attributes: This mode is used to differentiate log data that is generated in different frontend servers.</td>
</tr>
<tr>
<td></td>
<td>• File Path RegEx: If you select this mode, you must enter a value in the Custom RegEx field to extract part of the path as the topic. This mode is used to differentiate log data that is generated by users or instances.</td>
</tr>
<tr>
<td>Custom RegEx</td>
<td>If you select File Path RegEx as the topic generation mode, you must enter a custom regular expression.</td>
</tr>
<tr>
<td>Log File Encoding</td>
<td>• utf8: indicates UTF-8 encoding.</td>
</tr>
<tr>
<td></td>
<td>• gbk: indicates GBK encoding.</td>
</tr>
<tr>
<td>Configuration item</td>
<td>Description</td>
</tr>
<tr>
<td>--------------------</td>
<td>-------------</td>
</tr>
</tbody>
</table>
| **Timezone**       | Specify the time zone where logs are collected.  
|                     | - System Timezone: This option is selected by default, which indicates that the time zone where logs are collected is the same as the time zone to which the server belongs.  
|                     | - Custom: You can select a time zone. |
| **Timeout**        | If a log file is not updated within the specified period of time, the system considers that the file has timed out. The following Timeout options are available:  
|                     | - Never: All log files are continuously monitored and never time out.  
|                     | - 30 Minute Timeout: If a log file is not updated within 30 minutes, the system considers that the log file has timed out and no longer monitors the file. |
| **Filter Configuration** | Only log entries that meet all filtering conditions are collected.  
| Example: |  
| - Collect log entries that meet a condition: Set a condition Key:level Regex:WARNING|ERROR, which indicates that only log entries whose level is WARNING or ERROR are collected.  
| - Filter log entries that do not meet a condition:  
| - **Set a condition** Key:level Regex:^(?!.*(INFO|DEBUG)).*\|\*, which indicates that log entries whose level is INFO or DEBUG are not collected.  
| - **Set a condition** Key:url Regex:.*(?!.*(healthcheck)).*\|\*, which indicates that log entries with "healthcheck" in url are not collected. For example, log entries in which the key is url and the value is /inner/healthcheck/jiankong.html are not collected.  
| For more information about examples, see [regex-exclude-word](#) and [regex-exclude-pattern](#). |
8. Configure query and analytics settings.

The index is configured by default. You can reconfigure the index based on your business needs. For more information, see #unique_94.

After you specify the settings, you can start to collect logs by line.

3.5.3 Use regular expressions to collect logs

Log Service allows you to collect logs by using regular expressions and configure indexes. You can specify the required settings in the Log Service console.

Context

If you need to collect multi-line logs and extract fields from log entries, we recommend that you use regular expressions. Log Service provides the feature of generating a regular expression based on a sample log entry in the Import Data wizard. However, the expression must be manually modified for several times to fully conform to the sample log entry. For more information about how to specify a regular expression, see How do I modify a regular expression?

This topic describes how to configure Logtail when you need to use regular expressions to collect logs. For more information about the configuration procedure for log collection, see Configuration procedure for text log collection.

Configuration procedure for log collection

1. Log on to the Log Service console.

2. Select a log source.
   Select RegEx-Text Log.

   Select an existing project and Logstore. You can also click Create Now to create a project and Logstore. For more information, see #unique_93.

   If you enter the configuration procedure by clicking the plus sign (+) next to Data Import under a Logstore on the Logstores tab, the system automatically skips this step.
4. Create a server group.

Before creating a server group, you must ensure that Logtail has been installed on the servers.

- Install Logtail on ECS instances. For Linux-based instances, click Install and Logtail is automatically installed. Windows-based instances do not support automatic installation of Logtail. In this case, you need to manually install Logtail. For more information, see Install Logtail in Windows.
- Install Logtail on user-created servers. You can install Logtail by following the prompted instructions. For more information, see Install Logtail in Linux or Install Logtail in Windows.

After you have installed Logtail, click Complete Installation to create a server group. For more information, see Overview. If you have created a server group, click Use Existing Server Groups.

5. Configure server groups

Select a server group and move the group from Source Server Groups to Applied Server Groups.
6. Configure Logtail.

The following table lists the Logtail configuration items.

<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Config Name</strong></td>
<td>The name for a configuration set can only contain lowercase letters, digits, hyphens (-), and underscores (_). It must start and end with a lowercase letter or digit and must be 3 to 128 characters in length.</td>
</tr>
<tr>
<td><strong>Log Path</strong></td>
<td>Specifies the directory and name of the log file.</td>
</tr>
<tr>
<td></td>
<td>• The specified log file name can be a complete file name or a file name that contains wildcards. For more information about file naming conventions, see <a href="#">Wildcard matching</a>.</td>
</tr>
<tr>
<td></td>
<td>• Recursive directory matching is adopted in the log file search. If this matching method is applied, all files that match the specified file name under the specified directory and all its sub-directories are monitored.</td>
</tr>
<tr>
<td></td>
<td>- <strong>Example 1</strong>: <code>/apsara/nuwa/ ... /*.log</code> indicates the files whose extension is <code>.log</code> under the <code>/apsara/nuwa</code> directory and all its sub-directories are monitored.</td>
</tr>
<tr>
<td></td>
<td>- <strong>Example 2</strong>: <code>/var/logs/app_* ... /*.log</code> indicates the files whose file name contains <code>.log</code> in the following directories are monitored: the sub-directories of the <code>/var/logs</code> directory that match the <code>app_*</code> format and all the sub-directories of these matching sub-directories.</td>
</tr>
</tbody>
</table>

**Note:**
- A log file can only be collected by a single configuration set. |
- Only the asterisk (*) and question mark (?) can be used as wildcards in the log path.
<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Docker File</td>
<td>If the log file to be collected is in a Docker container, you can configure the internal path and container tag. Logtail automatically monitors the creation and destruction of the container, filters logs of the container based on the tag, and collects the filtered logs.</td>
</tr>
<tr>
<td>Mode</td>
<td>If you have specified RegEx-Text Log for the log source, the default mode is Full Regex Mode. You can manually change the mode.</td>
</tr>
<tr>
<td>Singleline</td>
<td>By default, the single-line feature is enabled, which indicates that log entries are separated by line. To collect multi-line logs, such as Java program logs, you must disable the Singleline feature and configure Regex to Match First Line.</td>
</tr>
<tr>
<td>Log Sample</td>
<td>Enter a sample log entry that is retrieved from the actual log source so that Log Service can automatically generate the regular expression.</td>
</tr>
<tr>
<td>Regex to Match First Line</td>
<td>You can click Auto Generate or Manual. After you enter a sample log entry and click Auto Generate, the system automatically generates a regular expression. If no regular expression is generated, you can switch to the manual mode and enter a regular expression for verification.</td>
</tr>
<tr>
<td>Extract Field</td>
<td>To analyze and process specific fields in log entries, you can use the Extract Field feature to convert the specified fields to key-value pairs that will be sent to Log Service. Therefore, you must specify a regular expression for parsing the log content.</td>
</tr>
<tr>
<td>Configuration item</td>
<td>Description</td>
</tr>
<tr>
<td>--------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| RegEx              | If you enable the Extract Field feature, you must specify this setting.  
|                    | • Automatically generate a regular expression  
|                    | You can select the fields to be extracted from the sample log entry and then click Generate Regular Expression. The system automatically generates a regular expression.  
|                    | • Manually input a regular expression  
|                    | You can also manually enter a regular expression. Click Manually to switch to the manual input mode. After you enter the regular expression, click Validate to check whether the regular expression can parse the log content. For more information about how to specify a regular expression, see [How do I modify a regular expression?](#). |
| Extracted Content  | If you enable the Extract Field feature, you must specify this setting.  
|                    | After a regular expression is automatically generated or manually specified, you must specify the key name for each extracted field. |
| Use System Time    | If you enable the Extract Field feature, you must specify this setting.  
<p>|                    | If you disable the Use System Time feature, you must specify a field as the time field during field extraction and name this field <code>time</code>. After you specify the <code>time</code> field, click Auto Generate in the Time Conversion Format field to automatically parse the time. For more information about time format conversion, see <a href="#">Configure the time format</a>. |</p>
<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
</table>
| Drop Failed to Parse           | Specifies whether to upload logs that fail to be parsed to Log Service.  
                                 |  · If you enable this feature, logs that fail to be parsed are not uploaded to Log Service.  
                                 |  · If you disable this feature, raw logs are uploaded to Log Service when logs fail to be parsed. |
| Maximum Monitor Directory Depth| Specifies the maximum number of directory layers that can be recursively monitored when logs are collected from the log source. The value ranges from 0 to 1000. The value 0 indicates that only the directory that is specified in the log path is monitored. |

7. Optional: Specify Advanced Options and click Next.

Specify Advanced Options based on your needs. We recommend that you keep the default settings unless otherwise required.

<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upload Raw Log</td>
<td>Specifies whether to upload raw logs. If you enable this feature, raw logs are written into the <strong>raw</strong> field and uploaded with the parsed logs.</td>
</tr>
</tbody>
</table>
| Topic Generation Mode       |  · Null - Do not generate topic: This mode is selected by default. In this mode, the topic is set to a null string and you can query logs without entering a topic.  
                                 |  · Server Group Topic Attributes: This mode is used to differentiate log data that is generated in different frontend servers.  
                                 |  · File Path RegEx: If you select this mode, you must enter a value in the Custom RegEx field to extract part of the path as the topic. This mode is used to differentiate log data that is generated by users or instances. |
| Custom RegEx                | If you select File Path RegEx as the topic generation mode, you must enter a custom regular expression.                                         |
| Log File Encoding           |  · utf8: indicates UTF-8 encoding.  
<pre><code>                             |  · gbk: indicates GBK encoding.                                                                                                             |
</code></pre>
<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Timezone</strong></td>
<td>Specify the time zone where logs are collected.</td>
</tr>
<tr>
<td></td>
<td>• System Timezone: This option is selected by default, which indicates that the time zone where logs are collected is the same as the time zone to which the server belongs.</td>
</tr>
<tr>
<td></td>
<td>• Custom: You can select a time zone.</td>
</tr>
<tr>
<td><strong>Timeout</strong></td>
<td>If a log file is not updated within the specified period of time, the system considers that the file has timed out. The following Timeout options are available:</td>
</tr>
<tr>
<td></td>
<td>• Never: All log files are continuously monitored and never time out.</td>
</tr>
<tr>
<td></td>
<td>• 30 Minute Timeout: If a log file is not updated within 30 minutes, the system considers that the log file has timed out and no longer monitors the file.</td>
</tr>
<tr>
<td><strong>Filter Configuration</strong></td>
<td>Only log entries that meet all filtering conditions are collected.</td>
</tr>
<tr>
<td></td>
<td>Example:</td>
</tr>
<tr>
<td></td>
<td>• Collect log entries that meet a condition: Set a condition `Key:level Regex:WARNING</td>
</tr>
<tr>
<td></td>
<td>• Filter log entries that do not meet a condition:</td>
</tr>
<tr>
<td></td>
<td>- <strong>Set a condition</strong> `Key:level Regex:^(?!.*(INFO</td>
</tr>
<tr>
<td></td>
<td>- <strong>Set a condition</strong> <code>Key:url Regex:.*^(?!.*(healthcheck)).*</code>, which indicates that log entries with &quot;healthcheck&quot; in url are not collected. For example, log entries in which the key is url and the value is <code>/inner/healthcheck/jiankong.html</code> are not collected.</td>
</tr>
</tbody>
</table>

For more information about examples, see [regex-exclude-word](#) and [regex-exclude-pattern](#).
8. Configure query and analytics settings.

The index is configured by default. You can reconfigure the index based on your business needs. For more information, see #unique_94.

After you specify the settings, you can start to use regular expressions to collect logs.

3.5.4 DSV formatted logs

Log Service allows you to collect delimiter-separated values (DSV) formatted logs and configure indexes. You can specify the required settings in the Log Service console.

Context

DSV formatted logs use line breaks as boundaries of log entries. Each line of log data indicates a log entry. The fields of each log entry are delimited by a fixed delimiter. The characters that can be used as delimiters include the tab (\t), space, vertical bar (|), comma (,), and semicolon (;). A field that contains a delimiter must be enclosed in a pair of double quotation marks (""), which are used as quotes.

This topic describes how to configure Logtail to collect DSV formatted logs. For more information about the configuration procedure for log collection, see Configuration procedure for text log collection.

Log formats

Common DSV formatted logs include comma-separated values (CSV) and tab-separated values (TSV) formatted logs.

A delimiter can contain a single character or multiple characters.

Single-character delimiter

You can specify a single-character delimiter and a quote in the console based on your needs.

· Delimiter: The fields of each log entry are delimited by a single-character delimiter, such as the tab (\t), vertical bar (|), space, comma (,), and semicolon (;). You can also specify a non-printable character for the delimiter.

Note:
The double quotation mark (") cannot be used as a delimiter.

If a double quotation mark (") is included in a log entry but not used as a quote, it must be escaped and processed as a pair of double quotation marks ("""). Log Service automatically restores the pair of double quotation marks (""") to a double quotation mark (") when parsing fields. You can use a double quotation mark (") on each border of a field as a quote, or use a double quotation mark (escaped as "") in the content of a field. If the use of double quotation marks (") does not comply with the defined format, you can use the Simple Mode or Full Regex Mode to parse fields.

An example is provided as follows. If you use commas as delimiters while double quotation marks and commas are included in a field, enclose the field with a pair of quotes and escape the double quotation marks into "". A log entry after being processed is 1999,Chevy,"Venture """"Extended Edition, Very Large"""","", 5000.00. The log entry can be parsed into five fields as follows: 1999, Chevy, Venture "Extended Edition, Very Large", empty field, and 5000.00.

- Quote: If a log field contains delimiters as part of the content, you must specify a quote to enclose the field so that the field can be parsed properly. Log Service parses the content enclosed in quotes as a complete field. Only delimiters can exist between fields.

  You can use one of the following characters as the quote: the tab (\t), the vertical bar (|), the space, the comma (,), the semicolon (;), and a non-printable character.

  For example, a log entry is 1997,Ford,E350,"ac, abs, moon",3000.00, in which the comma (,) is used as the delimiter and the double quotation mark (") is used as the quote. The log entry can be parsed into five fields as follows: 1997, Ford, E350, ac, abs, moon, and 3000.00. Among the five fields, ac, abs, moon enclosed in quotes is regarded as a complete field.

**Note:**
Log Service allows you to use a non-printable character as a delimiter or quote. Non-printable characters are those whose decimal ASCII codes are in the range of 1 to 31 and 127. If you use a non-printable character as the delimiter or quote, you need to find the hexadecimal ASCII code of this character and enter this character.
in the following format: 0xthe hexadecimal ASCII code of the non-printable character. For example, to use the non-printable character whose decimal ASCII code is 1 and hexadecimal ASCII code is 01, you need to enter 0x01.

A multi-character delimiter can contain two or three characters, such as ||, &&, and ^_^. If you specify a multi-character delimiter, Log Service only parses logs based on the delimiter. You do not need to use quotes to enclose log fields.

Note:
Ensure that log fields do not contain the delimiter. Otherwise, Log Service cannot properly parse these fields.

For example, if the delimiter is set to &&, the log entry 1997&&Ford&&E350&&ac&abs&moon&&3000.00 is parsed into the following five fields: 1997, Ford, E350, ac&abs&moon, and 3000.00.

Sample log entry

- Single-character delimiter

[Log entries]

05/May/2016:13:30:28,10.10. *. *, "POST /PutData? Category=YunOsAccountOpLog&AccessKeyId=****************&Date=Fri%2C%2028%20Jun%202013%2006%3A53%3A30%20GMT&Topic=raw&Signature=******************************** HTTP/1.1",200,18204,aliyun-sdk-java

05/May/2016:13:31:23,10.10. *. *, "POST /PutData? Category=YunOsAccountOpLog&AccessKeyId=****************&Date=Fri%2C%2028%20Jun%202013%2006%3A53%3A30%20GMT&Topic=raw&Signature=******************************** HTTP/1.1",200,18204,aliyun-sdk-java
Multi-character delimiter

05/May/2016:13:30:28&&10.200. **. **&&POST /PutData? Category=YunOsAccountOpLog&AccessKeyId=****************&Date=Fri%2C%2028%20Jun%202013%2006%3A53%3A30%20GMT&Topic=raw&Signature=pD12XYLmGxKQ%28mkd6x7hAgQ7b1c%3D%20HTTP/1.1&&200&&18204&&aliyun-sdk-java
05/May/2016:13:31:23&&10.200. **. **&&POST /PutData? Category=YunOsAccountOpLog&AccessKeyId=****************&Date=Fri%2C%2028%20Jun%202013%2006%3A53%3A30%20GMT&Topic=raw&Signature=pD12XYLmGxKQ%28mkd6x7hAgQ7b1c%3D%20HTTP/1.1&&200&&18204&&aliyun-sdk-java

Configuration procedure for log collection

1. Log on to the Log Service console.
2. Select a log source.
   Select Delimiter-Text Log.
   Select an existing project and Logstore. You can also click Create Now to create a project and Logstore. For more information, see #unique_93.
   If you enter the configuration procedure by clicking the plus sign (+) next to Data Import under a Logstore on the Logstores tab, the system automatically skips this step.
4. Create a server group.
   Before creating a server group, you must ensure that Logtail has been installed on the servers.

   • Install Logtail on ECS instances. For Linux-based instances, click Install and Logtail is automatically installed. Windows-based instances do not support automatic installation of Logtail. In this case, you need to manually install Logtail. For more information, see Install Logtail in Windows.
   • Install Logtail on user-created servers. You can install Logtail by following the prompted instructions. For more information, see Install Logtail in Linux or Install Logtail in Windows.

After you have installed Logtail, click Complete Installation to create a server group. For more information, see Overview. If you have created a server group, click Use Existing Server Groups.
5. Configure server groups

Select a server group and move the group from Source Server Groups to Applied Server Groups.

6. Configure Logtail.

The following table lists the Logtail configuration items.

<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Config Name</td>
<td>The name for a configuration set can only contain lowercase letters, digits, hyphens (-), and underscores (_). It must start and end with a lowercase letter or digit and must be 3 to 128 characters in length.</td>
</tr>
</tbody>
</table>

**Note:**
The name for a configuration set cannot be modified after it is created.
<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Path</td>
<td>Specifies the directory and name of the log file.</td>
</tr>
<tr>
<td></td>
<td>• The specified log file name can be a complete file name or a file name that contains wildcards. For more information about file naming rules, see <a href="#">Wildcard matching</a>.</td>
</tr>
<tr>
<td></td>
<td>• Recursive directory matching is adopted in the log file search. If this matching method is applied, all files that match the specified file name under the specified directory and all its sub-directories are monitored.</td>
</tr>
<tr>
<td></td>
<td>- <strong>Example 1:</strong> <code>/apsara/nuwa/ ... /*.log</code> indicates the files whose extension is <code>.log</code> under the <code>/apsara/nuwa</code> directory and all its sub-directories are monitored.</td>
</tr>
<tr>
<td></td>
<td>- <strong>Example 2:</strong> <code>/var/logs/app_* ... /*.log*</code> indicates the files whose file name contains <code>.log</code> in the following directories are monitored: the sub-directories of the <code>/var/logs</code> directory that match the <code>app_*</code> format and all the sub-directories of these matching sub-directories.</td>
</tr>
<tr>
<td></td>
<td><strong>Note:</strong></td>
</tr>
<tr>
<td></td>
<td>- A log file can only be collected by a single configuration set.</td>
</tr>
<tr>
<td></td>
<td>- Only the asterisk (*) and question mark (?) can be used as wildcards in the log path.</td>
</tr>
<tr>
<td>Docker File</td>
<td>If the log file to be collected is in a Docker container, you can configure the internal path and container tag. Logtail automatically monitors the creation and destruction of the container, filters logs of the container based on the tag, and collects the filtered logs.</td>
</tr>
<tr>
<td>Mode</td>
<td>If you have specified Delimiter-Text Log for the log source, the default mode is Delimiter Mode. You can manually change the mode.</td>
</tr>
<tr>
<td>Log Sample</td>
<td>Enter a sample log entry that is retrieved from the actual log source so that Log Service can automatically generate the regular expression.</td>
</tr>
<tr>
<td>Configuration item</td>
<td>Description</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| Delimiter           | Select a delimiter.  
Select an appropriate delimiter based on the log format. Otherwise, parsing may fail.  

**Note:**  
If you use a non-printable character as the delimiter, you need to find the hexadecimal ASCII code of this character and enter this character in the following format: `0x` the hexadecimal ASCII code of the non-printable character. For example, to use the non-printable character whose decimal ASCII code is 1 and hexadecimal ASCII code is 01, you need to enter `0x01`. |
| Quote               | Select a quote.  
Select an appropriate delimiter based on the log format. Otherwise, parsing may fail.  

**Note:**  
If you use a non-printable character as the quote, you need to find the hexadecimal ASCII code of this character and enter this character in the following format: `0x` the hexadecimal ASCII code of the non-printable character. For example, to use the non-printable character whose decimal ASCII code is 1 and hexadecimal ASCII code is 01, you need to enter `0x01`. |
| Extracted Content   | After you enter a sample log entry and select a delimiter, Log Service extracts log fields based on the delimiter and defines the fields as values. You need to specify a key for each value.                                                                                                                                                                                                                                                                                                                                 |

Note:  
If you use a non-printable character as the delimiter, you need to find the hexadecimal ASCII code of this character and enter this character in the following format: `0x` the hexadecimal ASCII code of the non-printable character. For example, to use the non-printable character whose decimal ASCII code is 1 and hexadecimal ASCII code is 01, you need to enter `0x01`.  

Note:  
If you use a non-printable character as the quote, you need to find the hexadecimal ASCII code of this character and enter this character in the following format: `0x` the hexadecimal ASCII code of the non-printable character. For example, to use the non-printable character whose decimal ASCII code is 1 and hexadecimal ASCII code is 01, you need to enter `0x01`.  

Extracted Content  
After you enter a sample log entry and select a delimiter, Log Service extracts log fields based on the delimiter and defines the fields as values. You need to specify a key for each value.
<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incomplete Entry Upload</td>
<td>Specifies whether to upload a log entry whose number of parsed fields is less than the number of the specified keys. If you enable this feature, the log entry is uploaded. If you disable this feature, the log entry is discarded. For example, if you set the delimiter to the vertical bar (</td>
</tr>
<tr>
<td>Use System Time</td>
<td>If you enable this feature, the system time when Logtail parses a log entry is used as the timestamp of the log entry. If you disable this feature, you must specify a field as the time field and name this field time. For more information about log time formats, see Configure the time format.</td>
</tr>
</tbody>
</table>
| Drop Failed to Parse       | Specifies whether to upload logs that fail to be parsed to Log Service. 
  • If you enable this feature, logs that fail to be parsed are not uploaded to Log Service. 
  • If you disable this feature, raw logs are uploaded to Log Service when logs fail to be parsed. |
### Configuration item

<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Monitor Directory Depth</td>
<td>Specifies the maximum number of directory layers that can be recursively monitored when logs are collected from the log source. The value ranges from 0 to 1000. The value 0 indicates that only the directory that is specified in the log path is monitored.</td>
</tr>
</tbody>
</table>

7. Optional: Specify Advanced Options and click Next.

Specify Advanced Options based on your needs. We recommend that you keep the default settings unless otherwise required.

<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upload Raw Log</td>
<td>Specifies whether to upload raw logs. If you enable this feature, raw logs are written into the <strong>raw</strong> field and uploaded with the parsed logs.</td>
</tr>
<tr>
<td>Configuration item</td>
<td>Description</td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| **Topic Generation Mode** | • Null - Do not generate topic: This mode is selected by default. In this mode, the topic is set to a null string and you can query logs without entering a topic.  
  • Server Group Topic Attributes: This mode is used to differentiate log data that is generated in different frontend servers.  
  • File Path RegEx: If you select this mode, you must enter a value in the Custom RegEx field to extract part of the path as the topic. This mode is used to differentiate log data that is generated by users or instances. |
| **Custom RegEx**     | If you select File Path RegEx as the topic generation mode, you must enter a custom regular expression.                                                                                                       |
| **Log File Encoding** | • utf8: indicates UTF-8 encoding.  
  • gbk: indicates GBK encoding.                                                                                                                                   |
| **Timezone**         | Specify the time zone where logs are collected.  
  • System Timezone: This option is selected by default, which indicates that the time zone where logs are collected is the same as the time zone to which the server belongs.  
  • Custom: You can select a time zone.                                                                                                                            |
| **Timeout**          | If a log file is not updated within the specified period of time, the system considers that the file has timed out. The following Timeout options are available:  
  • Never: All log files are continuously monitored and never time out.  
  • 30 Minute Timeout: If a log file is not updated within 30 minutes, the system considers that the log file has timed out and no longer monitors the file. |
### Filter Configuration

<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
</table>
| Filter Configuration | Only log entries that meet all filtering conditions are collected. Example:  
  * Collect log entries that meet a condition: Set a condition `Key:level Regex:WARNING|ERROR`, which indicates that only log entries whose level is WARNING or ERROR are collected.  
  * Filter log entries that do not meet a condition:  
    - Set a condition `Key:level Regex:^(?!.*(INFO|DEBUG)).*`, which indicates that log entries whose level is INFO or DEBUG are not collected.  
    - Set a condition `Key:url Regex:.*^(?!.*(healthcheck)).*`, which indicates that log entries with "healthcheck" in url are not collected. For example, log entries in which the key is url and the value is `/inner/healthcheck/jiankong.html` are not collected.  

For more information about examples, see `regex-exclude-word` and `regex-exclude-pattern`. |

8. Configure query and analytics settings.

The index is configured by default. You can reconfigure the index based on your business needs. For more information, see #unique_94.

After you specify the settings, you can start to collect DSV formatted logs.

3.5.5 JSON logs

Log Service allows you to collect JSON logs and configure indexes. You can specify the required settings in the Log Service console.

**Context**

Logtail supports parsing JSON objects from logs. It automatically extracts the keys and values from the first layer of an object as the names and values of log fields. The valid data types of field values include object, array, and primitive data types such as string or number.
JSON logs can be written in the following two types of structures:

- Object: a collection of key-value pairs.
- Array: an ordered list of values

\n is used to separate the lines of JSON logs. Each line is extracted as a single log entry.

Logtail can only automatically parse JSON logs of the object type. For JSON logs of other types, such as JSON arrays, you need to use regular expressions to extract the fields or specify the Simple Mode to collect logs by line.

This topic describes how to configure Logtail when you collect JSON logs. For more information about the configuration procedure for log collection, see Configuration procedure for text log collection.

Sample log entry

```
{"url": "POST /PutData? Category=YunOsAccountOpLog&AccessKeyId=U0Ujpek********&Date=Fri%2C%2028%20Jun%202013%2006%3A53%3A30%20GMT&Topic=raw&Signature=pD12XYLmGxKQ%2Bmkd6x7hAgQ7b1c%3D HTTP/1.1", "ip": "10.200.98.220", "user-agent": "aliyun-sdk-java", "request": {"status": "200", "latency": "18204"}, "time": "05/May/2016:13:30:28"}
{"url": "POST /PutData? Category=YunOsAccountOpLog&AccessKeyId=U0Ujpek********&Date=Fri%2C%2028%20Jun%202013%2006%3A53%3A30%20GMT&Topic=raw&Signature=pD12XYLmGxKQ%2Bmkd6x7hAgQ7b1c%3D HTTP/1.1", "ip": "10.200.98.210", "user-agent": "aliyun-sdk-java", "request": {"status": "200", "latency": "10204"}, "time": "05/May/2016:13:30:29"}
```

Configuration procedure for log collection

1. Log on to the Log Service console.
2. Select a log source.
   
   Select JSON-Text Log.
   
   Select an existing project and Logstore. You can also click Create Now to create a project and Logstore. For more information, see #unique_93.

If you enter the configuration procedure by clicking the plus sign (+) next to Data Import under a Logstore on the Logstores tab, the system automatically skips this step.
4. Create a server group.

Before creating a server group, you must ensure that Logtail has been installed on the servers.

- Install Logtail on ECS instances. For Linux-based instances, click Install and Logtail is automatically installed. Windows-based instances do not support automatic installation of Logtail. In this case, you need to manually install Logtail. For more information, see Install Logtail in Windows.
- Install Logtail on user-created servers. You can install Logtail by following the prompted instructions. For more information, see Install Logtail in Linux or Install Logtail in Windows.

After you have installed Logtail, click Complete Installation to create a server group. For more information, see Overview. If you have created a server group, click Use Existing Server Groups.

5. Configure server groups

Select a server group and move the group from Source Server Groups to Applied Server Groups.
6. Configure Logtail.

The following table lists the Logtail configuration items.

<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Config Name</strong></td>
<td>The name for a configuration set can only contain lowercase letters, digits, hyphens (-), and underscores (_). It must start and end with a lowercase letter or digit and must be 3 to 128 characters in length.</td>
</tr>
<tr>
<td><strong>Log Path</strong></td>
<td>Specifies the directory and name of the log file.</td>
</tr>
<tr>
<td></td>
<td>- The specified log file name can be a complete file name or a file name that contains wildcards. For more information about file naming conventions, see <a href="#">Wildcard matching</a>.</td>
</tr>
<tr>
<td></td>
<td>- Recursive directory matching is adopted in the log file search. If this matching method is applied, all files that match the specified file name under the specified directory and all its sub-directories are monitored.</td>
</tr>
</tbody>
</table>

- **Example 1**: `/apsara/nuwa/ ... /*.log` indicates the files whose extension is `.log` under the `/apsara/nuwa` directory and all its sub-directories are monitored.

- **Example 2**: `/var/logs/app_* ... /*.log*` indicates the files whose file name contains `.log` in the following directories are monitored: the sub-directories of the `/var/logs` directory that match the `app_*` format and all the sub-directories of these matching sub-directories.

<table>
<thead>
<tr>
<th>Note:</th>
</tr>
</thead>
<tbody>
<tr>
<td>- A log file can only be collected by a single configuration set.</td>
</tr>
<tr>
<td>- Only the asterisk (*) and question mark (?) can be used as wildcards in the log path.</td>
</tr>
<tr>
<td>Configuration item</td>
</tr>
<tr>
<td>--------------------</td>
</tr>
<tr>
<td><strong>Docker File</strong></td>
</tr>
<tr>
<td><strong>Mode</strong></td>
</tr>
<tr>
<td><strong>Use System Time</strong></td>
</tr>
</tbody>
</table>
| **Drop Failed to Parse** | Specifies whether to upload logs that fail to be parsed to Log Service.  
  - If you enable this feature, logs that fail to be parsed are not uploaded to Log Service.  
  - If you disable this feature, raw logs are uploaded to Log Service when logs fail to be parsed. |
<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Monitor Directory Depth</td>
<td>Specifies the maximum number of directory layers that can be recursively monitored when logs are collected from the log source. The value ranges from 0 to 1000. The value 0 indicates that only the directory that is specified in the log path is monitored.</td>
</tr>
</tbody>
</table>

- **Config Name:** test

- **Log Path:** C:\Program Files\Intel\*.*

  All files under the specified folder (including all directory levels) that conform to the file name convention will be monitored. The file name can be a complete name or a name that contains wildcards. The Linux file path must start with "/", for example, /apsarainuai/...app.Log. The Windows file path must start with a drive, for example, C:\Program Files\Intel\*.Log.

  **Docker File:**
  - For a Docker file, you can directly configure the log path and container tags. Container tags are specified by the configuration of the label whitelist and blacklist and environment variable whitelist and blacklist. Logtail will automatically monitor the creation and destruction of containers, and collect log entries of the specified containers according to the specified tags. For more information, see Documentation.

  **Mode:** JSON Mode

  **Use System Time:**

  - **Specify Time Key:** time
  - **Time Format:**

  **Drop Failed to Parse:**

  **Logs:** Enabled: Failed to parse logs will not be uploaded to Log Service. Disabled: Raw logs are uploaded to Log Service when the logs fail to parse.
7. Optional: Specify Advanced Options and click Next.

Specify Advanced Options based on your needs. We recommend that you keep the default settings unless otherwise required.

<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upload Raw Log</td>
<td>Specifies whether to upload raw logs. If you enable this feature, raw logs are written into the <code>__raw__</code> field and uploaded with the parsed logs.</td>
</tr>
</tbody>
</table>
| Topic Generation Mode    | • Null - Do not generate topic: This mode is selected by default. In this mode, the topic is set to a null string and you can query logs without entering a topic.  
                             • Server Group Topic Attributes: This mode is used to differentiate log data that is generated in different frontend servers.  
                             • File Path RegEx: If you select this mode, you must enter a value in the Custom RegEx field to extract part of the path as the topic. This mode is used to differentiate log data that is generated by users or instances. |
| Custom RegEx             | If you select File Path RegEx as the topic generation mode, you must enter a custom regular expression.                                           |
| Log File Encoding        | • utf8: indicates UTF-8 encoding.  
                             • gbk: indicates GBK encoding.                                                                                                           |
| Timezone                 | Specify the time zone where logs are collected.                                                                                             |
|                          | • System Timezone: This option is selected by default, which indicates that the time zone where logs are collected is the same as the time zone to which the server belongs.  
                             • Custom: You can select a time zone.                                                                                                      |
| Timeout                  | If a log file is not updated within the specified period of time, the system considers that the file has timed out. The following Timeout options are available:  
                             • Never: All log files are continuously monitored and never time out.  
                             • 30 Minute Timeout: If a log file is not updated within 30 minutes, the system considers that the log file has timed out and no longer monitors the file. |
<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filter Configuration</td>
<td>Only log entries that meet all filtering conditions are collected.</td>
</tr>
<tr>
<td>Example:</td>
<td>• Collect log entries that meet a condition: Set a condition Key:level Regex:WARNING</td>
</tr>
<tr>
<td></td>
<td>• Filter log entries that do not meet a condition:</td>
</tr>
<tr>
<td></td>
<td>- Set a condition Key:level Regex:^(!.*(INFO</td>
</tr>
<tr>
<td></td>
<td>- Set a condition Key:url Regex:^(!.<em>(healthcheck)].</em>), which indicates that log entries with &quot;healthcheck&quot; in url are not collected. For example, log entries in which the key is url and the value is /inner/healthcheck/jiankong.html are not collected.</td>
</tr>
</tbody>
</table>

For more information about examples, see [regex-exclude-word](#) and [regex-exclude-pattern](#).

8. Configure query and analytics settings.

The index is configured by default. You can reconfigure the index based on your business needs. For more information, see [unique_94](#).

After you specify the settings, you can start to collect JSON logs.

3.5.6 NGINX logs

Log Service allows you to collect NGINX logs and configure indexes. You can connect Log Service to NGINX and specify the required settings in the Log Service console.

Context

The NGINX log format and path are specified in the `/etc/nginx/nginx.conf` configuration file.
This topic describes how to configure Logtail when you collect NGINX logs. For more information about the configuration procedure for log collection, see Configuration procedure for text log collection.

NGINX log format

In the configuration file, the format of NGINX logs are defined as follows:

```bash
log_format main '$remote_addr - $remote_user [$time_local] "$request"
$request_time $request_length '
'$status $body_bytes_sent "$http_referer" '
'"$http_user_agent"';
```

The path of the log file is declared as follows. The "main" portion that follows the path indicates that logs are written in the preceding format.

```bash
access_log /var/logs/nginx/access.log main
```

Sample log entry

```
192.168.1.2 - - [10/Jul/2015:15:51:09 +0800] "GET /ubuntu.iso HTTP/1.0" 0.000 129 404 168 "-" "Wget/1.11.4 Red Hat modified"
```

Field description

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>remote_addr</td>
<td>The IP address of the client.</td>
</tr>
<tr>
<td>remote_user</td>
<td>The username of the client.</td>
</tr>
<tr>
<td>request</td>
<td>The request URL and HTTP protocol.</td>
</tr>
<tr>
<td>status</td>
<td>The response status of the request.</td>
</tr>
<tr>
<td>body_bytes_sent</td>
<td>The number of bytes in the response returned to the client, excluding the size of the header. The value of this field is the same as the value of bytes_sent in the mod_log_config module of Apache.</td>
</tr>
<tr>
<td>connection</td>
<td>The serial number of a connection.</td>
</tr>
<tr>
<td>connection_requests</td>
<td>The number of requests received from a connection.</td>
</tr>
<tr>
<td>msec</td>
<td>The time when the log entry is written, which is measured in seconds and accurate to milliseconds.</td>
</tr>
<tr>
<td>Field</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>pipe</td>
<td>Indicates whether the request is pipelined. If the request is pipelined, the field value is p. Otherwise, the field value is n.</td>
</tr>
<tr>
<td>http_referer</td>
<td>The URL of the webpage which is linked to the resource being requested.</td>
</tr>
<tr>
<td>&quot;http_user_agent&quot;</td>
<td>The browser information of the client, which must be enclosed by a pair of double quotation marks (&quot;&quot;).</td>
</tr>
<tr>
<td>request_length</td>
<td>The length of the request, which includes the request line, request header, and request body.</td>
</tr>
<tr>
<td>request_time</td>
<td>The time period for which the request is processed, which is measured in seconds and accurate to milliseconds. The time period starts when the first byte is read from the client and ends when the log entry is written after the last byte is sent to the client.</td>
</tr>
<tr>
<td>[$time_local]</td>
<td>The local time in the Common Log Format, which must be enclosed by brackets ([]).</td>
</tr>
</tbody>
</table>

Configuration procedure for log collection

1. Log on to the Log Service console.
2. Select a log source.
   Select Nginx-Text Log.
   Select an existing project and Logstore. You can also click Create Now to create a project and Logstore. For more information, see #unique_93.

If you enter the configuration procedure by clicking the plus sign (+) next to Data Import under a Logstore on the Logstores tab, the system automatically skips this step.
4. Create a server group.

Before creating a server group, you must ensure that Logtail has been installed on the servers.

- Install Logtail on ECS instances. For Linux-based instances, click Install and Logtail is automatically installed. Windows-based instances do not support automatic installation of Logtail. In this case, you need to manually install Logtail. For more information, see Install Logtail in Windows.
- Install Logtail on user-created servers. You can install Logtail by following the prompted instructions. For more information, see Install Logtail in Linux or Install Logtail in Windows.

After you have installed Logtail, click Complete Installation to create a server group. For more information, see Overview. If you have created a server group, click Use Existing Server Groups.

5. Configure server groups

Select a server group and move the group from Source Server Groups to Applied Server Groups.
6. Configure Logtail.

The following table lists the Logtail configuration items.

<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
</table>
| Config Name        | The name for a configuration set can only contain lowercase letters, digits, hyphens (-), and underscores (_). It must start and end with a lowercase letter or digit and must be 3 to 128 characters in length.  
   
   Note:  
   The name for a configuration set cannot be modified after it is created. |
| Log Path           | Specifies the directory and name of the log file.  
   - The specified log file name can be a complete file name or a file name that contains wildcards. For more information about file naming conventions, see Wildcard matching.  
   - Recursive directory matching is adopted in the log file search. If this matching method is applied, all files that match the specified file name under the specified directory and all its sub-directories are monitored.  
     - Example 1: /apsara/nuwa/ ... /*.log indicates the files whose extension is .log under the /apsara/nuwa directory and all its sub-directories are monitored.  
     - Example 2: /var/logs/app_* ... /*.log* indicates the files whose file name contains .log in the following directories are monitored: the sub-directories of the /var/logs directory that match the app_* format and all the sub-directories of these matching sub-directories.  
   
   Note:  
   - A log file can only be collected by a single configuration set.  
   - Only the asterisk (*) and question mark (?) can be used as wildcards in the log path. |
<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Docker File</td>
<td>If the log file to be collected is in a Docker container, you can configure the internal path and container tag. Logtail automatically monitors the creation and destruction of the container, filters logs of the container based on the tag, and collects the filtered logs.</td>
</tr>
<tr>
<td>Mode</td>
<td>If you have specified Nginx-Text Log for the log source, the default mode is NGINX Configuration Mode. You can manually change the mode.</td>
</tr>
<tr>
<td>NGINX Log Configuration</td>
<td>Enter the log configuration section that is specified in a standard NGINX configuration file. The section starts with <code>log_format</code>.</td>
</tr>
<tr>
<td>NGINX Key</td>
<td>Log Service automatically reads the keys of NGINX logs.</td>
</tr>
</tbody>
</table>
| Drop Failed to Parse       | Specifies whether to upload logs that fail to be parsed to Log Service.  
  · If you enable this feature, logs that fail to be parsed are not uploaded to Log Service.  
  · If you disable this feature, raw logs are uploaded to Log Service when logs fail to be parsed.                                               |
| Maximum Directory Monitoring Depth | Specifies the maximum number of directory layers that can be recursively monitored when logs are collected from the log source. The value ranges from 0 to 1000. The value 0 indicates that only the directory that is specified in the log path is monitored. |

7. Optional: Specify Advanced Options and click Next.

Specify Advanced Options based on your needs. We recommend that you keep the default settings unless otherwise required.

<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upload Raw Log</td>
<td>Specifies whether to upload raw logs. If you enable this feature, raw logs are written into the <strong>raw</strong> field and uploaded with the parsed logs.</td>
</tr>
<tr>
<td>Configuration item</td>
<td>Description</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| **Topic Generation Mode** | • Null - Do not generate topic: This mode is selected by default. In this mode, the topic is set to a null string and you can query logs without entering a topic.  
  • Server Group Topic Attributes: This mode is used to differentiate log data that is generated in different frontend servers.  
  • File Path RegEx: If you select this mode, you must enter a value in the Custom RegEx field to extract part of the path as the topic. This mode is used to differentiate log data that is generated by users or instances. |
| **Custom RegEx**         | If you select File Path RegEx as the topic generation mode, you must enter a custom regular expression.                                                                                                     |
| **Log File Encoding**    | • utf8: indicates UTF-8 encoding.  
  • gbk: indicates GBK encoding.                                                                                                                  |
| **Timezone**             | Specify the time zone where logs are collected.  
  • System Timezone: This option is selected by default, which indicates that the time zone where logs are collected is the same as the time zone to which the server belongs.  
  • Custom: You can select a time zone.                                                                                                             |
| **Timeout**              | If a log file is not updated within the specified period of time, the system considers that the file has timed out. The following Timeout options are available:  
  • Never: All log files are continuously monitored and never time out.  
  • 30 Minute Timeout: If a log file is not updated within 30 minutes, the system considers that the log file has timed out and no longer monitors the file. |
<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filter Configuration</td>
<td>Only log entries that meet all filtering conditions are collected.</td>
</tr>
<tr>
<td></td>
<td>Example:</td>
</tr>
<tr>
<td></td>
<td>· Collect log entries that meet a condition: Set a condition `Key:level Regex:WARNING</td>
</tr>
<tr>
<td></td>
<td>· <em>Filter log entries that do not meet a condition</em>:</td>
</tr>
<tr>
<td></td>
<td>- Set a condition `Key:level Regex:^(?!.*(INFO</td>
</tr>
<tr>
<td></td>
<td>- Set a condition <code>Key:url Regex:.*^(?!.*(healthcheck)).*</code>, which indicates that log entries with &quot;healthcheck&quot; in url are not collected. For example, log entries in which the key is url and the value is <code>/inner/healthcheck/jiankong.html</code> are not collected.</td>
</tr>
</tbody>
</table>

For more information about examples, see `regex-exclude-word` and `regex-exclude-pattern`.

8. Configure query and analytics settings.

The index is configured by default. You can reconfigure the index based on your business needs. For more information, see #unique_94.

After you specify the settings, you can start to collect NGINX logs.

3.5.7 IIS logs

Log Service allows you to collect Internet Information Services(IIS) logs and configure indexes. You can specify the required settings in the Log Service console.

Context

To meet the needs of analytics, we recommend that you use the W3C Extended Log Format. To use this format, click Select Fields in the IIS Manager, and then select sc-bytes and cs-bytes in the Standard Fields list. For more information about the W3C Extended Log Format, see #unique_91.
This topic describes how to configure Logtail when you collect IIS logs. For more information about the configuration procedure for log collection, see Configuration procedure for text log collection.

Configuration procedure for log collection

1. Log on to the Log Service console.
2. Select a log source.
   Select IIS-Text Log.
   Select an existing project and Logstore. You can also click Create Now to create a project and Logstore. For more information, see #unique_93.

   If you enter the configuration procedure by clicking the plus sign (+) next to Data Import under a Logstore on the Logstores tab, the system automatically skips this step.

4. Create a server group.
   Before creating a server group, you must ensure that Logtail has been installed on the servers.

   - Install Logtail on ECS instances. For Linux-based instances, click Install and Logtail is automatically installed. Windows-based instances do not support automatic installation of Logtail. In this case, you need to manually install Logtail. For more information, see Install Logtail in Windows.
   - Install Logtail on user-created servers. You can install Logtail by following the prompted instructions. For more information, see Install Logtail in Linux or Install Logtail in Windows.

   After you have installed Logtail, click Complete Installation to create a server group. For more information, see Overview. If you have created a server group, click Use Existing Server Groups.
5. Configure server groups

Select a server group and move the group from Source Server Groups to Applied Server Groups.

6. Configure Logtail.

The following table lists the Logtail configuration items.

<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Config Name</td>
<td>The name for a configuration set can only contain lowercase letters, digits, hyphens (-), and underscores (_). It must start and end with a lowercase letter or digit and must be 3 to 128 characters in length.</td>
</tr>
</tbody>
</table>

Note:
The name for a configuration set cannot be modified after it is created.
<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
</table>
| **Log Path**       | Specifies the directory and name of the log file.  
  - The specified log file name can be a complete file name or a file name that contains wildcards. For more information about file naming rules, see [Wildcard matching](#).  
  - Recursive directory matching is adopted in the log file search. If this matching method is applied, all files that match the specified file name under the specified directory and all its sub-directories are monitored.  
    - **Example 1**: `/apsara/nuwa/ ... /*.log` indicates the files whose extension is `.log` under the `/apsara/nuwa` directory and all its sub-directories are monitored.  
    - **Example 2**: `/var/logs/app_* ... /*.log*` indicates the files whose file name contains `.log` in the following directories are monitored: the sub-directories of the `/var/logs` directory that match the `app_*` format and all the sub-directories of these matching sub-directories.  
  
  **Note:**  
  - A log file can only be collected by a single configuration set.  
  - Only the asterisk (`*`) and question mark (`?`) can be used as wildcards in the log path. |
<p>| <strong>Docker File</strong>    | If the log file to be collected is in a Docker container, you can configure the internal path and container tag. Logtail automatically monitors the creation and destruction of the container, filters logs of the container based on the tag, and collects the filtered logs. For more information, see <a href="#">unique_97</a>. |
| <strong>Mode</strong>           | If you have specified IIS-Text Log for the log source, the default mode is IIS Configuration Mode. You can manually change the mode. |</p>
<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Log Format</strong></td>
<td>Select the log format of your IIS server.</td>
</tr>
<tr>
<td></td>
<td>• IIS: Microsoft IIS log file format</td>
</tr>
<tr>
<td></td>
<td>• NCSA: NCSA Common log file format</td>
</tr>
<tr>
<td></td>
<td>• W3C: W3C Extended Log Format</td>
</tr>
<tr>
<td><strong>IIS Configuration</strong></td>
<td>Enter the log configuration section that is specified in an IIS configuration file.</td>
</tr>
<tr>
<td></td>
<td>• If the specified log format is IIS or NCSA, the system automatically enters a commonly used syntax of the corresponding log format.</td>
</tr>
<tr>
<td></td>
<td>• If the specified log format is W3C, enter a syntax of the log format that is defined in the configuration file. For more information, see step 7.</td>
</tr>
<tr>
<td><strong>IIS Key Name</strong></td>
<td>Log Service automatically reads the keys of IIS logs.</td>
</tr>
<tr>
<td><strong>Drop Failed to Parse</strong></td>
<td>Specifies whether to upload logs that fail to be parsed to Log Service.</td>
</tr>
<tr>
<td></td>
<td>• If you enable this feature, logs that fail to be parsed are not uploaded to Log Service.</td>
</tr>
<tr>
<td></td>
<td>• If you disable this feature, raw logs are uploaded to Log Service when logs fail to be parsed.</td>
</tr>
<tr>
<td><strong>Maximum Monitor</strong></td>
<td>Specifies the maximum number of directory layers that can be recursively monitored when logs are collected from the log source. The value ranges from 0 to 1000. The value 0 indicates that only the directory that is specified in the log path is monitored.</td>
</tr>
<tr>
<td><strong>Directory Depth</strong></td>
<td></td>
</tr>
</tbody>
</table>
7. Specify the IIS Configuration field
   a) Open the IIS configuration file.
   
   - The default path for IIS5 configuration file: `C:\WINNT\system32\inetsrv\MetaBase.bin`
   - The default path for IIS6 configuration file: `C:\WINDOWS\system32\inetsrv\MetaBase.xml`
   - The default path for IIS7 configuration file: `C:\Windows\System32\inetsrv\config\applicationHost.config`

   b) Find the `logFile logExtFileFlags` field and copy the text inside the quotation marks that follows the field name.

   c) Paste the text inside the quotation marks in the IIS Configuration field.
8. Optional: Specify Advanced Options and click Next.

Specify Advanced Options based on your needs. We recommend that you keep the default settings unless otherwise required.

<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upload Raw Log</td>
<td>Specifies whether to upload raw logs. If you enable this feature, raw logs are written into the <strong>raw</strong> field and uploaded with the parsed logs.</td>
</tr>
</tbody>
</table>
| Topic Generation Mode  | • Null - Do not generate topic: This mode is selected by default. In this mode, the topic is set to a null string and you can query logs without entering a topic.  
  • Server Group Topic Attributes: This mode is used to differentiate log data that is generated in different frontend servers.  
  • File Path RegEx: If you select this mode, you must enter a value in the Custom RegEx field to extract part of the path as the topic. This mode is used to differentiate log data that is generated by users or instances. |
| Custom RegEx           | If you select File Path RegEx as the topic generation mode, you must enter a custom regular expression.                                      |
| Log File Encoding      | • utf8: indicates UTF-8 encoding.  
  • gbk: indicates GBK encoding. |
<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
</table>
| Timezone           | Specify the time zone where logs are collected.  
|                    | • System Timezone: This option is selected by default, which indicates that the time zone where logs are collected is the same as the time zone to which the server belongs.  
|                    | • Custom: You can select a time zone. |
| Timeout            | If a log file is not updated within the specified period of time, the system considers that the file has timed out. The following Timeout options are available:  
|                    | • Never: All log files are continuously monitored and never time out.  
|                    | • 30 Minute Timeout: If a log file is not updated within 30 minutes, the system considers that the log file has timed out and no longer monitors the file. |
| Filter Configuration | Only log entries that meet all filtering conditions are collected.  
| Example:           | Collect log entries that meet a condition: Set a condition Key:level Regex:WARNING|ERROR, which indicates that only log entries whose level is WARNING or ERROR are collected.  
|                    | Filter log entries that do not meet a condition:  
|                    | - Set a condition Key:level Regex:^(?!.*(INFO|DEBUG)).*, which indicates that log entries whose level is INFO or DEBUG are not collected.  
|                    | - Set a condition Key:url Regex:.^(?!.*(healthcheck)).*, which indicates that log entries with "healthcheck" in url are not collected. For example, log entries in which the key is url and the value is /inner/healthcheck/jiankong.html are not collected.  
|                    | For more information about examples, see regex-exclude-word and regex-exclude-pattern. |
9. **Configure query and analytics settings.**

   The index is configured by default. You can reconfigure the index based on your business needs. For more information, see #unique_94.

After you specify the settings, you can start to collect IIS logs.

### 3.5.8 Apache logs

Log Service allows you to collect Apache logs and configure indexes. You can specify the required settings in the Log Service console.

**Log formats**

The Apache configuration file defines two log formats: Combined Log Format and Common Log Format. You can also customize a log format based on your needs.

- **Syntax of the Combined Log Format:**
  
  ```markdown
  LogFormat "%h %l %u %t "%r" %>s %b "%{Referer}i" "%{User-Agent}i" combined
  ```

- **Syntax of the Common Log Format:**
  
  ```markdown
  LogFormat "%h %l %u %t "%r" %>s %b"
  ```

- **Syntax of a custom log format:**
  
  ```markdown
  LogFormat "%h %l %u %t "%r" %>s %b "%{Referer}i" "%{User-Agent}i" %D %f %k %p %q %R %T %I %O" customized
  ```

You must specify the log format, log file directory, and log file name in the Apache configuration file. For example, the following declaration in the configuration file indicates that the Combined Log Format is used, the log file directory is `/Var/log/apache2/access_log`, and the log file name is `access_log`.

```markdown
CustomLog "/var/log/apache2/access_log" combined
```

**Field description**

<table>
<thead>
<tr>
<th>Format string</th>
<th>Key name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>%a</td>
<td>client_addr</td>
<td>The IP address of the client in the request.</td>
</tr>
<tr>
<td>%A</td>
<td>local_addr</td>
<td>The local private IP address.</td>
</tr>
<tr>
<td>%b</td>
<td>response_size_bytes</td>
<td>The size of the response in bytes. If no bytes are sent, the value is &quot;-&quot;.</td>
</tr>
<tr>
<td>Format string</td>
<td>Key name</td>
<td>Description</td>
</tr>
<tr>
<td>---------------</td>
<td>------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>%B</td>
<td>response_bytes</td>
<td>The size of the response in bytes. If no bytes are sent, the value is 0.</td>
</tr>
<tr>
<td>%D</td>
<td>request_time_msec</td>
<td>The period of time for which the request is processed, which is measured in milliseconds.</td>
</tr>
<tr>
<td>%h</td>
<td>remote_addr</td>
<td>The name of the remote host.</td>
</tr>
<tr>
<td>%H</td>
<td>request_protocol_supple</td>
<td>The request protocol.</td>
</tr>
<tr>
<td>%l</td>
<td>remote_ident</td>
<td>The client log name, which is obtained from identd.</td>
</tr>
<tr>
<td>%m</td>
<td>request_method_supple</td>
<td>The request method.</td>
</tr>
<tr>
<td>%p</td>
<td>remote_port</td>
<td>The port number of the server.</td>
</tr>
<tr>
<td>%P</td>
<td>child_process</td>
<td>The ID of the child process.</td>
</tr>
<tr>
<td>%q</td>
<td>request_query</td>
<td>The query string. If it does not exist, the value is a null string.</td>
</tr>
<tr>
<td>“%r”</td>
<td>request</td>
<td>The request, which includes the method name, address, and HTTP protocol.</td>
</tr>
<tr>
<td>%s</td>
<td>status</td>
<td>The HTTP status code for the response.</td>
</tr>
<tr>
<td>%&gt;s</td>
<td>status</td>
<td>The HTTP status code for the final response.</td>
</tr>
<tr>
<td>%f</td>
<td>filename</td>
<td>The name of the requested file.</td>
</tr>
<tr>
<td>%k</td>
<td>keep_alive</td>
<td>The number of keep-alive requests.</td>
</tr>
<tr>
<td>%R</td>
<td>response_handler</td>
<td>The type of the handler that generates the response on the server.</td>
</tr>
<tr>
<td>%t</td>
<td>time_local</td>
<td>The local time when the server receives the request.</td>
</tr>
<tr>
<td>%T</td>
<td>request_time_sec</td>
<td>The time period for which the request is processed, which is measured in seconds.</td>
</tr>
<tr>
<td>Format string</td>
<td>Key name</td>
<td>Description</td>
</tr>
<tr>
<td>---------------</td>
<td>------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>%u</td>
<td>remote_user</td>
<td>The username that you used to log on to the client.</td>
</tr>
<tr>
<td>%U</td>
<td>request_uri_supple</td>
<td>The requested URL, excluding any query string.</td>
</tr>
<tr>
<td>%v</td>
<td>server_name</td>
<td>The name of the server.</td>
</tr>
<tr>
<td>%V</td>
<td>server_name_canonical</td>
<td>The server name based on the UseCanonicalName setting.</td>
</tr>
<tr>
<td>%I</td>
<td>bytes_received</td>
<td>The number of bytes that are received by the server. To use this field, you must enable the mod_logio module.</td>
</tr>
<tr>
<td>%O</td>
<td>bytes_sent</td>
<td>The number of bytes that are sent by the server. To use this field, you must enable the mod_logio module.</td>
</tr>
<tr>
<td>&quot;%{User-Agent}i&quot;</td>
<td>http_user_agent</td>
<td>The information about the client.</td>
</tr>
<tr>
<td>&quot;%{Referer}i&quot;</td>
<td>http_referer</td>
<td>The URL of the web page which is linked to the resource being requested.</td>
</tr>
</tbody>
</table>

Sample log entry

```
192.168.1.2 - - [02/Feb/2016:17:44:13 +0800] "GET /favicon.ico HTTP/1.1" 404 209 "http://localhost/x1.html" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_11_3) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/48.0.2564.97 Safari/537.36"
```

Configuration procedure for log collection

This topic describes how to configure Logtail to collect Apache logs. For more information about the configuration procedure for log collection, see Configuration procedure for text log collection.

1. Log on to the Log Service console.

2. Select a log source.
   - Select Apache-Text Log.

Select an existing project and Logstore. You can also click Create Now to create a project and Logstore. For more information, see #unique_93.

If you enter the configuration procedure by clicking the plus sign (+) next to Data Import under a Logstore on the Logstores tab, the system automatically skips this step.

4. Create a server group.

Before creating a server group, you must ensure that Logtail has been installed on the servers.

- Install Logtail on ECS instances. For Linux-based instances, click Install and Logtail is automatically installed. Windows-based instances do not support automatic installation of Logtail. In this case, you need to manually install Logtail. For more information, see Install Logtail in Windows.

- Install Logtail on user-created servers. You can install Logtail by following the prompted instructions. For more information, see Install Logtail in Linux or Install Logtail in Windows.

After you have installed Logtail, click Complete Installation to create a server group. For more information, see Overview. If you have created a server group, click Use Existing Server Groups.
5. Configure server groups

Select a server group and move the group from Source Server Groups to Applied Server Groups.

6. Configure Logtail.

The following table lists the Logtail configuration items.

<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Config Name</td>
<td>The name for a configuration set can only contain lowercase letters, digits, hyphens (-), and underscores (_). It must start and end with a lowercase letter or digit and must be 3 to 128 characters in length.</td>
</tr>
</tbody>
</table>

Note: The name for a configuration set cannot be modified after it is created.
<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
</table>
| Log Path | Specifies the directory and name of the log file.  
- The specified log file name can be a complete file name or a file name that contains wildcards. For more information about file naming rules, see *Wildcard matching*.  
- Recursive directory matching is adopted in the log file search. If this matching method is applied, all files that match the specified file name under the specified directory and all its sub-directories are monitored.  
  - Example 1: `/apsara/nuwa/ ... /*.log` indicates the files whose extension is `.log` under the `/apsara/nuwa` directory and all its sub-directories are monitored.  
  - Example 2: `/var/logs/app_* ... //*.log*` indicates the files whose file name contains `.log` in the following directories are monitored: the sub-directories of the `/var/logs` directory that match the `app_*` format and all the sub-directories of these matching sub-directories.  

**Note:**  
- A log file can only be collected by a single configuration set.  
- Only the asterisk (`*`) and question mark (`?`) can be used as wildcards in the log path. |
<p>| Docker File | If the log file to be collected is in a Docker container, you can configure the internal path and container tag. Logtail automatically monitors the creation and destruction of the container, filters logs of the container based on the tag, and collects the filtered logs. For more information, see <em>Container text logs</em>. |
| Mode | If you have specified Apache-Text Log for the log source, the default mode is Apache Configuration Mode. You can manually change the mode. |</p>
<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Format</td>
<td>Select the log format according to the format declared in your Apache log configuration file. To facilitate the query and analysis of log data, we recommend that you use a custom Apache log format.</td>
</tr>
<tr>
<td>APACHE Logformat Configuration</td>
<td>Enter the log configuration section that is specified in a standard Apache configuration file. The section starts with LogFormat.</td>
</tr>
<tr>
<td></td>
<td><strong>Note:</strong></td>
</tr>
<tr>
<td></td>
<td>If the specified log format is Common or Combined, the system automatically enters a commonly used syntax of the corresponding log format. Check whether the log format is the same as that defined in the local Apache configuration file.</td>
</tr>
<tr>
<td>APACHE Key Name</td>
<td>Log Service automatically reads the keys of Apache logs. Confirm the key names on the Logtail configuration page.</td>
</tr>
<tr>
<td>Drop Failed to Parse</td>
<td>Specifies whether to upload logs that fail to be parsed to Log Service.</td>
</tr>
<tr>
<td></td>
<td>• If you enable this feature, logs that fail to be parsed are not uploaded to Log Service.</td>
</tr>
<tr>
<td></td>
<td>• If you disable this feature, raw logs are uploaded to Log Service when logs fail to be parsed.</td>
</tr>
</tbody>
</table>
### Configuration item | Description
---|---
Maximum Monitor Directory Depth | Specifies the maximum number of directory layers that can be recursively monitored when logs are collected from the log source. The value ranges from 0 to 1000. The value 0 indicates that only the directory that is specified in the log path is monitored.

7. Optional: Specify Advanced Options and click Next.

Specify Advanced Options based on your needs. We recommend that you keep the default settings unless otherwise required.

### Configuration item | Description
---|---
Upload Raw Log | Specifies whether to upload raw logs. If you enable this feature, raw logs are written into the __raw__ field and uploaded with the parsed logs.
<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
</table>
| **Topic Generation Mode** | • Null - Do not generate topic: This mode is selected by default. In this mode, the topic is set to a null string and you can query logs without entering a topic.  
  • Server Group Topic Attributes: This mode is used to differentiate log data that is generated in different frontend servers.  
  • File Path RegEx: If you select this mode, you must enter a value in the Custom RegEx field to extract part of the path as the topic. This mode is used to differentiate log data that is generated by users or instances. |
| Custom RegEx            | If you select File Path RegEx as the topic generation mode, you must enter a custom regular expression.                                                                                                      |
| Log File Encoding       | • utf8: indicates UTF-8 encoding.  
  • gbk: indicates GBK encoding.                                                                                                                                         |
| Timezone                | Specify the time zone where logs are collected.  
  • System Timezone: This option is selected by default, which indicates that the time zone where logs are collected is the same as the time zone to which the server belongs.  
  • Custom: You can select a time zone.                                                                                                                                     |
| Timeout                 | If a log file is not updated within the specified period of time, the system considers that the file has timed out. The following Timeout options are available:  
  • Never: All log files are continuously monitored and never time out.  
  • 30 Minute Timeout: If a log file is not updated within 30 minutes, the system considers that the log file has timed out and no longer monitors the file. |
<table>
<thead>
<tr>
<th>Configuration item</th>
<th>Description</th>
</tr>
</thead>
</table>
| Filter Configuration | Only log entries that meet all filtering conditions are collected.  
Example:  
- Collect log entries that meet a condition: Set a condition `Key:level Regex:WARNING|ERROR`, which indicates that only log entries whose level is WARNING or ERROR are collected.  
- Filter log entries that do not meet a condition:  
  - Set a condition `Key:level Regex:^(?!.*(INFO|DEBUG)).*`, which indicates that log entries whose level is INFO or DEBUG are not collected.  
  - Set a condition `Key:url Regex:.*^(?!.*(healthcheck)).*`, which indicates that log entries with "healthcheck" in url are not collected. For example, log entries in which the key is url and the value is `/inner/healthcheck/jiankong.html` are not collected.  
For more information about examples, see `regex-exclude-word` and `regex-exclude-pattern`. |

8. Configure query and analytics settings.  
The index is configured by default. You can reconfigure the index based on your business needs. For more information, see #unique_94.  
After you specify the settings, you can start to collect Apache logs.

3.5.9 Configure and parse text logs  
Specify log line separation method  
A full access log is typically a row by line, such as the nginx's access log, each log is split with line breaks. For example, the following two access logs:

```
10.1.1.1 - - [13/Mar/2016:10:00:10 +0800] "GET / HTTP/1.1" 0.011 180 404 570 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; 360se)"
```
For Java applications, a program log usually spans several lines. The characteristic log header is used to separate two logs. For example, see the following Java program log:

```
[2016-03-18T14:16:16,000] [INFO] [SessionTracker] [SessionTrackerImpl. java:148] Expiring sessions
0x152436b9a12aeced, 50000
0x152436b9a12aed2, 50000
0x152436b9a12aedd, 50000
```
The preceding Java log has a starting field in the time format. The regular expression is `\d+-\d+-\w+:\d+\s.*`. You can complete the configurations in the console as follows.

Figure 3-14: Full mode parsing Regular Expression

Extract log fields

According to the Log Service data models, a log contains one or more key-value pairs. To extract specified fields for analysis, you must set a regular expression. If log content does not need to be processed, the log can be considered as a key-value pair.

For the access log in the previous example, you can choose to extract a field or not.
Log Service

Data Collection / 3 Logtail collection

- When fields are extracted
  
  Regular expression: `(\S+)\s-\s-\s\[(\S+)\s[^[\]]\]+\]\s"". *, Extracted contents: 10.1.1.1, 13/Mar/2016:10:00 and GET.

- When fields are not extracted
  
  Regular expression: (. *), Extracted contents: 10.1.1.1 - - [13/Mar/2016:10:00:10 +0800] "GET / HTTP/1.1" 0.011 180 404 570 "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; 360se)"

Specify log time

According to the Log Service data models, a log must have a time field in UNIX timestamp format. Currently, the log time can be set to the system time when Logtail collects the log or the time field in the log content.

For the access log in the previous example:

- Extract the time field in the log content
  Time expression: `%d/%b/%Y:%H:%M:%S`

- The system time when the log is collected
  Time: Timestamp when the log is collected.

3.5.10 Configure the time format

Each log in Log Service has a timestamp that records the log generation time. When collecting log data from your log files, Logtail must extract the timestamp string of each log and parse it into a timestamp. Therefore, you need to specify a timestamp format for parsing.

Logtail for Linux supports all time formats provided by the strftime function. Logtail only parses and uses the timestamp strings that can be expressed in the log formats defined by the strftime function.

Note:

- The log timestamp is accurate to seconds. Therefore, you only need to configure the time format to seconds, without the need for other information such as milliseconds or microseconds.
- In addition, you only need to configure the time field, instead of other information.
Common log time formats supported by Logtail

The timestamp strings of logs have diverse formats. To facilitate configuration, Logtail supports multiple log time formats, as described in the following table.

<table>
<thead>
<tr>
<th>Format</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>%a</td>
<td>The abbreviation of a day in a week.</td>
<td>Fri</td>
</tr>
<tr>
<td>%A</td>
<td>The full name of a day in a week.</td>
<td>Friday</td>
</tr>
<tr>
<td>%b</td>
<td>The abbreviation of a month.</td>
<td>Jan</td>
</tr>
<tr>
<td>%B</td>
<td>The full name of a month.</td>
<td>January</td>
</tr>
<tr>
<td>%d</td>
<td>The day in a month, in decimal format. Valid values: [01, 31].</td>
<td>07 or 31</td>
</tr>
<tr>
<td>%h</td>
<td>The abbreviation of a month, which is the same as %b.</td>
<td>Jan</td>
</tr>
<tr>
<td>%H</td>
<td>The hour in 24-hour format.</td>
<td>22</td>
</tr>
<tr>
<td>%I</td>
<td>The hour in 12-hour format.</td>
<td>11</td>
</tr>
<tr>
<td>%m</td>
<td>The month in decimal format.</td>
<td>08</td>
</tr>
<tr>
<td>%M</td>
<td>The minutes in decimal format. Valid values: [00, 59].</td>
<td>59</td>
</tr>
<tr>
<td>%n</td>
<td>The line break.</td>
<td>A line break</td>
</tr>
<tr>
<td>%p</td>
<td>The abbreviation of a period in 12-hour format.</td>
<td>AM or PM</td>
</tr>
<tr>
<td>%r</td>
<td>The time in 12-hour format, which is equivalent to %I:%M:%S %p.</td>
<td>11:59:59 AM</td>
</tr>
<tr>
<td>%R</td>
<td>The time expressed in hour and minutes, which is equivalent to %H:%M.</td>
<td>23:59</td>
</tr>
<tr>
<td>Format</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>---------</td>
</tr>
<tr>
<td><code>%S</code></td>
<td>The seconds in decimal format. Valid values: [00, 59].</td>
<td>59</td>
</tr>
<tr>
<td><code>%t</code></td>
<td>The tab character.</td>
<td>The tab character</td>
</tr>
<tr>
<td><code>%y</code></td>
<td>The year (excluding the century) in decimal format. Valid values: [00, 99].</td>
<td>04 or 98</td>
</tr>
<tr>
<td><code>%Y</code></td>
<td>The year in decimal format.</td>
<td>2004 or 1998</td>
</tr>
<tr>
<td><code>%C</code></td>
<td>The century in decimal format. Valid values: [00, 99].</td>
<td>16</td>
</tr>
<tr>
<td><code>%e</code></td>
<td>The day in a month, in decimal format. Valid values: [1, 31]. Prefix a space to a single-digit number.</td>
<td>7 or 31</td>
</tr>
<tr>
<td><code>%j</code></td>
<td>The day in a year, in decimal format. Valid values: [00, 366].</td>
<td>365</td>
</tr>
<tr>
<td><code>%u</code></td>
<td>The day in a week, in decimal format. Valid values: [1, 7]. A value of 1 indicates Monday.</td>
<td>2</td>
</tr>
<tr>
<td><code>%U</code></td>
<td>The week in a year, where Sunday is the first day of each week. Valid values: [00, 53].</td>
<td>23</td>
</tr>
<tr>
<td>Format</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>---------</td>
</tr>
<tr>
<td>%V</td>
<td>The week in a year, where Monday is the first day of each week. If the week at the beginning of January contains four or more days, this week is the first week of the year. If this week contains less than four days, the next week is considered as the first week of the year. Valid values: [01, 53].</td>
<td>24</td>
</tr>
<tr>
<td>%w</td>
<td>The day in a week, in decimal format. Valid values: [0, 6]. A value of 0 indicates Sunday.</td>
<td>5</td>
</tr>
<tr>
<td>%W</td>
<td>The week in a year, where Monday is the first day of each week. Valid values: [00, 53].</td>
<td>23</td>
</tr>
<tr>
<td>%c</td>
<td>The standard date and time.</td>
<td>To specify more information such as the long date or short date, you can use the preceding supported formats for more precise expression.</td>
</tr>
<tr>
<td>%x</td>
<td>The standard date.</td>
<td>To specify more information such as the long date or short date, you can use the preceding supported formats for more precise expression.</td>
</tr>
<tr>
<td>%X</td>
<td>The standard time.</td>
<td>To specify more information such as the long date or short date, you can use the preceding supported formats for more precise expression.</td>
</tr>
<tr>
<td>%s</td>
<td>The Unix timestamp.</td>
<td>1476187251</td>
</tr>
</tbody>
</table>
The following table lists the common log time formats, examples, and corresponding time expressions.

<table>
<thead>
<tr>
<th>Log time format</th>
<th>Example</th>
<th>Time expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Custom</td>
<td>2017-12-11 15:05:07</td>
<td>%Y-%m-%d %H:%M:%S</td>
</tr>
<tr>
<td>Custom</td>
<td>[2017-12-11 15:05:07.012]</td>
<td>[%Y-%m-%d %H:%M:%S]</td>
</tr>
<tr>
<td>RFC822</td>
<td>02 Jan 06 15:04 MST</td>
<td>%d %b %y %H:%M</td>
</tr>
<tr>
<td>RFC822Z</td>
<td>02 Jan 06 15:04-0700</td>
<td>%d %b %y %H:%M</td>
</tr>
<tr>
<td>RFC850</td>
<td>Monday, 02-Jan-06 15:04:05 MST</td>
<td>%A, %d-%b-%y %H:%M:%S</td>
</tr>
<tr>
<td>RFC1123</td>
<td>Mon, 02 Jan 2006 15:04:05 MST</td>
<td>%A, %d-%b-%y %H:%M:%S</td>
</tr>
<tr>
<td>RFC3339</td>
<td>2006-01-02T15:04:05Z07:00</td>
<td>%Y-%m-%dT%H:%M:%S</td>
</tr>
<tr>
<td>RFC3339Nano</td>
<td>2006-01-02T15:04:05.999999999Z07:00</td>
<td>%Y-%m-%dT%H:%M:%S</td>
</tr>
</tbody>
</table>

3.5.11 Import history logs

Logtail only collects incremental logs by default. If you want to import history logs, use the history log importing feature of Logtail.

Prerequisites

- Your Logtail must be v0.16.15 (Linux) or v1.0.0.1 (Windows) and later. If your Logtail is an earlier version, you must upgrade it to the latest version.
- Target history logs do not have to be covered by the collection configuration. If a log file has been collected by Logtail, Logtail will collect the log file again after you import history logs.
- The maximum interval between generating and importing local configurations is one minute.
- Due to the special action of loading local configurations, Logtail notifies you of this action by sending `LOAD_LOCAL_EVENT_ALARM` to your server.
- If you want to import a large number of log files, you must modify the startup parameters for Logtail. That is, increase the upper limit of the Logtail CPU and memory. We commend that you increase the CPU to 2.0 or any other
greater specification, and increase the memory to 512 MB or any other greater specification. For more information, see Configure startup parameters.

Context

Logtail collects logs based on the events that are detected by listening on or performing round robin for log modifications. Logtail can also load local configurations, and trigger log collections. Logtail collects history logs by loading local configurations.

You must perform the operations for importing history logs in the Logtail installation directory, which varies by operating systems.

- **Linux operating system:** /usr/local/ilogtail.
- **Windows operating system**
  - 32-bit Windows operating system: C:\Program Files\Alibaba\Logtail
  - 64-bit Windows operating system: C:\Program Files (x86)\Alibaba\Logtail

Procedure

1. **Create collection configurations**

   Configure the collection and apply the configuration to the machine group. If you only want to configure the collection for importing history logs, you can set a collection directory that does not exist. For more information about the collection configuration, see Configuration procedure for text log collection.

2. **Gets the configuration unique identity.**

   Obtain a unique identifier for the configuration from the user_log_config.json file in the Logtail installation directory. For more information, see Logtail installation directory. For the Linux operating system, run the grep command in this directory. For the Windows operating system, use a tool (for example, the text tool) to open the file.

   To view the identifier in the Linux operating system, do the following:

   ```bash
   grep "##" /usr/local/ilogtail/user_log_config.json | awk '{print $1}'
   
   "##1.0##log-config-test$multi"
   "##1.0##log-config-test$ecs-test"
   "##1.0##log-config-test$metric_system_test"
   "##1.0##log-config-test$redis-status"
   ```
3. Add local events.

Save local events to JSON file `local_event.json` in the Logtail installation directory (for more information, see Logtail installation directory). You must save the local events by using the following format:

```
[
  {
    "config" : "${your_config_unique_id}",
    "dir" : "${your_log_dir}",
    "name" : "${your_log_file_name}"
  },
  
  ...
]
```

- **Configuration items**

<table>
<thead>
<tr>
<th>Configuration items</th>
<th>Description:</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Config</td>
<td>Indicates the configuration unique identifier that is obtained in step 2.</td>
<td>##1.0##log-config-test$ecs-test</td>
</tr>
<tr>
<td>dir</td>
<td>Indicates the folder where logs are located.</td>
<td>/data/log</td>
</tr>
<tr>
<td>name</td>
<td>Indicates a log name that support wildcards.</td>
<td>access.log. 2018-08-08 or access.log*</td>
</tr>
</tbody>
</table>

Note:
The folder cannot end in `/`.

- **Configuration example**

For the Windows operating system, directly use a tool (for example, the text tool) to modify the file `local_event.json`. For the Linux operating system, add local events as follows:

```
$ cat /usr/local/ilogtail/local_event.json
[
  {
    "config": "##1.0##log-config-test$ecs-test",
    "dir": "//data/log",
```
• How can I check whether Logtail has loaded the configuration?

After you save local file `local_event.json`, Logtail loads this local configuration file to the memory within one minute, and clears the content in `local_event.json`.

You can check whether Logtail has read local events by following these methods:

- Check whether the content in `local_event.json` has been cleared. If cleared, Logtail has read the local configurations.

- Check whether the file `ilogtail.LOG` in the Logtail installation directory (for more information, see Logtail installation directory) includes `process local event` keywords. If the content in `local_event.json` has been cleared, but these keywords cannot be found, the local configuration file may be invalid and has been filtered out.

- Search the `Diagnose collection errors` result for the `LOAD_LOCAL_EVENT_ALARM` alarm.

• Logtail has loaded the configuration, but still cannot collect any data. How can I deal with this issue?

This issue may be caused by the following reasons:

- The configuration is invalid.
- The `config` item is available in the local configuration.
- The target log is not located in the specified path in the collection configuration.
- The target log has been collected.

3.5.12 Set a log topic

Log topics are used to classify logs. Topics can be specified when logs are written and serve as a filter when logs are queried.

Note:
No topic can be set for syslog logs.

Topic generation modes

You can set a topic when using Logtail to collect logs or when using API operations or SDKs to upload logs. The following topic generation modes are supported in the Log Service console: Null - Do not generate topic, Server Group Topic Attributes, and File Path RegEx.

- Null - Do not generate topic

When you configure Logtail to collect text logs in the Log Service console, the default log topic generation mode is Null - Do not generate topic. In this case, the topic is an empty string and you can query logs without entering a topic.

- Server Group Topic Attributes

This mode differentiates between logs that are generated on different servers. If the logs of different servers are stored in the same file or file path, you can divide the servers into different server groups to differentiate between the logs of different servers by topic. To do this, set different topic attributes for different server groups when you create server groups, and set Topic Generation Mode to Server Group Topic Attributes. Apply the previously created Logtail configuration to those server groups.

When the Server Group Topic Attributes mode is selected, Logtail uploads the topic attribute of the server group to which the current server belongs as the topic name to Log Service when reporting log data. When querying logs, you need to specify the topic (topic attribute) of the target server group as the query condition.

- File Path RegEx

This mode differentiates between logs that are generated by different users or instances. In some cases, the system stores logs in different directories for different users or instances but uses the same names for sub-directories and log files in these directories. As a result, Log Service cannot identify the user or instance that generates the logs when collecting log files. To address this problem, you can set Topic Generation Mode to File Path RegEx, enter the
regular expression that exactly matches the file path, and set the topic to the instance name.

- When File Path RegEx is selected, Logtail uploads the instance name as the topic name to Log Service when reporting log data. The topic generated varies based on your directory structure and configuration. You need to specify the instance name as the topic name when querying logs. For example, the following setting stores logs generated by different users or instances to different directories:

```
/logs
  | - /userA/serviceA
  | - service.log
  | - /userB/serviceA
  | - service.log
  | - /userC/serviceA
  | - service.log
```

- If multiple fields in the file path need to be extracted separately, you can use the multi-layer extraction method of `?P<key>`. The value of key can contain only lowercase letters and digits. Example:

```
/home/admin/serviceA/userB/access.log
\/(?P<service>\[^/\]+)/(?P<user>[^/\]+)/.*
```

The following custom tags are created for the logs:

```
"__tag__ : service : serviceA"
"__tag__ : user : userB"
```

**Note:**

Logtail version 0.16.19 and later are supported.

- If only the `/logs` file path and the `service.log` file name are configured, Logtail collects logs in the three directories that contain the service.log file. However, Log Service cannot identify the user or instance that generates specific logs. In this case, you can set Topic Generation Mode to File Path RegEx, and enter the `\/(.*)\/*
```

regular expression to extract instance names. After configuration, the following topics are generated for logs in different directories: userA, userB, and userC. You can then specify a topic to query logs.

**Note:**

In the regular expression of the file path, you need to escape forward slashes (`/`).
- **Static topic generation**

  **Set Topic Generation Mode to File Path RegEx, and enter** `customized:// + Custom topic` **in the Custom RegEx field.**

  ![Note: Logtail version 0.16.21 and later are supported.]

  **Set a log topic**

  1. **Configure Logtail in the Log Service console.** For more information, see the [Configuration procedure for text log collection](#).

  To set **Topic Generation Mode to Server Group Topic Attributes**, configure Topic on the server group creation or modification page.

  2. **On the page of configuring Logtail for data import,** expand Advanced Options and set **Topic Generation Mode.**

  ![Figure 3-15: Set a log topic](#)

  **Modify a log topic**

  To change the log topic generation mode, reconfigure **Topic Generation Mode** on the Logtail configuration page.
3.6 Container log collection

3.6.1 Kubernetes log collection process

Log Service uses Logtail to collect Kubernetes cluster logs and manages collection configuration through custom resource definition (CRD). This document describes how to install and use Logtail to collect Kubernetes cluster logs.

Configuration process

1. Run the installation command to install the alibaba-log-controller Helm package .

2. Choose the CRD or console to manage collection configuration as required.
Step 1 Install Logtail

Installation of managed Kubernetes clusters in Container Service

For more information about how to install Logtail in a managed Kubernetes cluster, see the Manually install Log Service components section in Use Log Service to collect Kubernetes cluster logs.

Installation of Kubernetes on Alibaba Cloud Container Service

Procedure

1. Log on to the master node of your Alibaba Cloud Container Service Kubernetes. For more information, see #unique_105.
2. Replace ${your_k8s_cluster_id} in the following command with your Kubernetes cluster ID and run the command:

   ```bash
   ```

   After installation, Log Service automatically creates a Log Service project in the same region of your Kubernetes cluster. The name of the created project is k8s-log-${your_k8s_cluster_id}. Under the project, machine group k8s-group-${your_k8s_cluster_id} is created automatically.

   **Note:**
   - The config-operation-log Logstore, which is used to store operational logs of alibaba-log-controller, is automatically created under the k8s-log-${your_k8s_cluster_id} project. You cannot delete this Logstore because it is required for alibaba-log-controller troubleshooting.
   - To store the collected logs in an existing project, run the installation command and confirm that the project belongs to the region where your Kubernetes cluster is deployed.

   **Installation example**

   After successful execution, the following information is displayed:

   ```bash
   [root@iZbp******biaZ ~]# wget http://logtail-release-cn-hangzhou.oss-cn-hangzhou.aliyuncs.com/kubernetes/alicloud-log-k8s-install.sh -O
   ```
You can run `helm Status alibaba-log-controller` to check the current states of pods. If all states are successful, installation is successful.

After successful installation, log on to the Log Service console. The Log Service project automatically created is displayed on the console. (If you have many projects, search the keyword `k8s-log`.)

Self-built Kubernetes installation

**Prerequisites**

1. The Kubernetes cluster must be version 1.8 or later.
2. Helm 2.6.4 or later has been installed.

Installation procedure

1. In the Log Service console, create a project. The project name must begin with k8s-log-custom-.

2. In the following command, replace the parameters with your own, and run the command.

```
wget http://logtail-release-cn-hangzhou.oss-cn-hangzhou.aliyuncs.com/kubernetes/alicloud-log-k8s-custom-install.sh; chmod 744 ./alicloud-log-k8s-custom-install.sh; sh ./alicloud-log-k8s-custom-install.sh {your-project-suffix} {region-id} {aliuid} {access-key-id} {access-key-secret}
```

The parameters and their descriptions are as follows:

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>{your-project-suffix}</td>
<td>The maid-later part of the project name that you created in the second step. k8s-log-custom- that you have created in the second step. For example, the created project is k8s-log-custom- xxxx, then you must enter xxxx.</td>
</tr>
<tr>
<td>{regionId}</td>
<td>The ID of the region where your project is located. You can view the #unique_26, for example, the region ID of China East 1 (Hangzhou) is cn-hangzhou</td>
</tr>
<tr>
<td>{aliuid}</td>
<td>User ID (AliUid). You need to replace the parameter with your Alibaba Cloud account ID AliUid.</td>
</tr>
</tbody>
</table>

Note: AliUid is a string of characters. For more information about how to view your AliUid, see step 1 in Configure AliUIds for ECS servers under other Alibaba Cloud accounts or on-premises IDCs.
<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>{access-key-id}</td>
<td>Your account access key ID. We recommend that you use the sub-account AccessKey and grant AliyunLogFullAccess permission. For more information, see #unique_106.</td>
</tr>
<tr>
<td>{access-key-secret}</td>
<td>Your account access key secret. We recommend that you use the sub-account AccessKey and grant AliyunLogFullAccess permission. For more information, see #unique_106.</td>
</tr>
</tbody>
</table>

After installation, Log Service automatically creates a machine group in the project. The machine group name is k8s-group-$\{your_k8s_cluster_id\}$.

**Note:**
- Logstore config-operation-log is automatically created in the project k8s-log-$\{your_k8s_cluster_id\}$. Do not delete this Logstore.
- After self-built kubernetes installation, Logtail is granted privileged permissions to avoid the error during the deletion of other pods' container text file busy error during the deletion of other pods. For more information, see bug 1468249, bug 1441737, and issue 34538.

**Installation example**

The output of the successful execution is as follows:

```
[root@iZbpldssxxqfbiaZ ~]# wget http://logtail-release-cn-hangzhou.oss-cn-hangzhou.aliyuncs.com/kubernetes/alicloud-log-k8s-custom-install.sh; chmod 744 ./alicloud-log-k8s-custom-install.sh; sh ./alicloud-log-k8s-custom-install.sh xxxx cn-hangzhou 165xxxxxxxx050 LTAxxxxxxxxxxx AIxxxxxxxxxxxxxxxxxxxxxxxxxxxxxe
....
....
....
NAME:   alibaba-log-controller
LAST DEPLOYED: Fri May 18 16:52:38 2018
NAMESPACE: default
STATUS: DEPLOYED
RESOURCES:
  => v1beta1/ClusterRoleBinding
  NAME                  AGE
  alibaba-log-controller  0s
  => v1beta1/DaemonSet
  NAME                  DESIRED CURRENT READY UP-TO-DATE AVAILABLE NODE
  SELECTOR AGE
```
You can use the `helm status alibaba-log-controller` to view the current pod status. If all the statuses are successful, the installation is complete.

Log on to the Log Service console after installation. You can view the automatically created Log Service project. If you have many projects, search by the keyword `k8s-log`.

Step 2 Configure

Log collection supports the console configuration mode by default. Meanwhile, CRD configuration mode for the Kubernetes microservice development is also provided. You can use kubectl to manage the configuration. The comparison of the two configurations is as follows:

<table>
<thead>
<tr>
<th></th>
<th>CRD Mode</th>
<th>Console mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operational complexity</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>Function</td>
<td>Supports advanced configuration with the exception of Console mode</td>
<td>Medium</td>
</tr>
<tr>
<td>Complexity</td>
<td>Medium</td>
<td>Low</td>
</tr>
<tr>
<td>Network connection</td>
<td>Connect to the Kubernetes cluster</td>
<td>Connect to the Internet</td>
</tr>
</tbody>
</table>
Log Service

<table>
<thead>
<tr>
<th>-</th>
<th>CRD Mode</th>
<th>Console mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integration with</td>
<td>Supported</td>
<td>Not supported</td>
</tr>
<tr>
<td>deployment components</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Authentication method</td>
<td>Kubernetes authorization</td>
<td>Cloud account authentication</td>
</tr>
</tbody>
</table>

We recommend you use the CRD method for collection configuration management, as this method is better integrated with the Kubernetes deployment and publishing process.

Manage collection configurations on the console

Create Logtail collection configurations on the console as required. For configuration steps, see:

- Container text log (recommended)
- Container standard output (recommended)
- Host text file

By default, the root directory of the host is mounted to the `/logtail_host` directory of the Logtail container. You must add this prefix when configuring the path. For example, to collect data in the `/home/logs/app_log/` directory of the host, you must set the log path on the configuration page to `/logtail_host/home/logs/app_log/`.

- #unique_47

Acquisition configuration through CRD Management

For the Kubernetes microservice development model, the logging service also provides a way to configure the CRD, you can directly use kubectl to manage the configuration, the integration of this approach with the Kubernetes deployment and release process is more complete.

For more information, see Configure CRDs to collect Kubernetes logs.

Other operations

DaemonSet deployment migration procedure

If you previously deployed the Log Service logtail by using the WebSphere set method that you used earlier, you will not be able to use CRD for configuration management. You can migrate to a new version in the following ways:
Note:
During the upgrade, some logs are duplicated. The CRD management configuration can be used only for the configuration created using the CRD. The historical configuration does not support the CRD management mode because the historical configuration is created using a non-CRD mode.

1. Install in the form of a new version, the installation command last adds a parameter for the Log Service Project name that was used by your previous kubernetes cluster.

   For example, the project name was k8s-log-demo, the cluster ID was c12ba2028cxxxxxxxxxx6939f0b, then the installation command is

   ```
   ```

2. After successful installation, in the Log Service console apply the historical collection configuration to the new machine group k8s-group-${your_k8s_cluster_id}.

3. In a minute, the historical collection configuration is bind to the historical machine group.

4. After the log collection is normalized, you can delete the previously installed Logtail DaemonSet.

Use multiple clusters in the same Log Service project

   You can use multiple clusters to collect logs to the same Log Service project. When installing other clusters Log Service components, you must replace ${your_k8s_cluster_id} in the installation parameters with the clusters ID you installed for the first time.

   For example, you now have three clusters with IDs: abc001, abc002, and abc003. The installation parameters for the three clusters, ${your_k8s_cluster_id}, must all be filled as abc001.

Note:
This method does not support Kubernetes multi-cluster sharing across regions.
Logtail container logs

Logtail logs are stored in the /usr/local/ilogtail/ directory in the Logtail container, the file name is ilogtail.LOG and ilogtail.plugin, the container stdout does not have the reference significance, so you can ignore the following stdout output:

```
start umount useless mount points, /shm$|/merged$|/mqueue$
umount: /logtail_host/var/lib/docker/overlay2/3fd0043af174cb0273c3c7869500fbe2b95d13b1e110172ef57fe840c82155/merged: must be superuser to umount
umount: /logtail_host/var/lib/docker/overlay2/d5b10aa19399992755d61f85d25009528daa749c1bf8c16edff44beab6e69718/merged: must be superuser to umount
umount: /logtail_host/var/lib/docker/overlay2/5c3125daddacedec29df72ad0c52fac8000c6e880dc4e8a640b1e16c22d2e/merged: must be superuser to umount
......
xargs: umount: exited with status 255; aborting
umount done
start logtail
ilogtail is running
logtail status:
ilogtail is running
```

View the status of log related components in the Kubernetes cluster

```
helm status alibaba-log-controller
```

alibaba-log-controller failed to start

Make sure that you perform the installation as follows:

1. The installation command is executed on the master node of the Kubernetes Cluster
2. The installation command parameter is entered in the cluster ID.

If the installation fails due to these problems, use `helm del --purge alibaba-log-controller` to remove the installation package and perform the installation again.

If the installation failure persists, open a ticket to contact technical support engineers of Log Service.

Check the status of the Logtail DaemonSet in the Kubernetes cluster

```
You can run the command kubectl get ds -n kube-system to check the running status of Logtail.
```

Note:
The default namespace of Logtail is kube-system.

Check the version number, IP address, and startup time of Logtail

An example is as follows:

```
[root@iZbp1dsu6v77zfb40qfbiaZ ~]# kubectl get po -n kube-system | grep logtail
NAME            READY     STATUS    RESTARTS   AGE
logtail-ds-gb92k   1/1       Running   0          2h
logtail-ds-wm7lw   1/1       Running   0          4d
[root@iZbp1dsu6v77zfb40qfbiaZ ~]# kubectl exec logtail-ds-gb92k -n kube-system cat /usr/local/ilogtail/app_info.json
{
  "UUID" : "",
  "hostname" : "logtail-ds-gb92k",
  "instance_id" : "0EBB2B0E-0A3B-11E8-B0CE-0A58AC140402_172.20.4.2_1517810940",
  "ip" : "172.20.4.2",
  "logtail_version" : "0.16.2",
  "os" : "Linux; 3.10.0-693.2.2.el7.x86_64; #1 SMP Tue Sep 12 22:26:13 UTC 2017; x86_64",
  "update_time" : "2018-02-05 06:09:01"
}
```

View the operational log for Logtail

Logtail running logs are stored in the /usr/local/ilogtail/ directory. The file name is ilogtail.LOG. The rotation file is compressed and stored as ilogtail.LOG.x.gz.

An example is as follows:

```
[root@iZbp1dsu6v77zfb40qfbiaZ ~]# kubectl exec logtail-ds-gb92k -n kube-system tail /usr/local/ilogtail/ilogtail.LOG
[2018-02-05 06:09:02.168693] [INFO] [9] [build/release64/sls/ilogtail/LogtailPlugin.cpp:104] logtail plugin Resume:start
[2018-02-05 06:09:02.168807] [INFO] [9] [build/release64/sls/ilogtail/LogtailPlugin.cpp:106] logtail plugin Resume:success
[2018-02-05 06:09:02.168822] [INFO] [9] [build/release64/sls/ilogtail/EventDispatcher.cpp:369] start add existed check point events, size:0
[2018-02-05 06:09:02.168827] [INFO] [9] [build/release64/sls/ilogtail/EventDispatcher.cpp:511] add existed check point events, size:0 cache size:0 event size:0 success count:0
```

Restart Logtail of a Pod

An example is as follows:

```
[root@iZbp1dsu6v77zfb40qfbiaZ ~]# kubectl exec logtail-gb92k -n kube-system /etc/init.d/ilogtailld stop
kill process Name: ilogtail pid: 7
kill process Name: ilogtail pid: 9
stop success
[root@iZbp1dsu6v77zfb40qfbiaZ ~]# kubectl exec logtail-gb92k -n kube-system /etc/init.d/ilogtailld start
```
3.6.2 Container text logs

Logtail collects text logs generated in containers and uploads these logs together with the container-related metadata information to Log Service.

Characteristics

Compared with basic log file collection, Docker file collection has the following characteristics:

- Allows you to configure the log path of a container, without the need to consider the mapping between the path and the host.
- Supports the use of labels to specify the containers whose logs are to be collected.
- Supports the use of labels to exclude specific containers.
- Supports the use of environment variables to specify the containers whose logs are to be collected.
- Supports the use of environment variables to exclude specific containers.
- Supports multiline logs such as Java Stack logs.
- Supports automatic labeling for container data.
- Supports automatic labeling for Kubernetes containers.

Limits

- Collection stop policy: When a container is stopped and Logtail detects the die event on the container, Logtail stops collecting logs of the container (with a latency of no more than 3 seconds). In this case, if a collection latency occurs, some logs generated before the stop action may be lost.
- Docker storage driver: Only overlay and overlay2 are supported. For other storage drivers, you must mount the log directory to the local host.
- Logtail running mode: Logtail must run in a container and must be deployed based on Logtail deployment solutions.
- Label: refers to the label information in docker inspect. It is not synonymous with labels in Kubernetes.
- Environment: refers to the environment information configured during container startup.
Configuration process

1. **Deploy and configure the Logtail container.**
2. **Configure log collection in Log Service.**

Logtail deployment and configuration

- **Kubernetes**
  
  For more information about Kubernetes log collection, see *Logtail deployment solution for collecting Kubernetes logs*.

- **Other container management methods**

  For more information about other container management methods such as Swarm and Mesos, see *Common deployment solution for collecting Docker logs*.

Collection configuration

1. Log on to the *Log Service console*, and click the name of the target project.
2. Click the Import Data button. On the Import Data page that appears, select Docker File.
3. Select a Logstore.
   
   Select an existing Logstore, or create a new project or Logstore.
4. Create and configure a server group.
   
   Before you create a server group, make sure that you have installed Logtail. After installing Logtail, click Complete Installation to create a server group. If you have created a server group, click Use Existing Server Groups.
   
   Select a server group and move the server group from Source Server Groups to Applied Server Groups.
5. **Configure a data source.**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Required</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Docker File</td>
<td>Yes</td>
<td>This parameter is used to check whether the collected target file is a Docker file.</td>
</tr>
<tr>
<td>Parameter</td>
<td>Required</td>
<td>Description</td>
</tr>
<tr>
<td>--------------------</td>
<td>----------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Label Whitelist</td>
<td>No</td>
<td>LabelKey is required. If LabelValue is not empty, only containers whose labels contain LabelKey = LabelValue are collected. If LabelValue is empty, all the containers whose labels contain LabelKey are collected.</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Note:</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>· Multiple key-value pairs are associated by the OR operator. When the label of a container contains any of the key-value pairs you specify, logs of the container are collected.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>· Labels described in this topic refer to the label information in docker inspect.</td>
</tr>
<tr>
<td>Label Blacklist</td>
<td>No</td>
<td>LabelKey is required. If LabelValue is not empty, only containers whose labels contain LabelKey = LabelValue are excluded. If LabelValue is empty, all containers whose labels contain LabelKey are excluded.</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Note:</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>· Multiple key-value pairs are associated by the OR operator. When the label of a container contains any of the key-value pairs you specify, the container is excluded.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>· Labels described in this topic refer to the label information in docker inspect.</td>
</tr>
<tr>
<td>Parameter</td>
<td>Required</td>
<td>Description</td>
</tr>
<tr>
<td>---------------------------</td>
<td>----------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Environment Variable Whitelist</td>
<td>No</td>
<td>EnvKey is required. If EnvValue is not empty, only containers whose environment variables contain EnvKey = EnvValue are collected. If EnvValue is empty, all containers whose environment variables contain EnvKey are collected.</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Note:</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Multiple key-value pairs are associated by the OR operator. When the environment variable of a container contains any of the key-value pairs you specify, logs of the container are collected.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The environment variable refers to the environment information configured in container startup.</td>
</tr>
<tr>
<td>Environment Variable Blacklist</td>
<td>No</td>
<td>EnvKey is required. If EnvValue is not empty, only containers whose environment variables contain EnvKey = EnvValue are excluded. If EnvValue is empty, all containers whose environment variables contain EnvKey are excluded.</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Note:</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Multiple key-value pairs are associated by the OR operator. When the environment variable of a container contains any of the key-value pairs you specify, the container is excluded.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The environment variable refers to the environment information configured in container startup.</td>
</tr>
<tr>
<td>Other</td>
<td>N/A</td>
<td>For more information about other collection configuration methods and parameters, see <a href="#">Configuration procedure for text log collection</a>.</td>
</tr>
</tbody>
</table>

**Note:**
• Labels in whitelist and blacklist are different from those defined in Kubernetes. Labels in this topic refer to the label information in docker inspect.

• A namespace and a container name in Kubernetes can be mapped to Docker labels. LabelKey corresponding to a namespace is io.kubernetes.pod.namespace. LabelKey corresponding to a container name is io.kubernetes.container.name. For example, the namespace of the pod you created is backend-prod and the container name is worker-server. In this case, you can configure a whitelist label: io.kubernetes.pod.namespace : backend-prod or io.kubernetes.container.name : worker-server, so that only logs of the container are collected.

• In Kubernetes, we recommend that you only use the io.kubernetes.pod.namespace and io.kubernetes.container.name labels. In other cases, use a environment variable whitelist or blacklist.

6. Configure query and analysis.

The index is set by default. You can modify it based on your needs.

Configuration examples

• Environment configuration

Collect the logs of the container whose environment variable is NGINX_PORT _80_TCP_PORT=80 but not POD_NAMESPACE=kube-system. The log file path is /var/log/nginx/access.log and logs are parsed in simple mode.

Note:
The environment variable refers to the environment information configured in container startup.

Figure 3-17: Environment configuration example
This example uses the following method to configure a data source. For more information about other collection configuration methods and parameters, see *Configuration procedure for text log collection*.

**Figure 3-18: Data source configuration example**

<table>
<thead>
<tr>
<th>* Config Name:</th>
<th>docker-file</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Path:</td>
<td>/var/log/nginx  /ssl/ access.log</td>
</tr>
<tr>
<td>Docker File:</td>
<td>off</td>
</tr>
<tr>
<td>Label Whitelist:</td>
<td>LabelKey</td>
</tr>
<tr>
<td>Environment Variable Whitelist:</td>
<td>EnvKey</td>
</tr>
<tr>
<td>Environment Variable Blacklist:</td>
<td>EnvKey</td>
</tr>
<tr>
<td>Mode:</td>
<td>Simple Mode</td>
</tr>
</tbody>
</table>
Label configuration

Collect the logs of a container that meets the following conditions: the container label is `io.kubernetes.container.name=nginx`, the log file path is `/var/log/nginx/access.log`, and logs are parsed in simple mode.

Note:
Labels described in this topic refer to the label information in docker inspect. They are not synonymous with labels in Kubernetes.

Figure 3-19: Label configuration example

```
"Onbuild": null,
"Labels": {
  "io.kubernetes.container.hash": "53073f5a",
  "io.kubernetes.container.restartCount": "0",
  "io.kubernetes.container.terminationMessagePath": "/dev/termination-log",
  "io.kubernetes.container.terminationMessagePolicy": "File",
  "io.kubernetes.container.terminationGracePeriod": "30",
  "io.kubernetes.container.logpath": "/var/log/pods/od00b078-4182-11e8-8414-00163f006885/nginx_0.log",
  "io.kubernetes.container.name": "nginx",
  "io.kubernetes.container.uid": "86ce654874-r4mff",
  "io.kubernetes.container.namespace": "default",
  "io.kubernetes.container.id": "od00b078-4182-11e8-8414-00163f006885",
  "io.kubernetes.container.sandbox.id": "52164e930e8b701493c259db877fa0a37513be55a204a8d0b6891df06da123969",
  "maintainer": "NGINX Docker Maintainers <docker-maint@nginx.com>"
},
"StopSignal": "SIGTERM"
```
This example uses the following method to configure a data source. For more information about other collection configuration methods and parameters, see Configuration procedure for text log collection.

Figure 3-20: Data source configuration

Default fields

Each uploaded log of a common Docker container contains the following fields.

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>image_name</em></td>
<td>The name of the image.</td>
</tr>
<tr>
<td><em>container_name</em></td>
<td>The name of the container.</td>
</tr>
</tbody>
</table>
3.6.3 Container standard output

Logtail takes the standard output (stdout) streams for containers as input sources and uploads logs together with the container-related metadata information to Log Service.

Characteristics

- Supports the collection of stdout and standard error (stderr) logs.
- Supports the use of labels to specify the containers whose logs are to be collected.
- Supports the use of labels to exclude specific containers.
- Supports the use of environment variables to specify the containers whose logs are to be collected.
- Supports the use of environment variables to exclude specific containers.
- Supports multiline logs such as Java Stack logs.
- Supports automatic labeling of container data.
- Supports automatic labeling for Kubernetes containers.

How it works

As shown in the following figure, Logtail communicates with the domain socket of the Docker engine to query containers running on the Docker engine. Logtail also locates the containers whose logs are to be collected based on the configured label.
and environment information. Logtail then uses the docker logs command to obtain the logs of specified containers.

Logtail periodically saves the log file position information to the checkpoint file when collecting stdout logs of a container. If Logtail is restarted after being stopped, it collects logs from the last saved log file position.

Limits

- This feature only applies to Logtail version 0.16.0 and later running on Linux. For more information about Logtail versions and version upgrade, see Install Logtail in Linux.
- By default, Logtail accesses the Docker engine through /var/run/docker.sock. Make sure that the domain socket exists and you have access permissions.
- Multiline logs: To ensure that a multiline log is not split into multiple logs due to output latency, the last collected multiline log is cached for a period of time. The default period of time a multiline log is cached is 3 seconds. You can specify the time by configuring the BeginLineTimeoutMs parameter. The parameter value cannot be less than 1000 ms. Otherwise, a false positive or negative error may occur.
- Collection stop policy: When a container is stopped and Logtail detects the die event on the container, Logtail stops collecting stdout logs of the container. In this case, if a collection latency occurs, some stdout logs that are generated before the stop action may be lost.
- Docker log driver: Only log drivers of the JSON type are supported in the collection of stdout logs.
• Context: By default, a collection configuration is in the same context. If you need to configure different types of containers in different contexts, configure each type separately.

• Data processing: The default field for collected data is content. A common processing configuration can be used.

• Label: refers to the label information in docker inspect. It is not synonymous with labels in Kubernetes.

• Environment: refers to the environment information configured during container startup.

Configuration process

1. Deploy and configure the Logtail container.

2. Configure log collection in Log Service.

Logtail deployment and configuration

• Kubernetes

For more information about Kubernetes log collection, see Logtail deployment solution for collecting Kubernetes logs.

• Other container management methods

For more information about other container management methods such as Swarm and Mesos, see Common Logtail deployment solution for collecting Docker logs.

Data source configuration

1. Log on to the Log Service console, and click the name of the target project.

2. Click the Import Data button. On the Import Data page that appears, select Docker Standard Output.

3. Select a Logstore.

Select an existing Logstore, or create a new project or Logstore.

4. Create and configure a server group.

Before creating a server group, make sure that you have installed Logtail. After installing Logtail, click Complete Installation to create a server group. If you have created a server group, click Use Existing Server Groups.

Note:
The server that receives syslog logs must belong to a server group and must have Logtail version 0.16.13 or later installed.

Select a server group and move the server group from Source Server Groups to Applied Server Groups.

5. Configure a data source.

   In the Plug-in Config section, configure collection information. An example is as follows. For more information about configuration parameters, see Parameters.

   ```json
   {
   "inputs": [
   {
   "type": "service_docker_stdout",
   "detail": {
   "Stdout": true,
   "Stderr": true,
   "IncludeLabel": {
   "io.kubernetes.container.name": "nginx"
   },
   "ExcludeLabel": {
   "io.kubernetes.container.name": "nginx-ingress-controller"
   },
   "IncludeEnv": {
   "NGINX_SERVICE_PORT": "80"
   },
   "ExcludeEnv": {
   "POD_NAMESPACE": "kube-system"
   }
   }
   }
   ]
   }
   
6. Configure query and analysis.

   The index is set by default. You can modify it based on your needs.

Parameters

   The type of input sources is service_docker_stdout.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Type</th>
<th>Required</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IncludeLabel</td>
<td>Map. Key: string. Value: string</td>
<td>Yes</td>
<td>This parameter is empty by default. If the parameter is empty, logs of all containers are collected. If the key is not empty but the value is empty, logs of all containers whose labels include this key are collected.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Note:</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Key-value pairs are associated by the OR operator. When the label of a container contains any of the key-value pairs, logs of the container are collected.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Labels described in this topic refer to the label information in docker inspect.</td>
</tr>
<tr>
<td>ExcludeLabel</td>
<td>Map. Key: string. Value: string</td>
<td>No</td>
<td>This parameter is empty by default. If the parameter is empty, no containers are excluded. If the key is not empty but the value is empty, the containers whose labels include this key are excluded.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Note:</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Key-value pairs are associated by the OR operator. When the label of a container includes any of the key-value pairs, the container is excluded.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Labels described in this topic refer to the label information in docker inspect.</td>
</tr>
<tr>
<td>Parameter</td>
<td>Type</td>
<td>Required</td>
<td>Description</td>
</tr>
<tr>
<td>-----------</td>
<td>------------</td>
<td>----------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>IncludeEnv</td>
<td>Map. Key: string, Value: string</td>
<td>No</td>
<td>This parameter is empty by default. If the parameter is empty, logs of all containers are collected. If the key is not empty but the value is empty, logs of the containers whose environment variables include this key are collected.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Note:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>·  Key-value pairs are associated by the OR operator. When the environment variable of a container includes any of the key-value pairs, logs of the container are collected.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>·  The environment variable refers to the environment information configured in container startup.</td>
</tr>
<tr>
<td>ExcludeEnv</td>
<td>Map. Key: string, Value: string</td>
<td>No</td>
<td>This parameter is empty by default. If the parameter is empty, no containers are excluded. If the key is not empty but the value is empty, the containers whose environment variables include this key are excluded.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Note:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>·  Key-value pairs are associated by the OR operator. When the environment variable of a container includes any of the key-value pairs, the container is excluded.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>·  The environment variable refers to the environment information configured in container startup.</td>
</tr>
<tr>
<td>Stdout</td>
<td>Boolean</td>
<td>No</td>
<td>Default value: true. If the parameter is set to false, stdout data is not collected.</td>
</tr>
<tr>
<td>Stderr</td>
<td>Boolean</td>
<td>No</td>
<td>Default value: true. If the parameter is set to false, stderr data is not collected.</td>
</tr>
</tbody>
</table>
### Log Service

#### Data Collection / 3 Logtail collection

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Type</th>
<th>Required</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BeginLineRegEx</td>
<td>String</td>
<td>No</td>
<td>This parameter is empty by default. If the parameter is not empty, the regular expression that is used to match the first line of each log is applied. If a line matches this regular expression, this line is assumed as the start of a new log. Otherwise, this line is assumed as part of the previous log.</td>
</tr>
<tr>
<td>BeginLineTimeoutMs</td>
<td>Integer</td>
<td>No</td>
<td>The timeout period for the regular expression to match a line. Default value: 3000. Unit: ms. If no new log appears within 3 seconds, the last log is uploaded.</td>
</tr>
<tr>
<td>BeginLineCheckLength</td>
<td>Integer</td>
<td>No</td>
<td>The length of data that matches the regular expression. Default value: $10 \times 1024$. Unit: bytes. You can set this parameter to check whether the beginning part of a line can match the regular expression. This improves matching efficiency.</td>
</tr>
<tr>
<td>MaxLogSize</td>
<td>Integer</td>
<td>No</td>
<td>The maximum log length. Default value: $512 \times 1024$. Unit: bytes. If the log length exceeds this value, the log data is uploaded directly without finding the first line of logs.</td>
</tr>
</tbody>
</table>

**Note:**

- Labels in IncludeLabel and ExcludeLabel are different from those defined in Kubernetes. Labels in this topic refer to the label information in docker inspect.
- A namespace and a container name in Kubernetes can be mapped to Docker labels. The LabelKey parameter corresponding to a namespace is `io.kubernetes.pod.namespace`. The LabelKey parameter corresponding to a container name is `io.kubernetes.container.name`. For example, the namespace of the pod you created is backend-prod and the container name is worker-server. In this case, you can configure a whitelist label: `io.kubernetes.pod.namespace : backend-prod` or `io.kubernetes.container.name : worker-server`, so that only logs of the container are collected.
In Kubernetes, we recommend that you only use the `io.kubernetes.pod.namespace` and `io.kubernetes.container.name` labels. In other cases, use `IncludeEnv` or `ExcludeEnv`.

### Default fields

- **Common Docker containers**
  Each uploaded log contains the following fields.

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>_time_</code></td>
<td>The data upload time. Example: <code>2018-02-02T02:18:41.979147844Z</code></td>
</tr>
<tr>
<td><code>_source_</code></td>
<td>The type of input sources. Valid values: stdout and stderr.</td>
</tr>
<tr>
<td><code>_image_name_</code></td>
<td>The name of the image.</td>
</tr>
<tr>
<td><code>_container_name_</code></td>
<td>The name of the container.</td>
</tr>
<tr>
<td><code>_container_ip_</code></td>
<td>The IP address of the container.</td>
</tr>
</tbody>
</table>

- **Kubernetes containers**
  Each uploaded log contains the following fields.

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>_time_</code></td>
<td>The data upload time. Example: <code>2018-02-02T02:18:41.979147844Z</code></td>
</tr>
<tr>
<td><code>_source_</code></td>
<td>The type of input sources. Valid values: stdout and stderr.</td>
</tr>
<tr>
<td><code>_image_name_</code></td>
<td>The name of the image.</td>
</tr>
<tr>
<td><code>_container_name_</code></td>
<td>The name of the container.</td>
</tr>
<tr>
<td><code>_pod_name_</code></td>
<td>The name of the pod.</td>
</tr>
<tr>
<td><code>_namespace_</code></td>
<td>The namespace where the pod is located.</td>
</tr>
<tr>
<td><code>_pod_uid_</code></td>
<td>The unique identifier of the pod.</td>
</tr>
<tr>
<td><code>_container_id_</code></td>
<td>The IP address of the pod.</td>
</tr>
</tbody>
</table>
Common configuration examples

- **Environment configuration**

  Collect the logs of the container whose environment variable is `NGINX_PORT_80_TCP_PORT=80` but not `POD_NAMESPACE=kube-system`.

  **Note:**
  
  The environment variable refers to the environment information configured in container startup.

  **Figure 3-21: Environment configuration example**

  ```json
  "inputs": [
    {
      "type": "service_docker_stdout",
      "detail": {
        "Stdout": true,
        "Stderr": true,
        "IncludeEnv": {
          "NGINX_PORT_80_TCP_PORT": "80"
        }
      }
    }
  ]
  ```

- **Collection configuration**

  ```json
  {
    "inputs": [
      {
        "type": "service_docker_stdout",
        "detail": {
          "Stdout": true,
          "Stderr": true,
          "IncludeEnv": {
            "NGINX_PORT_80_TCP_PORT": "80"
          }
        }
      }
    ]
  }
  ```
Label configuration

Collect the stdout and stderr logs of the container whose label is `io.kubernetes.container.name=nginx` but not type=pre.

Note:
Labels described in this topic refer to the label information in docker inspect. They are not synonymous with labels in Kubernetes.
Example of configuring multiline log collection

Configuring multiline log collection is important for the collection of Java exception stack logs. The following section introduces a standard collection configuration for Java stdout logs.

• Sample log

```plaintext
2018-02-03 14:18:41.968  INFO [spring-cloud-monitor] [nio-8080-exec-4] c.g.s.web.controller.DemoController : service start
  at org.apache.catalina.core.ApplicationFilterChain.internalDoFilter(ApplicationFilterChain.java:193)
  at org.apache.catalina.core.ApplicationFilterChain.doFilter(ApplicationFilterChain.java:166)
  at org.apache.catalina.core.StandardWrapperValve.invoke(StandardWrapperValve.java:199)
  at org.apache.catalina.core.StandardContextValve.invoke(StandardContextValve.java:96)
  ...
2018-02-03 14:18:41.968  INFO [spring-cloud-monitor] [nio-8080-exec-4] c.g.s.web.controller.DemoController : service start done
```

• Collection configuration

Collect logs of the container whose label is `app=monitor`. The first line of each log to be collected is of the date type. To improve matching efficiency, only the first 10 bytes of each line are checked.

```json
{
  "inputs": [
    {
      "detail": {
        "BeginLineCheckLength": 10,
        "BeginLineRegex": "\d+-\d+-\d+. *",
        "IncludeLabel": {
          "app": "monitor"
        }
      },
      "type": "service.docker.stdout"
    }
  ]
}
```

Processing of collected data

We recommend that you use the multiline log format and a regular expression to parse logs into the following fields: time, module, thread, class, and info.
• Collection configuration

Collect logs of the container whose label is `app=monitor`. The first line of each log to be collected is of the date type. To improve matching efficiency, only the first 10 bytes of each line are checked.

```json
{
  "inputs": [
    {
      "detail": {
        "BeginLineCheckLength": 10,
        "BeginLineRegex": "[\d-]+[\d-]+.*",
        "IncludeLabel": {
          "app": "monitor"
        }
      },
      "type": "service_docker_stdout"
    }
  ],
  "processors": [
    {
      "type": "processor_regex",
      "detail": {
        "SourceKey": "content",
        "Regex": "([\d-]+[\d-]+[\d-]+[\d-]+[\d-]+[\d-]+\d+([\w\s\d\-]+))[\s\d\S]+([\s\d\S]+)",
        "Keys": [
          "time",
          "module",
          "thread",
          "class",
          "info"
        ],
        "NoKeyError": true,
        "NoMatchError": true,
        "KeepSource": false
      }
    }
  ]
}
```

• Sample output

After the 2018-02-03 14:18:41.968 INFO [spring-cloud-monitor] [nio-8080-exec-4] c.g.s.web.controller.DemoController : service start done log is processed, the following output is returned:

```
__tag__:__hostname__:logtail-dfgef
__container_name__:monitor
__image_name__:registry.cn-hangzhou.aliyuncs.xxxxxxxxxxxxxxxx
__namespace__:default
__pod_name__:monitor-6f54bd5d74-rtzc7
__pod_uid__:7f012b72-04c7-11e8-84aa-00163f00c369
__source__:stdout
__time__:2018-02-02T14:18:41.979147844Z
time:2018-02-02 02:18:41.968
level:INFO
module:spring-cloud-monitor
```

3.6.4 Configure CRDs to collect Kubernetes logs

To collect Kubernetes logs, you can complete the relevant configurations in the Log Service console. For Kubernetes microservices, you can also use custom resource definitions (CRDs) for log collection. To configure CRDs, you can use the kubectl command-line tool.

We recommend that you configure CRDs for log collection. CRDs are fully integrated with Kubernetes in application deployment and publishing.

Procedure

You can run the installation command to install the alibaba-log-controller Helm package. During the installation process, the following operations are performed:

1. Create a CRD named aliyunlogconfigs.
2. Create a Deployment controller named alibaba-log-controller.
3. Deploy Logtail as a DaemonSet.

The log collection procedure is described as follows:

1. You use kubectl or other tools to create the aliyunlogconfigs CRD.
2. The alibaba-log-controller Deployment controller detects the creation of the CRD.

3. The alibaba-log-controller Deployment controller sends requests to Log Service for creating Logstores, applying a server group, and configuring Logtail based on the CRD.

4. In DaemonSet mode, Logtail requests the configuration server for new or updated configurations at regular intervals, and then validates the configurations in the hot reload mode.

5. Logtail collects standard outputs from each pod based on updated configurations.


Configure CRDs

**Note:**
If you have deployed Logtail as a DaemonSet, you cannot use CRDs to modify or delete Logtail settings. You can switch the configuration method from DaemonSets to CRDs. For more information, see the Kubernetes log collection process topic.

You can define the AliyunLogConfig CRD to create Log Service resources and complete the log collection configurations. If you delete the CRD, the corresponding resources and log collection configurations are also deleted. CRD format:

```yaml
apiVersion: log.alibabacloud.com/v1alpha1      ## The default setting that you do not need to modify.
kind: AliyunLogConfig                          ## The default setting that you do not need to modify.
metadata:
  name: simple-stdout-example                  ## The resource name, which must be unique in the cluster.
spec:
  project: k8s-my-project                      ## Optional. The project name. You can specify a project that is not occupied. Otherwise, the project created when you installed the Helm package is selected by default.
  logstore: k8s-stdout                         ## The Logstore name. The specified Logstore is automatically created if it does not exist.
  shardCount: 2                                ## Optional. The number of Logstore shards. Valid values: 1 to 10. Default value: 2.
  lifeCycle: 90                                ## Optional. The validity period of log data in the Logstore. Valid values: 1 to 7300. Default value: 90. The value of 7300 indicates that log data is permanently stored in the Logstore.
  logtailConfig:                               ## The Logtail settings.
```

---

**Issue:** 20200323
After you define the CRD and apply the specified settings, a Deployment controller named alibaba-log-controller is automatically created.

For more information about logtailConfig, see #unique_108.

View the log collection settings

You can view the log collection settings by using Kubernetes CRDs or the Log Service console.

For more information about how to view the log collection settings in the console, see Manage collection configurations.

Note:
If you use CRDs and you modify the Logtail configurations in the console, the modifications will be overwritten after you update the CRDs.

• Use the kubectl get aliyunlogconfigs command to view all Logtail configuration sets.

<table>
<thead>
<tr>
<th>NAME</th>
<th>AGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>regex-file-example</td>
<td>10s</td>
</tr>
<tr>
<td>regex-stdout-example</td>
<td>4h</td>
</tr>
<tr>
<td>simple-file-example</td>
<td>5s</td>
</tr>
</tbody>
</table>

• Use the kubectl get aliyunlogconfigs ${config_name} -o yaml command to view the detailed settings and status of a Logtail configuration set.

The status field indicates whether the configuration set is applied. If the configuration set is applied, the statusCode parameter in the status field is 200. If the statusCode parameter is not 200, the configuration set fails to be applied.

```
[root@iZbp1dsbiaZ ~]# kubectl get aliyunlogconfigs simple-file-example -o yaml
apiVersion: log.alibabacloud.com/v1alpha1
kind: AliyunLogConfig
metadata:
  annotations:
    kubectl.kubernetes.io/last-applied-configuration: |
    
```
To configure container standard outputs, set `inputType` to `plugin` and enter the details in the `plugin` field under `inputDetail`. For more information, see Container standard output.

- Collect logs in the simple mode

The following example shows how to collect standard outputs and errors (stdout and stderr) of all containers. The containers whose environment variable contains the `COLLECT_STDOUT_FLAG=false` key are excluded.

```yaml
apiVersion: log.alibabacloud.com/v1alpha1
kind: AliyunLogConfig
metadata:
  # The resource name, which must be unique in the cluster.
  name: simple-stdout-example
spec:
  # The name of the Logstore to which logs are uploaded.
  logstore: k8s-stdout
  # The detailed Logtail settings.
  logtailConfig:
    # The input type.
    inputType: plugin
```
# The name of the Logtail configuration set. This name must be the same as the resource name.
configName: simple-stdout-example

inputDetail:
  plugin:
    inputs:
      -
        # The input type.
        type: service_docker_stdout
detail:
          # Collect standard outputs and errors.
          Stdout: true
          Stderr: true
          # Collect standard outputs and errors of all containers. The containers whose environment variable contains the COLLECT_STDOUT_FLAG: false key are excluded.
          ExcludeEnv:
            COLLECT_STDOUT_FLAG: "false"

• Collect logs in the custom mode

The following example shows how to collect access logs of Grafana and parse the access logs into structured data.

The Grafana container has the environment variable GF_INSTALL_PLUGINS=grafana-piechart-...... You can set IncludeEnv to GF_INSTALL_PLUGINS: '' to enable Logtail to collect standard outputs only from this container.

Figure 3-24: Collect logs in the custom mode

The format of Grafana access logs is as follows:

t=2018-03-09T07:14:03+0000 lvl=info msg="Request Completed" logger=context userId=0 orgId=0 uname= method=GET path=/ status=302 remote_addr=172.16.64.154 time_ms=0 size=29 referer=

To use regular expressions to parse access logs, use the following configurations:

apiVersion: log.alibabacloud.com/v1alpha1
kind: AliyunLogConfig
metadata:
    # The resource name, which must be unique in the cluster.
    name: regex-stdout-example
spec:
# The name of the Logstore to which logs are uploaded.
logstore: k8s-stdout-regex

# The detailed Logtail settings.
logtailConfig:
  # The input type.
inputType: plugin
  # The name of the Logtail configuration set, which must be the
same as the resource name.
configName: regex-stdout-example
  inputDetail:
    plugin:
      inputs:
        # The input type.
        type: service_docker_stdout
        detail:
          # Collect standard outputs, but do not collect
          standard errors.
          Stdout: true
          Stderr: false
          # Collect standard outputs of the containers whose
          environment variable contains the GF_INSTALL_PLUGINS key.
          IncludeEnv:
            GF_INSTALL_PLUGINS: ''
          processors:
            # Use a regular expression.
            type: processor_regex
            detail:
              # The key of the data collected from the docker is
              content by default.
              SourceKey: content
              # Use a regular expression to extract fields.
              Regex: 't=(\d+-\d+\-\w+:\d+:\d+:\d+) lvl=(\w+) msg
             ="([^"\]+)" logger=(\w+) userId=(\w+) orgId=(\w+) method
              =((\w+) path=((\S+) status=((\d+) remote_addr=((\S+) time_ms=((\d+) size
              =((\d+) referer=((\S*) .*'
              # The extracted keys.
              Keys: ['time', 'level', 'message', 'logger', 'userId
            ', 'orgId', 'uname', 'method', 'path', 'status', 'remote_addr', 'time_ms', 'size', 'referer']
              # Remain original fields.
              KeepSource: true
              NoKeyError: true
After the configurations are applied, the data collected by Log Service is as follows:

Figure 3-25: Collected log data

Configuration example of container files

- **Collect logs in the simple mode**

The following example shows how to collect log files from containers whose environment variable contains the `ALIYUN_LOGTAIL_USER_DEFINED_ID` key. The log path is `/data/logs/app_1` and the file name is `simple.LOG`.

```yaml
apiVersion: log.alibabacloud.com/v1alpha1
kind: AliyunLogConfig
metadata:
    # The resource name, which must be unique in the cluster.
    name: simple-file-example
spec:
    # The name of the Logstore to which logs are uploaded.
    logstore: k8s-file
    # The detailed Logtail settings.
    logtailConfig:
        # The input type.
        inputType: file
        # The name of the Logtail configuration set. This name must be the same as the resource name.
        configName: simple-file-example
        inputDetail:
            # Set logType to common_reg_log.
            logType: common_reg_log
            # The log path.
            logPath: /data/logs/app_1
            # The file name. The name can contain wildcards, for example, log_*.log.
            filePattern: simple.LOG
            # To collect files from the container, set dockerFile to true.
            dockerFile: true
            # Only collect logs from containers whose environment variable contains the ALIYUN_LOGTAIL_USER_DEFINED_ID key.
            dockerIncludeEnv:
```
• Use regular expressions to collect logs

The following example shows a log entry that is generated by a Java program:

```
[2018-05-11T20:10:16,000] [INFO] [SessionTracker] [SessionTrackerImpl.java:148] Expiring sessions
java.sql.SQLException: Incorrect string value: '\'xF0\x9F\x8E\x8F ','...' for column 'data' at row 1
at org.springframework.jdbc.support.AbstractFallbackSQLExceptionTranslator.translate(AbstractFallbackSQLExceptionTranslator.java:84)
at org.springframework.jdbc.support.AbstractFallbackSQLExceptionTranslator.translate(AbstractFallbackSQLExceptionTranslator.java:84)
```

A log entry may be divided into multiple lines because the log entry contains error stack traces. Therefore, you must specify a regular expression to match the header of the log entry. To extract fields by using a regular expression, you can specify the configurations as follows:

```yaml
apiVersion: log.alibabacloud.com/v1alpha1
custom:
  kind: AliyunLogConfig
metadata:
  # The resource name, which must be unique in the cluster.
  name: regex-file-example
spec:
  # The name of the Logstore to which logs are uploaded.
  logstore: k8s-file
  logtailConfig:
    # The input type.
    inputType: file
    # The name of the Logtail configuration set. This name must be the same as the resource name.
    configName: regex-file-example
    inputDetail:
      # Set logType to common_reg_log.
      logType: common_reg_log
      # The log path.
      logPath: /app/logs
      # The file name. The name can contain wildcards, for example, log_*.
      filePattern: error.LOG
      # The regular expression that matches log headers.
      logBeginRegex: '\[\d+-\d+-\w+:\d+:\d+:\d+]?\s\[\w+\]\s.*'
      # The regular expression that parses logs.
      regex: '[[^\[]][^\[]\s\[(\w+)]\s\[(\w+)]\s\[(^[:]+):(\d+)]\s(.*)'
      
      # The extracted keys.
      key: ["time", "level", "method", "file", "line", "message"]
      # The time key is optional. If you configure this key, the value of the time field in logs is extracted as the key value.
      timeFormat: '%Y-%m-%dT%H:%M:%S'
      # To collect files from the container, set dockerFile to true.
      dockerFile: true
      # Collect logs from containers whose environment variable contains the ALIYUN_LOGTAIL_USER_DEFINED_ID key.
      dockerIncludeEnv:
```
After the configurations are applied, the data collected by Log Service is as follows:

Figure 3-26: Collected log data

- **Collect DSV formatted logs**

To use Logtail to parse delimiter-separated value (DSV) formatted logs, use the following configurations:

```yaml
apiVersion: log.alibabacloud.com/v1alpha1
class: AliyunLogConfig

metadata:
  # The resource name, which must be unique in the cluster.
  name: delimiter-file-example

spec:
  # The name of the Logstore to which logs are uploaded.
  logstore: k8s-file
  logtailConfig:
    # The input type.
    inputType: file
    configName: delimiter-file-example
    # The name of the Logtail configuration set. This name must be the same as the resource name.
    inputDetail:
      # Set logType to delimiter_log.
      logType: delimiter_log
      # The log path.
      logPath: /usr/local/ilogtail
      # The file name. The name can contain wildcards, for example, log_*,.log
      filePattern: delimiter_log.LOG
      # Use multi-character delimiters.
      separator: '|&|
      # The extracted keys.
      key: ['time', 'level', 'method', 'file', 'line', 'message']
      # Optional. The key used to parse time.
      timeKey: 'time'
      # Optional. The method used to parse time.
      timeFormat: '%Y-%m-%dT%H:%M:%S'
      # To collect files from the container, set dockerFile to true.
      dockerFile: true
      # Collect logs from containers whose environment variable contains the ALIYUN_LOGTAIL_USER_DEFINED_ID key.
      dockerIncludeEnv:
```

Issue: 20200323
• Collect JSON logs

If each line of log data in a file is a JSON object, you can parse the file as follows:

```yaml
apiVersion: log.alibabacloud.com/v1alpha1
kind: AliyunLogConfig
metadata:
  # The resource name, which must be unique in the cluster.
  name: json-file-example
spec:
  # The name of the Logstore to which logs are uploaded.
  logstore: k8s-file
  logtailConfig:
    # The input type.
    inputType: file
    # The name of the Logtail configuration set. This name must be the same as the resource name.
    configName: json-file-example
    inputDetail:
      # Set logType to json_log.
      logType: json_log
      # The log path.
      logPath: /usr/local/ilogtail
      # The file name. The name can contain wildcards, for example, log_*.log
      filePattern: json_log.LOG
      # Optional. The key used to parse time.
      timeKey: 'time'
      # Optional. The method used to parse time.
      timeFormat: '%Y-%m-%dT%H:%M:%S'
      # To collect files from the container, set dockerFile to true.
      dockerFile: true
      # Collect logs from containers whose environment variable contains the ALIYUN_LOGTAIL_USER_DEFINED_ID key.
      dockerIncludeEnv:
        ALIYUN_LOGTAIL_USER_DEFINED_ID: ''
```

3.6.5 Collect Kubernetes logs in sidecar mode

Logtail of Log Service can collect logs from Kubernetes clusters in sidecar mode. A sidecar container is created for each business container from which you want to collect logs, achieving multi-tenant isolation and high collection performance.

Context

By default, Logtail installed in Kubernetes clusters run in DaemonSet mode. This mode features simple O&M, low resource usage, and flexible configuration, and can be used to collect container standard output (stdout) logs and container files.

However, in DaemonSet mode, Logtail needs to collect logs from all containers on a node. This mode has performance bottlenecks and causes a loosely isolated environment for business logs. To address this problem, Logtail supports the sidecar mode. In this mode, a sidecar container is created for each business
container you want to collect logs from. This mode provides multi-tenant isolation and excellent performance. We recommend that you use this mode for large Kubernetes clusters or clusters that serve multiple business parties as PaaS platforms.

Characteristics

- Supports Alibaba Cloud Container Service for Kubernetes, ECS-based user-created Kubernetes clusters, and IDC-based user-created Kubernetes clusters.
- Supports the collection of Pod metadata, including the Pod name, Pod IP address, Pod namespace, and the name and IP address of the node where the Pod is located.
- Supports automatic creation of Log Service resources by using a Custom Resource Definition (CRD). The resources include projects, Logstores, indexes, Logtail configurations, and server groups.
- Supports dynamic resizing. The number of replicas can be changed at any time, and changes take effect immediately.

How it works

In log collection using the sidecar mode, Logtail shares the log directory with business containers. Business containers write logs to the shared directory. Logtail monitors the changes of log files in the shared directory and collects logs. For more information, see the official community documentation:

- Log collection in sidecar mode
- Sidecar mode example

Prerequisites

- Log Service is activated. 
  
  *Activate Log Service* if it has not been activated.
- A CRD is created if you need to use it for log collection configuration. For more information, see *Kubernetes log collection process*.

Limits

- Logtail must share the log directory with business containers.
- The sidecar mode does not support the collection of container stdout logs.
Sidecar configuration

The sidecar configuration involves:

- **Configuration 1**: Set basic running parameters
- **Configuration 2**: Set the mount point
- **Configuration 3**: Set the stop latency

An example of sidecar mode configuration is as follows:

```yaml
apiVersion: batch/v1
kind: Job
metadata:
  name: nginx-log-sidecar-demo
  namespace: default
spec:
template:
  metadata:
    name: nginx-log-sidecar-demo
  spec:
    restartPolicy: Never
    containers:
    - name: nginx-log-demo
      image: registry.cn-hangzhou.aliyuncs.com/log-service/docker-log-test:latest
      command: ["/bin/mock_log"]
      args: ["--log-type=nginx", "--stdout=false", "--stderr=true", "--path=/var/log/nginx/access.log", "--total-count=1000000000", "--logs-per-sec=100"]
      volumeMounts:
      - name: nginx-log
        mountPath: /var/log/nginx
    # logtail sidecar container
    - name: logtail
      image: registry.cn-hangzhou.aliyuncs.com/log-service/logtail:latest
      command:
      - sh
      - -c
      - /usr/local/ilogtail/run_logtail.sh 10
      livenessProbe:
        exec:
          command:
          - /etc/init.d/ilogtaild
          - status
        initialDelaySeconds: 30
        periodSeconds: 30
      resources:
        limits:
          memory: 512Mi
        requests:
          cpu: 10m
          memory: 30Mi
      env:
      - name: logtail
        # more info: https://cr.console.aliyun.com/repository/cn-hangzhou/log-service/logtail/detail
        # this images is released for every region
        image: registry.cn-hangzhou.aliyuncs.com/log-service/logtail:latest
        # when recevie sigterm, logtail will delay 10 seconds and then stop
        command:
        - sh
        - -c
        - /usr/local/ilogtail/run_logtail.sh 10
```

---

**Issue**: 20200323
Configuration 1: Set basic running parameters

The following configuration is included:

```yaml
### base config
# user id
- name: "ALIYUN_LOGTAIL_USER_ID"
  value: "${your_aliyun_user_id}"
# user defined id
- name: "ALIYUN_LOGTAIL_USER_DEFINED_ID"
  value: "${your_machine_group_user_defined_id}"
# config file path in logtail's container
- name: "ALIYUN_LOGTAIL_CONFIG"
  value: "/etc/ilogtail/conf/${your_region_config}/ilogtail_config.json"
### env tags config
- name: "ALIYUN_LOG_ENV_TAGS"
  value: "_pod_name_|_pod_ip_|_namespace_|_node_name_|_node_ip_"
  valueFrom:
    fieldRef:
      fieldPath: metadata.name
- name: "_pod_name_"
  valueFrom:
    fieldRef:
      fieldPath: status.podIP
- name: "_namespace_"
  valueFrom:
    fieldRef:
      fieldPath: metadata.namespace
- name: "_node_name_"
  valueFrom:
    fieldRef:
      fieldPath: spec.nodeName
- name: "_node_ip_"
  valueFrom:
    fieldRef:
      fieldPath: status.hostIP
volumeMounts:
- name: nginx-log
  mountPath: /var/log/nginx
### share this volume
volumes:
- name: nginx-log
  emptyDir: {}
```

Issue: 20200323
value: "/etc/ilogtail/conf/${your_region_config}/ilogtail_config.json"

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
</table>
| ${your_region_config}        | The region where the Log Service project is located. Make sure that you use the correct value format based on the network type. Value formats:  
  · For the Internet, specify the region in the region-internet format. For example, the value for the China (Hangzhou) region is cn-hangzhou-internet.  
  · For the Alibaba Cloud internal network, specify the region in the region format. For example, the value for the China (Hangzhou) region is cn-hangzhou.  
  For more information about specifying a region based on the involved project, see Table 3-3: Region names for Logtail installation. |
| ${your_aliyun_user_id}       | The user ID. Set this parameter to the ID of your Alibaba Cloud account, which is in string format. For more information about how to view the ID, see Configure AliUids for ECS servers under other Alibaba Cloud accounts or on-premises IDCs.          |
| ${your_machine_group_user_defined_id} | The custom identifier of your server group. The ID must be unique in the region where your Log Service is located. For more information, see #unique_110.                                                                 |

Configuration 2: Set the mount point

  · Make sure that the Logtail container and business containers are mounted to the same directory.  
  · We recommend that you use the emptyDir method.  

A mount point is described in the preceding example of sidecar mode configuration.

Configuration 3: Set the stop latency

The following example sets the stop latency to 10 seconds. After receiving an external stop signal, the Logtail container waits for 10 seconds before it stops to prevent incomplete data collection.

command:
Log collection configuration

You can use the CRD mode or the Log Service console to configure log collection. The CRD mode can automatically create resources such as projects, Logstores, indexes, server groups, and collection configurations, and is deeply integrated with Kubernetes. We recommend that you use the CRD mode. The console mode is easier than CRD, and is suitable for users who debug or use Kubernetes log collection for the first time.

CRD-based configuration

For more information about CRD-based configuration, see Configure CRDs to collect Kubernetes logs. Compared with the DaemonSet collection mode, the CRD mode is subject to the following limits:

- You must specify the name of the project that requires log collection. Otherwise, logs are collected and sent to the project where Logtail is installed.
- You must specify the server group to which the configuration is applied. Otherwise, the configuration is applied to the server group to which the DaemonSet belongs.
- The sidecar mode supports only file collection. In file collection, dockerFile must be set to false.

For more information, see Examples.

Console-based configuration

Configure a server group.

In the Log Service console, create a Logtail server group, and set the server group identifier to a custom identifier to dynamically adapt to changes of the Pod IP address. To do so, perform the following steps:

1. Activate Log Service and create a project and a Logstore. For more information, see #unique_93.

2. On the Server Groups page in the Log Service console, click the icon next to Server Groups and select Create Server Group.
3. Select Custom ID from the Identifier drop-down list. In the Custom Identifier field, enter the value of `ALIYUN_LOGTAIL_USER_DEFINED_ID` set in the preceding step of setting basic running parameters.

Specify a collection mode.

After a server group is created, you can specify a collection mode for the corresponding files. Logtail supports various modes, including the simple, NGINX access log, delimiter log, JSON log, and regular expression log modes. For more information, see [Configuration procedure for text log collection](#).

The following figure shows an example configuration.

![Create Server Group](image)

Note:

You must turn off Docker File.
Examples

Scenario

- The Kubernetes cluster is an on-premises cluster in an IDC, and the region where Log Service resides is China (Hangzhou). Logs are collected from the Internet.
- In the following examples, the name of the volume to mount is `nginx-log` and the type of the volume is `emptyDir`. The volume is mounted to the `/var/log/nginx` directory in the nginx-log-demo and Logtail containers.
- Set the file path for access logs to be collected to `/var/log/nginx/access.log` and the destination Logstore to `nginx-access`.
- Set the file path for error logs to be collected to `/var/log/nginx/error.log` and the destination Logstore to `nginx-error`.
- Sidecar-based configuration

```
apiVersion: batch/v1
```
kind: Job
metadata:
  name: nginx-log-sidecar-demo
namespace: default
spec:
template:
  metadata:
    name: nginx-log-sidecar-demo
spec:
  restartPolicy: Never
  containers:
  - name: nginx-log-demo
    image: registry.cn-hangzhou.aliyuncs.com/log-service/docker-log-test:latest
    command: ["/bin/mock_log"]
    args: ["--log-type=nginx", "--stdout=false", "--stderr=true", "--path=/var/log/nginx/access.log", "--total-count=1000000000", "--logs-per-sec=100"]
    volumeMounts:
    - name: nginx-log
      mountPath: /var/log/nginx
  - name: logtail
    # more info: https://cr.console.aliyun.com/repository/cn-hangzhou/log-service/logtail/detail
    # this images is released for every region
    image: registry.cn-hangzhou.aliyuncs.com/log-service/logtail:latest
    command:
    - sh
    - -c
      - /usr/local/ilogtail/run_logtail.sh 10
    livenessProbe:
    exec:
      command:
      - /etc/init.d/ilogtailld
      - status
      initialDelaySeconds: 30
      periodSeconds: 30
    env:
    - name: "ALIYUN_LOGTAIL_USER_ID"
      value: "xxxxxxxxxx"
    - name: "ALIYUN_LOGTAIL_USER_DEFINED_ID"
      value: "nginx-log-sidecar"
    - name: "ALIYUN_LOGTAIL_CONFIG"
      value: "/etc/ilogtail/conf/cn-hangzhou-internet/ilogtail_config.json"
    - name: "ALIYUN_LOG_ENV_TAGS"
      value: "_pod_name_|_pod_ip_|_namespace_|_node_name_|_node_ip_"
    - name: "_pod_name_"
      valueFrom:
        fieldRef:
          fieldPath: metadata.name
    - name: "_pod_ip_"
      valueFrom:
        fieldRef:
fieldPath: status.podIP
- name: "_namespace_
  valueFrom:
    fieldRef:
      fieldPath: metadata.namespace
- name: "_node_name_
  valueFrom:
    fieldRef:
      fieldPath: spec.nodeName
- name: "_node_ip_
  valueFrom:
    fieldRef:
      fieldPath: status.hostIP

volumeMounts:
- name: nginx-log
  mountPath: /var/log/nginx
  ### share this volume
volumes:
- name: nginx-log
  emptyDir: {}

• CRD-based configuration

```yaml
# config for access log
apiVersion: log.alibabacloud.com/v1alpha1
kind: AliyunLogConfig
metadata:
  # your config name, must be unique in you k8s cluster
  name: nginx-log-access-example
spec:
  # project name to upload log
  project: k8s-nginx-sidecar-demo
  # logstore name to upload log
  logstore: nginx-access
  # machine group list to apply config, should be same with your sidecar' [ALIYUN_LOGTAIL_USER_DEFINED_ID]
  machineGroups:
  - nginx-log-sidecar

  # logtail config detail
  logtailConfig:
    # log file's input type is 'file'
    inputType: file
    # logtail config name, should be same with [metadata.name]
    configName: nginx-log-access-example
    inputDetail:
      # Simple mode with logType set to "common_reg_log"
      logType: common_reg_log
      # Log folder
      logPath: /var/log/nginx
      # File name which can include wildcards, such as log_*.log
      filePattern: access.log
      # Sidecar mode with dockerFile set to false
      dockerFile: false
      # Regular expression to match the first line of each log. For single-line mode, the value is set to .*
      logBeginRegex: '.*
      # Regular expression for parsing logs
      regex: '([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)\s([^\s]+)
      # Extracted keys
```
Query of log collection results

After the configuration in the preceding example is applied to the Kubernetes cluster, the Logtail container automatically creates the corresponding resources such as the project, Logstore, server group, and Logtail configuration, and automatically collects the generated logs to Log Service. You can log on to the Log Service console to view collected logs.
3.6.6 Collect standard Docker logs

This topic describes how to use Logtail to collect standard Docker logs and upload these logs together with the container-related metadata to Log Service.

Procedure

1. **Deploy a Logtail container.**

2. **Configure a Logtail server group.**

   Create a server group with a custom ID in the Log Service console. The container cluster can automatically scale up or down based on data traffic.

3. **Create a collection configuration.**

   Create collection configurations in the Log Service console. All the collection-related parameters are configured on the server side. No local configuration is needed.
Deploy a Logtail container

1. Run the following command to pull the Logtail image:

   ```bash
docker pull registry.cn-hangzhou.aliyuncs.com/log-service/logtail
   ```

2. Start the Logtail container.

   Set the $\{your\_region\_name\}$, $\{your\_aliyun\_user\_id\}$, and $\{your\_machine\_group\_user\_defined\_id\}$ parameters in the startup template.

   ```bash
docker run -d -v /:/logtail_host:ro -v /var/run:/var/run --env
   ALIYUN_LOGTAIL_CONFIG=/etc/logtail/conf/$\{your\_region\_name\}/
   logtail_config.json
   --env ALIYUN_LOGTAIL_USER_ID=$\{your\_aliyun\_user\_id\} --env
   ALIYUN_LOGTAIL_USER_DEFINED_ID=$\{your\_machine\_group\_user\_defined\_id\}
   registry.cn-hangzhou.aliyuncs.com/log-service/logtail
   ```

   **Note:**

   Before you set the parameters, you must complete either of the following configurations. Otherwise, the container text file busy error may occur when you delete another container.

   - For CentOS 7.4 and later versions, set fs.may_detach_mounts to 1. For more information, visit [Bug 1468249](https://example.com/bug1468249), [Bug 1441737](https://example.com/bug1441737), and [Issue 34538](https://example.com/issue34538).
- Grant the privileged permission to Logtail by adding the --privileged flag to the startup parameters. For more information, visit Docker run reference.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>${your_region_name}</code></td>
<td>The region to which the Log Service project belongs. Set this parameter based on the network type. Valid values:</td>
</tr>
<tr>
<td></td>
<td>• For the public network, set the region in the <code>region-internet</code> format. For example, the value for the China (Hangzhou) region is <code>cn-hangzhou-internet</code>.</td>
</tr>
<tr>
<td></td>
<td>• For the Alibaba Cloud internal network, set the region in the <code>region</code> format. For example, the value for the China (Hangzhou) region is <code>cn-hangzhou</code>.</td>
</tr>
<tr>
<td></td>
<td>For more information about installation parameters in each region, see <a href="#">Table 1 Logtail installation parameters</a>. Set this parameter based on the region of the project.</td>
</tr>
<tr>
<td><code>${your_aliyun_user_id}</code></td>
<td>The user ID. Set this parameter to the ID of your Alibaba Cloud account, which is a string. For information about how to view the ID, see Step 1 in Configure AliUids for ECS servers under other Alibaba Cloud accounts or on-premises IDCs.</td>
</tr>
<tr>
<td><code>${your_machine_group_user_defined_id}</code></td>
<td>The custom ID of your server group. For information about how to set the custom ID, see Step 1 in unique_44.</td>
</tr>
</tbody>
</table>

After you set the parameters, the command that you run to start the Logtail container is as follows:

```
docker run -d -v /:/logtail_host:ro -v /var/run:/var/run --env ALIYUN_LOGTAIL_CONFIG=/etc/logtail/conf/cn_hangzhou/logtail_config.json --env ALIYUN_LOGTAIL_USER_ID=1654218******** --env ALIYUN_LOGTAIL_USER_DEFINED_ID=log-docker-demo registry.cn-hangzhou.aliyuncs.com/log-service/logtail
```

**Note:**
You can customize the startup parameters of the Logtail container if the following conditions are met:

a. The following environment variables exist before you start the Logtail container: ALIYUN_LOGTAIL_USER_DEFINED_ID, ALIYUN_LOGTAIL_USER_ID, and ALIYUN_LOGTAIL_CONFIG.

b. The /var/run directory is mounted to the /var/run directory of the Logtail container.

c. To collect container standard output, container logs, or host files, you must mount the root directory to the /logtail_host directory of the Logtail container.

d. If the error log entry The parameter is invalid : uuid=none exists in the Logtail log file /usr/local/ilogtail/ilogtail.LOG, you must create a file named product_uuid on the host. Then, you must enter a legal UUID (for example, 169E98C9-ABC0-4A92-B1D2-AA6239C0D261) in the file and mount the file to the /sys/class/dmi/id/product_uuid directory of the Logtail container.

Configure a Logtail server group

1. Activate Log Service, and then create a project and a Logstore. For more information, see #unique_111.

2. In the left-side navigation pane, click the Server Groups icon to show the server group list.

3. Click the icon next to Server Groups, and then choose Create Server Group. You can also create a server group when you import data to Log Service.
4. In the dialog box that appears, select Custom ID from the Identifier drop-down list. Enter the value of `ALIYUN_LOGTAIL_USER_DEFINED_ID` set in the previous step in the Custom Identifier field.

Figure 3-28: Configure a Logtail server group

Click OK. One minute later, click the name of the server group in the Server Groups list. On the Server Group Settings page that appears, you can view the heartbeat status of the Logtail container. For more information, see the View the server group status section in *Manage server groups*.

Create a collection configuration

Create a Logtail collection configuration in the console as needed. For more information, see:

- *Container text logs (recommended)*
- *Container standard output (recommended)*
- *Host text logs*

The root directory of a host is mounted to the `/logtail_host` directory of the Logtail container by default. You must add the `/logtail_host` prefix to the log path. For example, if you want to collect data from the `/home/logs/app_log/`
directory of the host, you must set the log path as /logtail_host/home/logs/app_log/.

• #unique_47

What to do next

• Check the running status of the Logtail container.
  
  You can run the `docker exec ${logtail_container_id} /etc/init.d/ilogtaild status` command to check the running status of Logtail.

• View the version number, IP address, and startup time of Logtail.
  
  You can run the `docker exec ${logtail_container_id} cat /usr/local/ilogtail/app_info.json` command to view Logtail information.

• View the operational logs of Logtail.

  The operational logs of Logtail are stored in the ilogtail.LOG file in the /usr/local/ilogtail/ directory. If the log file is rotated and compressed, it is stored as a file named ilogtail.LOG.x.gz.

Example:

```
[root@iZbp17enxc2us3624wexh2Z ilogtail]# docker exec a287de895e40 tail -n 5 /usr/local/ilogtail/ilogtail.LOG
[2018-02-06 08:13:35.721864] [INFO] [8] [build/release64/sls/ilogtail/LogtailPlugin.cpp:104] logtail plugin Resume:start
[2018-02-06 08:13:35.722135] [INFO] [8] [build/release64/sls/ilogtail/LogtailPlugin.cpp:106] logtail plugin Resume:success
[2018-02-06 08:13:35.722149] [INFO] [8] [build/release64/sls/ilogtail/EventDispatcher.cpp:369] start add existed checkpoint events, size:0
[2018-02-06 08:13:35.722155] [INFO] [8] [build/release64/sls/ilogtail/EventDispatcher.cpp:511] add existed checkpoint events, size:0 event size:0 success count:0
[2018-02-06 08:13:39.725417] [INFO] [8] [build/release64/sls/ilogtail/ConfigManager.cpp:3776] check container path update flag:0 size:1
```

Ignore the following standard output:

```
start umount useless mount points, /shm$|/merged$|/mqueue$
umount: /logtail_host/var/lib/docker/overlay2/3fd0043af174cb0273c3c7869500fbeb2d95d13b1e110172ef57fe840c82155/merged: must be superuser to umount
umount: /logtail_host/var/lib/docker/overlay2/d5b10aa19399992755de1f85d25009528daa749c1bf8c16ed4f44beab6e69718/merged: must be superuser to umount
umount: /logtail_host/var/lib/docker/overlay2/5c3125daddacedec29df72ad0c52fac800cd56c6e880dc4e8a640b1e16c22dbe/merged: must be superuser to umount
...```
xargs: umount: exited with status 255; aborting
umount done
start logtail
ilogtail is running
logtail status:
ilogtail is running

- Restart Logtail.

To restart Logtail, use the following sample code:

```
[root@iZbp17enxc2us3624wexh2Z ilogtail]# docker exec a287de895e40 /
/etc/init.d/ilogtailld stop
kill process Name: ilogtail pid: 7
kill process Name: ilogtail pid: 8
stop success
[root@iZbp17enxc2us3624wexh2Z ilogtail]# docker exec a287de895e40 /
/etc/init.d/ilogtailld start
ilogtail is running
```

3.6.7 Kubernetes event collection

This topic describes how to use the eventer component to collect events from Kubernetes to Log Service. For more information about the source code for Kubernetes event collection, visit GitHub.

Collection configuration method

- **Note:**
  - If you use Alibaba Cloud Kubernetes, you only need to configure the endpoint, project, and logStore parameters.
  - If you use self-built Kubernetes, you need to configure the endpoint, project, logStore, regionId, internal, accessKeyId, and accessKeySecret parameters.

Deploy the YAML template as follows:

```
apiVersion: extensions/v1beta1
kind: Deployment
metadata:
  name: eventer-sls
  namespace: kube-system
spec:
  replicas: 1
  template:
    metadata:
      labels:
        task: monitoring
        k8s-app: eventer-sls
      annotations:
        scheduler.alpha.kubernetes.io/critical-pod: ''
    spec:
      serviceAccount: admin
      containers:
```
- name: eventer-sls
  image: registry.cn-hangzhou.aliyuncs.com/ringtail/eventer:v1.6.1.3
  imagePullPolicy: IfNotPresent
  command:
    - /eventer
    - --source=kubernetes:https://kubernetes.default
    - --sink=sls:https://${endpoint}? project=${project}&logStore=${logstore}

Parameter settings

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Type</th>
<th>Required</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>endpoint</td>
<td>String</td>
<td>Yes</td>
<td>The endpoint of Log Service. For more information, see #unique_26.</td>
</tr>
<tr>
<td>project</td>
<td>String</td>
<td>Yes</td>
<td>The project in Log Service.</td>
</tr>
<tr>
<td>logStore</td>
<td>String</td>
<td>Yes</td>
<td>The Logstore in Log Service.</td>
</tr>
<tr>
<td>internal</td>
<td>String</td>
<td>Yes</td>
<td>Specifies whether Alibaba Cloud Kubernetes is used. Set this parameter to false.</td>
</tr>
<tr>
<td>regionId</td>
<td>String</td>
<td>Yes</td>
<td>The region ID of Log Service. For more information, see #unique_26.</td>
</tr>
<tr>
<td>accessKeyId</td>
<td>String</td>
<td>Yes</td>
<td>The AccessKey ID of your account. We recommend that you use the AccessKey ID of a RAM user.</td>
</tr>
<tr>
<td>accessKeySecret</td>
<td>String</td>
<td>Yes</td>
<td>The AccessKey Secret of your account. We recommend that you use the AccessKey Secret of a RAM user.</td>
</tr>
</tbody>
</table>

Log sample

The following log sample is collected by Log Service. For more information, see Kubernetes.

The following table describes the content of the log.
<table>
<thead>
<tr>
<th>Log field</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>hostname</td>
<td>String</td>
<td>The hostname of the event.</td>
</tr>
<tr>
<td>level</td>
<td>String</td>
<td>The level of the log. Valid values: Normal and Warning.</td>
</tr>
<tr>
<td>pod_id</td>
<td>String</td>
<td>The unique identifier of the pod. This field is available only when the event type is related to the pod.</td>
</tr>
<tr>
<td>pod_name</td>
<td>String</td>
<td>The name of the pod. This field is available only when the event type is related to the pod.</td>
</tr>
<tr>
<td>eventId</td>
<td>JSON</td>
<td>The details of the event, which is a JSON string.</td>
</tr>
</tbody>
</table>

hostname:  cn-hangzhou.i-***********
level:  Normal
pod_id:  2a360760-1c82-11e9-9ddf-00163e0c7cbe
pod_name:  logtail-ds-blkkr
eventId:  

  "metadata":{
    "name":"logtail-ds-blkkr. 157b7cc90de7e192",
    "namespace":"kube-system",
    "selfLink":/api/v1/namespaces/kube-system/events/logtail-ds-blkkr.
    157b7cc90de7e192",
    "uid":"2aaf75ab-1c82-11e9-9ddf-00163e0c7cbe",
    "resourceVersion":6129169",
    "creationTimestamp":"2019-01-20T07:08:19Z"
  },
  "involvedObject":{
    "kind":"Pod",
    "namespace":"kube-system",
    "name":"logtail-ds-blkkr",
    "uid":"2a360760-1c82-11e9-9ddf-00163e0c7cbe",
    "apiVersion":"v1",
    "resourceVersion":6129161",
    "fieldPath":spec.containers{logtail}"
  },
  "reason":"Started",
  "message":"Started container",
  "source":{
    "component":"kubelet",
    "host":cn-hangzhou.i-***********
  },
  "firstTimestamp":"2019-01-20T07:08:19Z",
  "lastTimestamp":"2019-01-20T07:08:19Z",
  "count":1,
  "type":"Normal",
  "eventTime":null,
  "reportingComponent":",
  "reportingInstance":""
### 3.7 Logtail limits

#### Table 3-4: File collection

<table>
<thead>
<tr>
<th>Item</th>
<th>Capability and limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>File encoding</td>
<td>Log files encoded in UTF-8 and GBK are supported. We recommend that you use UTF-8 encoding for better processing performance. If log files are encoded in other formats, errors such as garbled characters and data losses may occur.</td>
</tr>
<tr>
<td>Log file size</td>
<td>Unlimited.</td>
</tr>
<tr>
<td>Log file rotation</td>
<td>Supported. Both <code>.log</code> and <code>.log</code> are supported for file names.</td>
</tr>
<tr>
<td>Log collection behavior when log parsing is blocked</td>
<td>When log parsing is blocked, Logtail keeps the log file descriptor (FD) in the open state. If log file rotation occurs multiple times during the block, Logtail attempts to keep the log parsing sequence of each rotation. If more than 20 unparsed logs are rotated, Logtail does not process subsequent log files. For more information, see the related technical document.</td>
</tr>
<tr>
<td>Symbolic link</td>
<td>Monitored directories can be symbolic links.</td>
</tr>
<tr>
<td>Single log size</td>
<td>The maximum size of a single log is 512 KB. If cross-line logs are divided by the regular expression for specifying the starting header of a cross-line log, the maximum size of each log is still 512 KB. If the size of a log exceeds 512 KB, the log is forcibly split into multiple parts for collection. For example, if the size of a log is 1,025 KB, the first 512 KB, the subsequent 512 KB, and the last 1 KB are processed sequentially.</td>
</tr>
<tr>
<td>Regular expression</td>
<td>Regular expressions can be Perl-compatible regular expressions.</td>
</tr>
</tbody>
</table>
### Log Service

#### Data Collection / 3 Logtail collection

<table>
<thead>
<tr>
<th>Item</th>
<th>Capability and limit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Multiple collection configurations for the same file</strong></td>
<td>Not supported. We recommend that you collect only one copy of log files to a Logstore and configure multiple subscriptions. If you need to collect multiple copies of log files, configure symbolic links for log files to bypass this limit.</td>
</tr>
<tr>
<td><strong>File opening behavior</strong></td>
<td>Logtail keeps a file to be collected in the open state. Logtail closes the file if the file is not modified within 5 minutes (when rotation does not occur).</td>
</tr>
<tr>
<td><strong>First log collection behavior</strong></td>
<td>Logtail collects only incremental log files. If modifications are found in a file for the first time and the file size exceeds 1 MB, Logtail collects the logs from the last 1 MB. Otherwise, Logtail collects the logs from the beginning. If a log file is not modified after the configuration is issued, Logtail does not collect this file.</td>
</tr>
<tr>
<td><strong>Non-standard text log</strong></td>
<td>If a log contains the '\0' lines, the log is truncated at the position of the first '\0' line.</td>
</tr>
</tbody>
</table>

Table 3-5: Checkpoint management

<table>
<thead>
<tr>
<th>Item</th>
<th>Capability and limit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Checkpoint timeout period</strong></td>
<td>If a file is not modified for more than 30 days, the checkpoint is deleted.</td>
</tr>
<tr>
<td><strong>Checkpoint storage policy</strong></td>
<td>Checkpoints are regularly saved (every 15 minutes) and are automatically saved when you exit the application.</td>
</tr>
<tr>
<td><strong>Checkpoint storage path</strong></td>
<td>By default, checkpoints are stored in the <code>/tmp/logtail_checkpoint</code> directory. For more information about how to modify the directory, see <code>Set startup parameters</code>.</td>
</tr>
</tbody>
</table>
### Table 3-6: Configuration

<table>
<thead>
<tr>
<th>Item</th>
<th>Capability and limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configuration update</td>
<td>Your updated configuration takes effect with a latency of about 30 seconds.</td>
</tr>
<tr>
<td>Dynamic configuration loading</td>
<td>Supported. The update of a Logtail configuration does not affect other Logtail configurations.</td>
</tr>
<tr>
<td>Number of configurations</td>
<td>Theoretically unlimited. We recommend that you create a maximum of 100 collection configurations on a server.</td>
</tr>
<tr>
<td>Multi-tenant data segregation</td>
<td>Collection configurations for different tenants are isolated.</td>
</tr>
</tbody>
</table>

### Table 3-7: Resources and performance

<table>
<thead>
<tr>
<th>Item</th>
<th>Capability and limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log processing throughput</td>
<td>The default traffic of raw logs is limited to 2 Mbit/s. (Data is uploaded after it is encoded and compressed, with a general compression ratio of 5 to 10 times.) Logs may be lost if the log traffic exceeds the limit. For more information about how to modify the parameters, see <em>Set startup parameters</em>.</td>
</tr>
<tr>
<td>Maximum performance</td>
<td>Single-core capability: The maximum processing capability is 100 Mbit/s for logs in simple mode, 20 Mbit/s by default for logs in full regex mode (depending on the complexity of regular expressions), 40 Mbit/s for logs in delimiter mode, and 30 Mbit/s for logs in JSON mode. After multiple processing threads are enabled, the performance can be improved by 1.5 to 3 times.</td>
</tr>
<tr>
<td>Number of monitored directories</td>
<td>Logtail limits the depth of monitored directories to conserve your resources. If the upper limit is reached, Logtail stops monitoring more directories and log files. Logtail monitors a maximum of 3,000 directories (including subdirectories).</td>
</tr>
</tbody>
</table>
### Number of monitored files

A Logtail configuration on each server monitors a maximum of 10,000 files, and the Logtail client on each server monitors a maximum of 100,000 files. Excessive files are not monitored. When the upper limit is reached, you can:

- Improve the depth of the monitored directory in each Logtail configuration.
- Modify the value of the `mem_usage_limit` parameter to increase the Logtail memory usage threshold. For more information, see [Set startup parameters](#).

The Logtail memory usage threshold can be set to a maximum of 2 GB, indicating that each Logtail configuration can monitor a maximum of 100,000 files, and each Logtail client can monitor a maximum of 1,000,000 files.

### Default resources

By default, Logtail occupies up to 40% of CPU usage and 256 MB of memory. If logs are generated at a high speed, you can modify relevant parameters. For more information, see [Set startup parameters](#).

### Resource out-of-limit processing policy

If the resources occupied by Logtail within 3 minutes exceed the upper limit, Logtail is forcibly restarted, which may cause data loss or duplication.

## Table 3-8: Error handling

<table>
<thead>
<tr>
<th>Item</th>
<th>Capability and limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network error handling</td>
<td>If the network is disconnected, Logtail retries and automatically adjusts the retry interval.</td>
</tr>
<tr>
<td>Resource quota out-of-limit processing</td>
<td>If the data transmission rate exceeds the quota of the Logstore, Logtail blocks log collection and automatically retries.</td>
</tr>
<tr>
<td>Maximum retry period before timeout</td>
<td>If data transmission fails for more than 6 successive hours, Logtail discards the data.</td>
</tr>
</tbody>
</table>
### Status self-check

Logtail automatically restarts if an exception occurs, for example, an application unexpectedly exits or resource usage exceeds the quota.

<table>
<thead>
<tr>
<th>Item</th>
<th>Capability and limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Status self-check</td>
<td>Logtail automatically restarts if an exception occurs, for example, an application unexpectedly exits or resource usage exceeds the quota.</td>
</tr>
</tbody>
</table>

### Log collection latency

Except for the block state, the latency in log collection by Logtail does not exceed 1 second after logs are flushed to a disk.

### Log upload policy

Logtail automatically aggregates logs in the same file before uploading the logs. Logs are uploaded if the number of logs exceeds 2,000, the total size of the log file exceeds 2 MB, or the log collection duration exceeds 3 seconds.

<table>
<thead>
<tr>
<th>Item</th>
<th>Capability and limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log collection latency</td>
<td>Except for the block state, the latency in log collection by Logtail does not exceed 1 second after logs are flushed to a disk.</td>
</tr>
<tr>
<td>Log upload policy</td>
<td>Logtail automatically aggregates logs in the same file before uploading the logs. Logs are uploaded if the number of logs exceeds 2,000, the total size of the log file exceeds 2 MB, or the log collection duration exceeds 3 seconds.</td>
</tr>
</tbody>
</table>
4 Cloud product collection

4.1 Cloud service logs

Log Service can collect logs from various cloud services, such as Elastic Compute Service (ECS), Object Storage Service (OSS), and Server Load Balancer (SLB). The logs record cloud service information including operation information, running statuses, and business dynamics.

The following table lists the Alibaba Cloud services from which Log Service can collect logs.

<table>
<thead>
<tr>
<th>Type</th>
<th>Cloud service</th>
<th>Activation method</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elastic computing</td>
<td>ECS</td>
<td>Install Logtail.</td>
<td>Logtail introduction</td>
</tr>
<tr>
<td></td>
<td>Container Service/Container Service for Kubernetes</td>
<td>Activate the service in the Container Service or Container Service for Kubernetes console.</td>
<td>Text logs and output</td>
</tr>
<tr>
<td>Storage</td>
<td>OSS</td>
<td>Activate the service in the OSS console.</td>
<td>Overview</td>
</tr>
<tr>
<td>Network</td>
<td>SLB</td>
<td>Activate the service in the SLB console.</td>
<td>Access logs of Layer-7 SLB</td>
</tr>
<tr>
<td></td>
<td>VPC</td>
<td>Activate the service in the VPC console.</td>
<td>#unique_18</td>
</tr>
<tr>
<td></td>
<td>API Gateway</td>
<td>Activate the service in the API Gateway console.</td>
<td>Access logs of API Gateway</td>
</tr>
<tr>
<td>Security</td>
<td>ActionTrail</td>
<td>Activate the service in the ActionTrail console.</td>
<td>Overview</td>
</tr>
<tr>
<td>Type</td>
<td>Cloud service</td>
<td>Activation method</td>
<td>Remarks</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>Threat Detection Service</td>
<td>Purchase Threat Detection Service Enterprise Edition and activate the service in the Threat Detection Service console.</td>
<td>TDS logs</td>
<td></td>
</tr>
<tr>
<td>Anti-Bot Service</td>
<td>Activate the service in the Anti-Bot Service console.</td>
<td>Anti-Bot Service logs</td>
<td></td>
</tr>
<tr>
<td>Application</td>
<td>Log Service</td>
<td>Activate the service in the Log Service console.</td>
<td>#unique_24</td>
</tr>
</tbody>
</table>

4.2 API Gateway access logs

Alibaba Cloud API Gateway provides API hosting service to facilitate micro-service aggregation, frontend and backend isolation, and system integration. Each API request corresponds to an access record, which contains information such as the IP address of the API caller, requested URL, response latency, returned status code, and number of bytes for each request and response. With the preceding information, you can understand the operating status of your web services.

Figure 4-1: API Gateway

Log Service allows you to collect API Gateway access logs by using the data import wizard.
Characteristics

1. Online log querying: You can use any keywords to query logs. Both exact query and fuzzy matching are supported. Log query can be used for troubleshooting or statistical query.

2. Detailed API call logs: You can search for detailed API call logs.

3. Customized analysis charts: You can customize charts based on any log data to meet your statistical and business requirements.

4. Preset analysis reports: API Gateway has predefined global statistical charts, including those for the number of requests, success rate, failure rate, latency, the number of applications that call APIs, failure statistics, top groups, top APIs, and top latency.

Field description

<table>
<thead>
<tr>
<th>Log field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>apiGroupUid</td>
<td>The ID of the API group.</td>
</tr>
<tr>
<td>apiGroupName</td>
<td>The name of the API group.</td>
</tr>
<tr>
<td>apiUid</td>
<td>The ID of the API.</td>
</tr>
<tr>
<td>apiName</td>
<td>The name of the API.</td>
</tr>
<tr>
<td>apiStageUid</td>
<td>The ID of the API stage.</td>
</tr>
<tr>
<td>apiStageName</td>
<td>The name of the API stage.</td>
</tr>
<tr>
<td>httpMethod</td>
<td>The HTTP method of the request.</td>
</tr>
<tr>
<td>path</td>
<td>The request path.</td>
</tr>
<tr>
<td>domain</td>
<td>The domain name that sends the request.</td>
</tr>
<tr>
<td>statusCode</td>
<td>The HTTP status code.</td>
</tr>
<tr>
<td>errorMessage</td>
<td>The error message.</td>
</tr>
<tr>
<td>appId</td>
<td>The ID of the application that sends the request.</td>
</tr>
<tr>
<td>appName</td>
<td>The name of the application that sends the request.</td>
</tr>
<tr>
<td>clientIp</td>
<td>The IP address of the client that sends the request.</td>
</tr>
<tr>
<td>exception</td>
<td>The specific error message returned by the backend server.</td>
</tr>
<tr>
<td>Log field</td>
<td>Description</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>providerAliUid</td>
<td>The account ID of the API provider.</td>
</tr>
<tr>
<td>region</td>
<td>The ID of the region, such as cn-hangzhou.</td>
</tr>
<tr>
<td>requestHandleTime</td>
<td>The time when the request is sent. It must be in GMT.</td>
</tr>
<tr>
<td>requestId</td>
<td>The request ID. It must be globally unique.</td>
</tr>
<tr>
<td>requestSize</td>
<td>The size of the request message. Unit: bytes.</td>
</tr>
<tr>
<td>responseSize</td>
<td>The size of the response message. Unit: bytes.</td>
</tr>
<tr>
<td>serviceLatency</td>
<td>The backend latency. Unit: ms.</td>
</tr>
</tbody>
</table>

**Procedure**

1. Create a project and a Logstore.

   For more information about the creation, see Create a project and Create a Logstore.

   Skip this step if a Logstore already exists.

2. Expand the target Logstore and click the plus sign (+) icon next to Data Import.

   You can also click the Import Data button on the right of the page, and select a Logstore in the data import process.

3. Select a data type.

   Click API Gateway on the Import Data page.
4. Select a Logstore.

The system skips this step if you enter the log collection configuration process by clicking the plus sign (+) icon next to Data Import under a Logstore on the Logstores page.

5. Specify the data source.

In the step of specifying the data source, check whether you have completed the following configurations:

a. Activate API Gateway.

API Gateway provides a complete suite of API hosting services, helping you share services and data with partners in the form of APIs. It also allows you to release your APIs in the API marketplace for more developers to purchase and use.

If you have not activated API Gateway, activate it as instructed on the relevant page.

b. Complete RAM authorization.

Use RAM to authorize Log Service before you establish a dispatch rule. This allows Log Service to collect your API Gateway logs.

Click Authorize in the upper-right corner for quick authorization.

c. Create a dispatch rule.

If this is the first dispatch rule that you create, the system automatically imports API Gateway logs and creates a dispatch rule. If you have configured API Gateway log collection, a message indicating that the log dispatch rule already exists is displayed. You can delete the existing dispatch rule.

6. Configure query and analysis.

By default, Log Service creates indexes for you. To create or modify indexes, choose Index Attributes > Modify on the search and analysis page.

You have completed data import initialization. You can select the api-gateway-access-log Logstore to query and analyze logs, or go to the dashboard page to view reports.
4.3 MNS logs

Message Service (MNS) log feature pushes your message operation logs to the specified Logging Bucket. You can configure to push logs to Log Service in the console, and enable the log feature for a queue/topic in the region. MNS automatically pushes operation logs of the queue/topic messages to the specified Logging Bucket.

- If you delete a Log Service project or Logstore corresponding to the Logging Bucket, or revoke the permission granted to MNS, logs cannot be pushed to Log Service normally.
- The log delay time is about five minutes.
- A Logging Bucket is configured for each region. All the message operation logs of the queues/topics with the log feature enabled in the region are pushed to this Logging Bucket.
- You can set whether to enable the log feature for each queue/topic separately. The function is disabled by default.

Precautions

- If you delete a Log Service project or Logstore corresponding to the Logging Bucket, or revoke the permission granted to MNS, logs cannot be pushed to Log Service normally.
- The log delay time is about five minutes.
- A Logging Bucket is configured for each region. All the message operation logs of the queues/topics with the log feature enabled in the region are pushed to this Logging Bucket.
- You can set whether to enable the log feature for each queue/topic separately. The function is disabled by default.

Prerequisites

1. You have already activated Log Service and Message Service MNS.
2. Message service logs can only be pushed to Log Service project in the corresponding region. Make sure that you have created a project and Logstore in the appropriate region.
Procedure

1. Enable the log feature for Queue and Topic.

   Take the log feature for Queue as an example.

   a. In the MNS console, click Queues in the left-side navigation pane, and select a region.

   b. Click Modify Settings at the right of queue for log collecting.

   c. In the displayed Modify Queue dialog box, turn on the Enable Logging switch.

   Note:
   - This feature is disabled by default. Make sure this feature is enabled for all the queues where logs are to be collected.
   - As the procedure for enabling the log feature for Topic is similar to one for Queue, follow the preceding steps to enable the log feature for Topic.

2. enter the log query page.

   In the MNS console, click Logs in the left-side navigation pane to enter the Log Management page.

3. Select the region

   At the top of the page, select the region for the Queue or Topic. Then, click Configure in the Actions column.

4. Confirm authorization and Log Service project.

   - If you configure MNS logs push for the first time, follow the prompts on the Quick authorization page.

   - If you do not have a proper project and Logstore, go to the Log Service console, and follow the page prompts to create a project and Logstore. For more information, see #unique_93.

5. Push configuration.

   Click Configure at the right of a region. In the Push Logs to LogService tab, select the corresponding Project and Logstore name.

   Note:
   - Do not cancel authorization or delete the Resource Access Management (RAM) role. Otherwise, MNS logs cannot be pushed to Log Service normally.
• Make sure Logging Bucket and Log Service project are in the same region.
• After creating a project and Logstore, return to the configuration page of the MNS console, and click Refresh to view the new project and Logstore.

Click OK after completing the configuration.

Log format

Queue message operation logs

Queue message operation logs are generated when queue messages are operated, such as sending a message, consuming a message, or deleting a message.

A message operation log contains multiple fields. Fields contained in message operation logs vary with operations.

Log field description

The description of each field is shown in the following table.

<table>
<thead>
<tr>
<th>Field</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>The operation occurrence time.</td>
</tr>
<tr>
<td>MessageId</td>
<td>The ID of the message processed in this operation.</td>
</tr>
<tr>
<td>QueueName</td>
<td>The name of the queue corresponding to this operation.</td>
</tr>
<tr>
<td>AccountId</td>
<td>The account of the queue corresponding to this operation.</td>
</tr>
<tr>
<td>RemoteAddress</td>
<td>The address of the client that initiates this operation.</td>
</tr>
<tr>
<td>NextVisibleTime</td>
<td>The next visible time of the message after this operation is complete.</td>
</tr>
<tr>
<td>ReceiptHandleInRequest</td>
<td>The ReceiptHandle parameter passed in when this operation is performed.</td>
</tr>
<tr>
<td>ReceiptHandleInResponse</td>
<td>The ReceiptHandle parameter returned when this operation is complete.</td>
</tr>
</tbody>
</table>

Field description for each operation

Logs of different operations contain different fields. The following table lists fields contained in each operation.
### Log Service

#### Data Collection / 4 Cloud product collection

<table>
<thead>
<tr>
<th>Operation</th>
<th>Time</th>
<th>QueueName</th>
<th>Account</th>
<th>MessageId</th>
<th>RemoteAddress</th>
<th>NextVisibleTime</th>
<th>ReceiptHandleInResponse</th>
<th>ReceiptHandleInRequest</th>
</tr>
</thead>
<tbody>
<tr>
<td>SendMessage/BatchSendMessage</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PeekMessage/BatchPeekMessage</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ReceiveMessage/BatchReceiveMessage</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>-</td>
</tr>
<tr>
<td>ChangeMessageVisibility</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>DeleteMessage/BatchDeleteMessage</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>-</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Topic message operation logs**

Topic message operation logs are generated when topic messages are operated, including publishing and pushing a message.

The following sections describe the meaning of each field in the topic message operation logs, and fields contained in different operations.

**Log field description**

A message operation log contains multiple fields. The following table lists the description of each field.

<table>
<thead>
<tr>
<th>Field</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>The operation occurrence time.</td>
</tr>
<tr>
<td>MessageId</td>
<td>The ID of the message processed in this operation.</td>
</tr>
</tbody>
</table>
Field description for each operation

Logs of different operations contain different fields. The following table lists fields contained in each operation.

<table>
<thead>
<tr>
<th>Operation</th>
<th>Time</th>
<th>Message</th>
<th>TopicName</th>
<th>SubscriptionName</th>
<th>AccountId</th>
<th>RemoteAddress</th>
<th>NotifyStatus</th>
<th>SubscriptionName</th>
</tr>
</thead>
<tbody>
<tr>
<td>PublishMessage</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>-</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>-</td>
</tr>
<tr>
<td>Notify</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>-</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

NotifyStatus

NotifyStatus is a field contained only in logs for pushing messages. This field helps you investigate the reason why MNS fails to push messages to the endpoint.

Depending on the different NotifyStatus values, you can perform troubleshooting according to the suggestions provided in the following table.

<table>
<thead>
<tr>
<th>Error code</th>
<th>Description</th>
<th>Troubleshooting method</th>
</tr>
</thead>
<tbody>
<tr>
<td>2xx</td>
<td>The message is pushed successfully.</td>
<td>-</td>
</tr>
<tr>
<td>Other HTTP status code</td>
<td>After a message is pushed to an endpoint, the endpoint returns a non-2xx status code.</td>
<td>Check the processing logic of the endpoint.</td>
</tr>
<tr>
<td>Error code</td>
<td>Description</td>
<td>Troubleshooting method</td>
</tr>
<tr>
<td>------------------</td>
<td>------------------------------------------------------</td>
<td>------------------------------------------------------------------</td>
</tr>
<tr>
<td>InvalidHost</td>
<td>The endpoint specified in the subscription is invalid.</td>
<td>Use curl or telnet to check whether the endpoint in the subscription is valid.</td>
</tr>
<tr>
<td>ConnectTimeout</td>
<td>The connection to the endpoint specified in the subscription times out.</td>
<td>Use curl or telnet to check whether the endpoint in the subscription is accessible.</td>
</tr>
<tr>
<td>ConnectFailure</td>
<td>The connection to the endpoint specified in the subscription fails.</td>
<td>Use curl or telnet to check whether the endpoint in the subscription is accessible.</td>
</tr>
<tr>
<td>UnknownError</td>
<td>Unknown error.</td>
<td>Contact MNS technical support.</td>
</tr>
</tbody>
</table>

4.4 OSS access logs

4.4.1 Overview

When you access Object Storage Service (OSS), the system generates a large number of access logs. After you enable the logging feature for a bucket, OSS automatically generates an object by hour based on the predefined naming rules to store access logs for the bucket. Afterward, OSS writes the object to the specified target bucket.

Enable OSS log storage

1. Log on to the **OSS console**.

2. In the list of buckets on the left, click the name of the target bucket to go to the Overview tab page of the bucket.

3. Click the Basic Settings tab, click Configure in the Log field, enable Log, and then set Destination Bucket and Log Prefix.

   - Destination Bucket: Select the name of the bucket where you want to store logs from the drop-down list. You must select your own bucket that remains in the same data center as your Logstore.
   - Log prefix: Enter the directory where the log is generated and the prefix of the log. The log is stored in the specified directory.

4. Click Save.
Log naming rules

The following example shows the naming rules for objects that store access logs:

<TargetPrefix><SourceBucket>YYYY-MM-DD-HH-MM-SS-<UniqueString>

- <TargetPrefix>: the prefix that you have specified.
- <SourceBucket>: the name of the source bucket.
- YYYY-MM-DD-HH-MM-SS: the time in China Standard Time (UTC+8) when the log is created. YYYY indicates a 4-digit year, MM indicates a 2-digit month, DD indicates a 2-digit day, HH indicates a 2-digit hour, MM indicates a 2-digit minute, and SS indicates a 2-digit second.
- <UniqueString>: the string that OSS generates to identify the object.

For example, the name of an object used to store OSS access logs can be:

MyLog-OSS-example2015-09-10-04-00-00-0000

- MyLog is the log prefix that you have specified.
- oss-example is the name of the source bucket.
- 2015-09-10-04-00-00 is the time in China Standard Time (UTC+8) when the log is created.
- 0000 indicates the string that OSS generates to identify the object.

4.4.2 Collect OSS access logs

Log Service can collect Object Storage Service (OSS) access logs, query and analyze the collected OSS access logs in real time, and clearly display analysis results by using multiple visual charts. These professional collection and analysis features for OSS access logs simplify your operation log auditing and event tracing, and allow you to work more efficiently.

Prerequisites

1. You have activated Log Service.
2. You have activated OSS, and created one or more buckets.
3. The project in Log Service and the bucket in OSS belong to the same Alibaba Cloud account and remain in the same region.

Procedure
1. Authorize log collection.

   Click [here](#) to authorize Log Service to store OSS access logs to your Logstore.

   You can also choose Log Service > Authorize Log Service in the OSS console to complete authorization.

2. Associate one or more buckets with your Logstore.

   a. In the OSS console, choose Log Service > Configure to go to the Log Service page.

   b. Click Associate to continue with the next step.

   ![Figure 4-2: Associate one or more buckets with your Logstore](image)

   A. Select or create a project.

   Select the region where the Log Service project is located and the Log Service project name, and click Next.

   ![Note:](image)

   - Your project in Log Service and the bucket in OSS must remain in the same region. Log Service can collect and store logs from multiple buckets to the same Logstore for real-time queries and analysis.
   - If you have not created any Log Service project in the current region, you can click Create Project to create a project.

   B. Select or create a Logstore.

   Select a Logstore name and click Next.

   ![Note:](image)
If you have not created any Logstore of Log Service in the current region, you can click Create Logstore to create a Logstore.

C. Associate one or more buckets with your Logstore.

Select a bucket name and click Submit. You can also select multiple bucket names to associate these buckets with the Logstore in the current region.

Figure 4-3: Associate one or more buckets with your Logstore

You have created association rules. Click Configure Index in the dialog box that appears to go to the Log Service console and configure indexes.
4.4.3 Query OSS access logs

Log Service can query and analyze collected Object Storage Service (OSS) access logs in real time, and displays analysis results by using various visualized charts.

Configure query and analysis

1. After you have associated buckets with a Logstore, click Configure Index in the dialog box that appears to go to the Log Service console.

Figure 4-4: Configure query and analysis

The buckets have been associated with the logstore. Various data report templates are provided for real-time log analysis.

To enable Log Service, configure a logstore index.
2. **Log Service** provides several preset indexes for querying OSS access logs. For more information about the related fields, see *Log fields*.

   a. Find the target project and click the project name.
   
   b. Double-click the Logstore name and choose Index Attributes > Modify.

   Figure 4-5: Configure indexes

   ![Figure 4-5: Configure indexes](image)

   **Note:**

   By default, Log Service creates four dashboards for the Logstore associated with one or more buckets. After you complete the configuration, you can view these dashboards on the Dashboard page. You can also click Analyze Log next.
to a target Logstore on the Log Service page in the OSS console, and click a dashboard name in the left-side navigation pane to view the dashboard.

Figure 4-6: Analyze logs

Log Service provides the following default dashboards:

- **oss_operation_center**: displays the overall operation information.

Figure 4-7: oss_operation_center
- **oss_access_center**: displays statistics of access logs.

  Figure 4-8: oss_access_center

- **oss_performance_center**: displays statistics of performance.

  Figure 4-9: oss_performance_center
• **oss_audit_center**: displays statistics of object deletion and modification.

Figure 4-10: oss_audit_center

4.4.4 Log fields

This topic describes all fields of Object Storage Service (OSS) logs.

Types of OSS logs

Table 4-1: Types of OSS logs

<table>
<thead>
<tr>
<th>Log type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access logs</td>
<td>Record all access to the corresponding OSS buckets. Log Service collects logs of this type in real time.</td>
</tr>
</tbody>
</table>
Log type | Description
---|---
Batch deletion logs | Record the OSS objects deleted in batch deletion operations. Log Service collects logs of this type in real time.

**Note:**
When you call the DeleteObjects API operation, a request record is generated in an access log. The information about the objects that you requested to delete is stored in the HTTP body of a request. Therefore, a hyphen (-) is used to indicate the corresponding object in the access log. To retrieve a list of the deleted objects, check the corresponding batch deletion log. You can use the `request_id` parameter to associate the batch deletion request with the objects that you want to delete.

Hourly metering logs | Record hourly metering statistics in a specific bucket to support analysis. A delay of several hours exists between log generation and log collection.

OSS built-in logging and OSS access logging

Log Service provides the OSS access logging feature to record, collect, store, and analyze access data, data deleted in batch deletion operations, and hourly metering data of OSS. **OSS built-in logging** is a built-in feature of OSS, which is also used to record and store access data of OSS.

OSS access logs provided by Log Service contain all information about OSS access operations. However, these logs have different log fields from OSS built-in logs. The following table lists the differences of fields between the two types of logs.

<table>
<thead>
<tr>
<th>Field of OSS built-in logs</th>
<th>Field of OSS access logs (Log Service)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remote IP</td>
<td>client_ip</td>
<td>The IP address from which a request originates. The proxy or the client's firewall may block this field.</td>
</tr>
<tr>
<td>Time</td>
<td>time</td>
<td>The time when OSS receives a request.</td>
</tr>
<tr>
<td>Field of OSS built-in logs</td>
<td>Field of OSS access logs (Log Service)</td>
<td>Description</td>
</tr>
<tr>
<td>----------------------------</td>
<td>----------------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Request-URI</td>
<td>request-uri</td>
<td>The requested URL, including the query-string parameter.</td>
</tr>
<tr>
<td>HTTP Status</td>
<td>http_status</td>
<td>The HTTP status code that OSS returns.</td>
</tr>
<tr>
<td>SentBytes</td>
<td>response_body_length</td>
<td>The traffic consumed when a user downloads logs from OSS.</td>
</tr>
<tr>
<td>RequestTime (ms)</td>
<td>response_time</td>
<td>The time consumed to complete a request, in milliseconds.</td>
</tr>
<tr>
<td>Referer</td>
<td>referer</td>
<td>The HTTP header that identifies the address of the web page which is linked to the resource being requested.</td>
</tr>
<tr>
<td>User-Agent</td>
<td>User-Agent</td>
<td>The user-agent header in an HTTP request.</td>
</tr>
<tr>
<td>HostName</td>
<td>host</td>
<td>The domain name that is requested.</td>
</tr>
<tr>
<td>Request ID</td>
<td>request_id</td>
<td>The unique ID used to identify a request.</td>
</tr>
<tr>
<td>LoggingFlag</td>
<td>logging_flag</td>
<td>Indicates whether the access logging feature has been enabled.</td>
</tr>
<tr>
<td>Requester Aliyun ID</td>
<td>requester_id</td>
<td>The Alibaba Cloud ID of the requester. The value of this field is a hyphen (-) for anonymous access.</td>
</tr>
<tr>
<td>Operation</td>
<td>operation</td>
<td>The type of a request.</td>
</tr>
<tr>
<td>Bucket</td>
<td>bucket</td>
<td>The name of the bucket that is requested.</td>
</tr>
<tr>
<td>Field of OSS built-in logs</td>
<td>Field of OSS access logs (Log Service)</td>
<td>Description</td>
</tr>
<tr>
<td>---------------------------</td>
<td>----------------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Key</td>
<td>Object</td>
<td>The key that is requested.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Note: The object field of Log Service is URL encoded.</td>
</tr>
<tr>
<td>ObjectSize</td>
<td>object_size</td>
<td>The size of an object.</td>
</tr>
<tr>
<td>Server Cost Time (ms)</td>
<td>server_cost_time</td>
<td>The time consumed by the OSS server to process a request, in milliseconds.</td>
</tr>
<tr>
<td>Error Code</td>
<td>error_code</td>
<td>The error code that OSS returns.</td>
</tr>
<tr>
<td>Request Length</td>
<td>request_length</td>
<td>The length of a request, in bytes.</td>
</tr>
<tr>
<td>UserID</td>
<td>owner_id</td>
<td>The ID of a bucket owner.</td>
</tr>
<tr>
<td>Delta DataSize</td>
<td>delta_data_size</td>
<td>The variation of the size of a bucket. The value of this field is a hyphen (−) if the bucket size does not change.</td>
</tr>
<tr>
<td>Sync Request</td>
<td>sync_request</td>
<td>Indicates whether the request is forwarded to the origin of Content Delivery Network (CDN). The value of this field is a hyphen (−) if this is not a back-to-origin request.</td>
</tr>
</tbody>
</table>

Access log

Table 4-2: Access log

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>access_id</td>
<td>The AccessKey ID of the user's Alibaba Cloud account.</td>
<td>mEEJX*******</td>
</tr>
<tr>
<td>Field</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td><strong>topic</strong></td>
<td>The name of the topic in a log. The value of this field is fixed to oss_access_log</td>
<td>-</td>
</tr>
<tr>
<td>time</td>
<td>The time when OSS receives a request. Use the value of <strong>time</strong> if a timestamp is required.</td>
<td>27/Feb/2018:13:58:45</td>
</tr>
<tr>
<td>owner_id</td>
<td>The Alibaba Cloud ID of the bucket owner.</td>
<td>12345678</td>
</tr>
<tr>
<td>User-Agent</td>
<td>The user-agent header in an HTTP request.</td>
<td>curl/7.15.5</td>
</tr>
<tr>
<td>logging_flag</td>
<td>Indicates whether logging has been enabled to periodically export logs to OSS buckets.</td>
<td>true</td>
</tr>
<tr>
<td>bucket</td>
<td>The name of a bucket.</td>
<td>bucket123</td>
</tr>
<tr>
<td>content_length_in</td>
<td>The value of Content-Length in a request header, in bytes.</td>
<td>12345</td>
</tr>
<tr>
<td>content_length_out</td>
<td>The value of Content-Length in a response header, in bytes.</td>
<td>12345</td>
</tr>
<tr>
<td>Object</td>
<td>The URL encoded object of the request. You can use select url_decode (object) to decode the object when querying logs.</td>
<td>data%2Fcur_file.txt</td>
</tr>
<tr>
<td>object_size</td>
<td>The size of a requested object, in bytes.</td>
<td>1234</td>
</tr>
<tr>
<td>operation</td>
<td>The type of an access operation. For more information about access types and descriptions, see Access types.</td>
<td>GetObject</td>
</tr>
<tr>
<td>Field</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>---------------------------------------------------</td>
</tr>
<tr>
<td>bucket_location</td>
<td>The cluster where a bucket is located. The field values are in the format of <code>oss-&lt;region&gt;-id</code></td>
<td>oss-cn-beijing-f</td>
</tr>
<tr>
<td>request_uri</td>
<td>The URL encoded URI of the request, including the query-string parameter. You can use <code>url_decode(request_uri)</code> to decode the URI when querying logs.</td>
<td>/1518085703067732%/2Fcur_file.txt HTTP/1.1</td>
</tr>
<tr>
<td>error_code</td>
<td>The error code that OSS returns. For more information about error codes and descriptions, see #unique_122.</td>
<td>NoSuchKey</td>
</tr>
<tr>
<td>request_length</td>
<td>The size of an HTTP request, including the header, in bytes.</td>
<td>376</td>
</tr>
<tr>
<td>client_ip</td>
<td>The IP address from which a request originates.</td>
<td>1.2.3.4</td>
</tr>
<tr>
<td>response_body_length</td>
<td>The size of the body in an HTTP response, excluding the header.</td>
<td>123</td>
</tr>
<tr>
<td>http_method</td>
<td>The method of an HTTP request.</td>
<td>GET</td>
</tr>
<tr>
<td>referer</td>
<td>The HTTP header that identifies the address of the web page which is linked to the resource being requested.</td>
<td><a href="http://www.abc.com">http://www.abc.com</a></td>
</tr>
<tr>
<td>requester_id</td>
<td>The Alibaba Cloud ID of the requester. The value of this field is a hyphen (-) for anonymous access.</td>
<td>12345678</td>
</tr>
<tr>
<td>Field</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>request_id</td>
<td>The request ID that can be used by OSS technical support to troubleshoot issues.</td>
<td>5A7C39674857FB9FFFFFF</td>
</tr>
<tr>
<td>response_time</td>
<td>The response time of a request, in milliseconds.</td>
<td>123</td>
</tr>
<tr>
<td>server_cost_time</td>
<td>The time consumed by the OSS server to process a request, in milliseconds.</td>
<td>123</td>
</tr>
<tr>
<td>http_type</td>
<td>The type of an HTTP request. The value of this field is HTTP or HTTPS.</td>
<td>http</td>
</tr>
<tr>
<td>sign_type</td>
<td>The type of a signature. For more information about signature types and descriptions, see Signature types.</td>
<td>NormalSign</td>
</tr>
<tr>
<td>http_status</td>
<td>The status code of an HTTP connection returned by the OSS server.</td>
<td>200</td>
</tr>
<tr>
<td>sync_request</td>
<td>The type of a synchronization request. For more information about synchronization request types and descriptions, see Synchronization request types.</td>
<td>cn</td>
</tr>
<tr>
<td>bucket_storage_type</td>
<td>The type of bucket storage. For more information about bucket storage types and descriptions, see Bucket storage types.</td>
<td>standard</td>
</tr>
<tr>
<td>host</td>
<td>The domain name that is requested.</td>
<td>bucket123.oss-cn-beijing.aliyuncs.com</td>
</tr>
<tr>
<td>Field</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>vpc_addr</td>
<td>The virtual IP address (VIP) that corresponds to the domain name that you requested. The data type of this field is in Integer. This field can be used by OSS technical support to troubleshoot issues.</td>
<td>1234567890</td>
</tr>
<tr>
<td>vpc_id</td>
<td>The ID of the Virtual Private Cloud (VPC) ID when the requester accessed. This field can be used by OSS technical support to troubleshoot issues.</td>
<td>1234</td>
</tr>
<tr>
<td>delta_data_size</td>
<td>The variation of the size of an object. The value of this field is 0 if the object size does not change. The value of this field is a hyphen (-) for requests other than uploads.</td>
<td>280</td>
</tr>
<tr>
<td>acc_access_region</td>
<td>For a request in CDN, this field is the domain name corresponding to the region where the requested access point is located. Otherwise, The value of this field is a hyphen (-).</td>
<td>cn-hangzhou</td>
</tr>
</tbody>
</table>

Batch deletion logs

When you call the DeleteObjects API operation, a request record is generated in an access log. The information about the objects that you requested to delete is stored in the HTTP body of a request. Therefore, a hyphen (−) is used to indicate the corresponding object in the access log. To retrieve a list of the deleted objects, check the corresponding batch deletion log. The following table lists the fields and descriptions of batch deletion logs. You can use the request_id parameter to associate the batch deletion request with the objects that you want to delete.
Table 4-3: Batch deletion logs

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>topic</strong></td>
<td>The name of the topic in a log. The value of this field is oss_batch_delete_log.</td>
<td>-</td>
</tr>
<tr>
<td>client_ip</td>
<td>The IP address from which a request originates.</td>
<td>1.2.3.4</td>
</tr>
<tr>
<td>user_agent</td>
<td>The user-agent header in an HTTP request.</td>
<td>curl/7.15.5</td>
</tr>
<tr>
<td>bucket</td>
<td>The name of a bucket.</td>
<td>bucket123</td>
</tr>
<tr>
<td>error_code</td>
<td>The error code that OSS returns. For more information about error codes and descriptions, see #unique_122.</td>
<td>NoSuchKey</td>
</tr>
<tr>
<td>request_length</td>
<td>The size of an HTTP request, including the header, in bytes.</td>
<td>376</td>
</tr>
<tr>
<td>response_body_length</td>
<td>The size of the body in an HTTP response, excluding the header.</td>
<td>123</td>
</tr>
<tr>
<td>Object</td>
<td>The URL encoded object of the request. You can use select url_decode (object) to decode the object when querying logs.</td>
<td>data%2Fcur_file.txt</td>
</tr>
<tr>
<td>object_size</td>
<td>The size of a request object, in bytes.</td>
<td>1234</td>
</tr>
<tr>
<td>operation</td>
<td>The type of an access operation. For more information about access types and descriptions, see Access types.</td>
<td>GetObject</td>
</tr>
<tr>
<td>Field</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------------------------------------------</td>
</tr>
<tr>
<td>bucket_location</td>
<td>The cluster where a bucket is located. The field values are in the format of <code>oss-&lt;region&gt;-id</code>.</td>
<td>oss-cn-beijing-f</td>
</tr>
<tr>
<td>http_method</td>
<td>The method of an HTTP request.</td>
<td>POST</td>
</tr>
<tr>
<td>referer</td>
<td>The HTTP header that identifies the address of the webpage which is linked to the resource being requested.</td>
<td><a href="http://www.abc.com">http://www.abc.com</a></td>
</tr>
<tr>
<td>request_id</td>
<td>The request ID that can be used by OSS technical support to troubleshoot issues.</td>
<td>5A7C39674857FB9FFFFFF</td>
</tr>
<tr>
<td>http_status</td>
<td>The status code of an HTTP connection returned by the OSS server.</td>
<td>200</td>
</tr>
<tr>
<td>sync_request</td>
<td>The type of a synchronization request. For more information about synchronization request types and descriptions, see Synchronization request types.</td>
<td>cdn</td>
</tr>
<tr>
<td>request_uri</td>
<td>The URL encoded URI of a request, including the query-string parameter. You can use <code>select url_decode(request_uri)</code> to decode the URI when querying logs.</td>
<td>/1518085703067732%/2Fcur_file.txt HTTP/1.1</td>
</tr>
<tr>
<td>host</td>
<td>The domain name that is requested.</td>
<td>bucket123.oss-cn-beijing.aliyuncs.com</td>
</tr>
</tbody>
</table>
### Log Service

Data Collection / 4 Cloud product collection

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>logging_flag</td>
<td>Indicates whether logging has been enabled to periodically export logs to OSS buckets.</td>
<td>true</td>
</tr>
<tr>
<td>server_cost_time</td>
<td>The time consumed by the OSS server to process a request, in milliseconds.</td>
<td>123</td>
</tr>
<tr>
<td>owner_id</td>
<td>The Alibaba Cloud ID of the bucket owner.</td>
<td>12345678</td>
</tr>
<tr>
<td>requester_id</td>
<td>The Alibaba Cloud ID of the requester. The value of this field is a hyphen (-) for anonymous access.</td>
<td>12345678</td>
</tr>
<tr>
<td>delta_data_size</td>
<td>The variation of the size of an object. The value of this field is 0 if the object size does not change. The value of this field is a hyphen (-) for requests other than uploads.</td>
<td>280</td>
</tr>
</tbody>
</table>

Hourly metering logs

This type of logs record hourly metering statistics in a specific bucket to support analysis. A delay of several hours exists between log generation and log collection.

Table 4-4: Hourly metering log

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>topic</strong></td>
<td>The name of the topic in a log. The value of this field is oss_metering_log.</td>
<td>-</td>
</tr>
<tr>
<td>owner_id</td>
<td>The Alibaba Cloud ID of the bucket owner.</td>
<td>12345678</td>
</tr>
<tr>
<td>bucket</td>
<td>The name of a bucket.</td>
<td>bucket123</td>
</tr>
<tr>
<td>cdn_in</td>
<td>The CDN inbound traffic, in bytes.</td>
<td>123</td>
</tr>
<tr>
<td>cdn_out</td>
<td>The CDN outbound traffic, in bytes.</td>
<td>123</td>
</tr>
<tr>
<td>Field</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>get_request</td>
<td>The number of GET requests.</td>
<td>123</td>
</tr>
<tr>
<td>intranet_in</td>
<td>The inbound traffic of the internal network, in bytes.</td>
<td>123</td>
</tr>
<tr>
<td>intranet_out</td>
<td>The outbound traffic of the internal network, in bytes.</td>
<td>123</td>
</tr>
<tr>
<td>network_in</td>
<td>The Internet inbound traffic, in bytes.</td>
<td>123</td>
</tr>
<tr>
<td>network_out</td>
<td>The Internet outbound traffic, in bytes.</td>
<td>123</td>
</tr>
<tr>
<td>put_request</td>
<td>The number of PUT requests.</td>
<td>123</td>
</tr>
<tr>
<td>storage_type</td>
<td>The type of bucket storage. For more information about bucket storage types and descriptions, see <a href="#">Bucket storage types</a>.</td>
<td>standard</td>
</tr>
<tr>
<td>storage</td>
<td>The storage capacity of a bucket, in bytes.</td>
<td>123</td>
</tr>
<tr>
<td>metering_datasize</td>
<td>The size of metering data in non-standard storage.</td>
<td>123</td>
</tr>
<tr>
<td>process_img_size</td>
<td>The size of a processed image, in bytes.</td>
<td>123</td>
</tr>
<tr>
<td>process_img</td>
<td>The name of a processed image.</td>
<td>123</td>
</tr>
<tr>
<td>sync_in</td>
<td>The synchronous inbound traffic, in bytes.</td>
<td>123</td>
</tr>
<tr>
<td>sync_out</td>
<td>The synchronous outbound traffic, in bytes.</td>
<td>123</td>
</tr>
<tr>
<td>start_time</td>
<td>The timestamp when a metering operation starts.</td>
<td>1518084000</td>
</tr>
<tr>
<td>Field</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>end_time</td>
<td>The timestamp when a metering operation ends. The metering data is hourly collected.</td>
<td>1518087600</td>
</tr>
<tr>
<td>region</td>
<td>The region where a bucket is located.</td>
<td>cn-beijing</td>
</tr>
<tr>
<td>bucket_location</td>
<td>The cluster where a bucket is located. The field values are in the format of oss-&lt;region&gt;-id.</td>
<td>oss-cn-beijing-f</td>
</tr>
</tbody>
</table>

**Operation types**

Table 4-5: Operation types

<table>
<thead>
<tr>
<th>Operation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AbortMultiPartUpload</td>
<td>Stops a multipart upload.</td>
</tr>
<tr>
<td>AppendObject</td>
<td>Appends an object.</td>
</tr>
<tr>
<td>CommitTransition</td>
<td>Commits a transition.</td>
</tr>
<tr>
<td>CompleteUploadPart</td>
<td>Completes a multipart upload.</td>
</tr>
<tr>
<td>CopyObject</td>
<td>Copies an object.</td>
</tr>
<tr>
<td>DeleteBucket</td>
<td>Deletes a bucket.</td>
</tr>
<tr>
<td>DeleteLiveChannel</td>
<td>Deletes a LiveChannel.</td>
</tr>
<tr>
<td>DeleteObject</td>
<td>Deletes an object.</td>
</tr>
<tr>
<td>DeleteObjects</td>
<td>Deletes multiple objects.</td>
</tr>
<tr>
<td>ExpireObject</td>
<td>Makes an object expire.</td>
</tr>
<tr>
<td>GetBucket</td>
<td>Queries objects.</td>
</tr>
<tr>
<td>GetBucketAcl</td>
<td>Obtains permissions of a bucket.</td>
</tr>
<tr>
<td>GetBucketCors</td>
<td>Queries the cross-origin resource sharing (CORS) rules of a bucket.</td>
</tr>
<tr>
<td>GetBucketEventNotification</td>
<td>Queries the notification configurations of a bucket.</td>
</tr>
<tr>
<td>GetBucketInfo</td>
<td>Queries the information about a bucket.</td>
</tr>
<tr>
<td>Operation</td>
<td>Description</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>GetBucketLifecycle</td>
<td>Queries the lifecycle configurations of a bucket.</td>
</tr>
<tr>
<td>GetBucketLocation</td>
<td>Queries the region where a bucket is located.</td>
</tr>
<tr>
<td>GetBucketLog</td>
<td>Queries the access log configurations of a bucket.</td>
</tr>
<tr>
<td>GetBucketReferer</td>
<td>Queries the hotlink protection configurations of a bucket.</td>
</tr>
<tr>
<td>GetBucketReplication</td>
<td>Queries the cross-region replication configurations.</td>
</tr>
<tr>
<td>GetBucketReplicationProgress</td>
<td>Queries the progress of a cross-region replication.</td>
</tr>
<tr>
<td>GetBucketStat</td>
<td>Queries the information about a bucket.</td>
</tr>
<tr>
<td>GetBucketWebSite</td>
<td>Queries the static website hosting status of a bucket.</td>
</tr>
<tr>
<td>GetLiveChannelStat</td>
<td>Queries the status of a LiveChannel.</td>
</tr>
<tr>
<td>GetObject</td>
<td>Reads an object.</td>
</tr>
<tr>
<td>GetObjectAcl</td>
<td>Obtains the access control list (ACL) of an object.</td>
</tr>
<tr>
<td>GetObjectInfo</td>
<td>Queries the information about an object.</td>
</tr>
<tr>
<td>GetObjectMeta</td>
<td>Queries the metadata of an object.</td>
</tr>
<tr>
<td>GetObjectSymlink</td>
<td>Queries the details of the object that a symbolic link refers to.</td>
</tr>
<tr>
<td>GetPartData</td>
<td>Queries the data in all parts of an object.</td>
</tr>
<tr>
<td>GetPartInfo</td>
<td>Queries the information about all parts of an object.</td>
</tr>
<tr>
<td>GetProcessConfiguration</td>
<td>Queries the image processing configurations of a bucket.</td>
</tr>
<tr>
<td>GetService</td>
<td>Queries buckets.</td>
</tr>
<tr>
<td>HeadBucket</td>
<td>Queries the information about a bucket.</td>
</tr>
<tr>
<td>HeadObject</td>
<td>Queries the information about an object.</td>
</tr>
<tr>
<td>Operation</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>InitiateMultipartUpload</td>
<td>Initializes the object for multipart upload.</td>
</tr>
<tr>
<td>ListMultiPartUploads</td>
<td>Queries multipart upload events.</td>
</tr>
<tr>
<td>ListParts</td>
<td>Queries status of all parts of an object.</td>
</tr>
<tr>
<td>Options</td>
<td>Queries the options.</td>
</tr>
<tr>
<td>PostObject</td>
<td>Uploads an object by using a form.</td>
</tr>
<tr>
<td>PostProcessTask</td>
<td>Commits data processing operations, such as taking snapshots.</td>
</tr>
<tr>
<td>PostVodPlaylist</td>
<td>Creates a video-on-demand (VOD) playlist of a LiveChannel.</td>
</tr>
<tr>
<td>ProcessImage</td>
<td>Processes an image.</td>
</tr>
<tr>
<td>PutBucket</td>
<td>Creates a bucket.</td>
</tr>
<tr>
<td>PutBucketCors</td>
<td>Specifies the CORS rule for a bucket.</td>
</tr>
<tr>
<td>PutBucketLifecycle</td>
<td>Specifies the lifecycle configurations of a bucket.</td>
</tr>
<tr>
<td>PutBucketLog</td>
<td>Specifies the access log for a bucket.</td>
</tr>
<tr>
<td>PutBucketWebSite</td>
<td>Specifies the static website hosting mode for a bucket.</td>
</tr>
<tr>
<td>PutLiveChannel</td>
<td>Creates a LiveChannel.</td>
</tr>
<tr>
<td>PutLiveChannelStatus</td>
<td>Specifies the status of a LiveChannel.</td>
</tr>
<tr>
<td>PutObject</td>
<td>Upload an object.</td>
</tr>
<tr>
<td>PutObjectAcl</td>
<td>Modifies the ACL of an object.</td>
</tr>
<tr>
<td>PutObjectSymlink</td>
<td>Creates the object by using the symbolic link.</td>
</tr>
<tr>
<td>RedirectBucket</td>
<td>Redirects the request to a bucket endpoint.</td>
</tr>
<tr>
<td>RestoreObject</td>
<td>Restores an object.</td>
</tr>
<tr>
<td>UploadPart</td>
<td>Resumes uploading an object from a checkpoint.</td>
</tr>
<tr>
<td>UploadPartCopy</td>
<td>Copies a part.</td>
</tr>
<tr>
<td>get_image_exif</td>
<td>Queries the exchangeable image file format (Exif) data of an image.</td>
</tr>
</tbody>
</table>
Log Service

<table>
<thead>
<tr>
<th>Operation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>get_image_info</td>
<td>Queries the length and width of an image.</td>
</tr>
<tr>
<td>get_image_infoexif</td>
<td>Queries the length, width, and Exif data of an image.</td>
</tr>
<tr>
<td>get_style</td>
<td>Queries the picture processing rule of a bucket.</td>
</tr>
<tr>
<td>list_style</td>
<td>Queries all picture processing rules of a bucket.</td>
</tr>
<tr>
<td>put_style</td>
<td>Creates a picture processing rule for a bucket.</td>
</tr>
</tbody>
</table>

For more information about each operation, see #unique_126.

Synchronization request types

Table 4-6: Synchronization request type

<table>
<thead>
<tr>
<th>Synchronization request type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>Indicates a common request.</td>
</tr>
<tr>
<td>cdn</td>
<td>Indicates a back-to-origin request from CDN.</td>
</tr>
</tbody>
</table>

Signature types

Table 4-7: Signature type

<table>
<thead>
<tr>
<th>Signature type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NotSign</td>
<td>Indicates that a request was not signed.</td>
</tr>
<tr>
<td>NormalSign</td>
<td>Indicates that a request was signed with a normal signature.</td>
</tr>
<tr>
<td>UriSign</td>
<td>Indicates that a request was signed with a URL signature.</td>
</tr>
<tr>
<td>AdminSign</td>
<td>Indicates that a request was signed by an administrator.</td>
</tr>
</tbody>
</table>

For more information about signatures, see #unique_127.
Bucket storage classes

Table 4-8: Bucket storage classes

<table>
<thead>
<tr>
<th>Storage class</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>standard</td>
<td>The default storage class for frequently accessed objects.</td>
</tr>
<tr>
<td>archive</td>
<td>The storage class for archiving objects.</td>
</tr>
<tr>
<td>infrequent_access</td>
<td>The storage class for infrequently accessed objects.</td>
</tr>
</tbody>
</table>

For more information about each storage class, see #unique_128.

4.5 SLB Layer 7 access log

Alibaba Cloud Server Load Balancer (SLB) distributes traffic among multiple ECS instances to improve the service capabilities of your applications. You can use SLB to prevent single point of failures (SPOFs) and improve the availability of your applications.

The access log feature of SLB can be applied to HTTP- and HTTPS-based Layer 7 load balancing. Access logs can contain about 30 fields such as the time when a request is received, the IP address of the client, processing latency, request URI, backend server (ECS instance) address, and returned status code. As an Internet access point, SLB needs to distribute a large number of access requests. You can use access logs to analyze the behaviors and geographical distribution of client users, and troubleshoot issues.

Note:

The access log feature can only be applied to Layer 7 load balancing. This feature is available in all regions.

Advantages

- Simple: The access log feature enables developers and operations and maintenance (O&M) personnel to break free from tedious and time-consuming log processing so that they can concentrate on business development and technical research.
• Massive: Access logs are directly proportional to request PVs of SLB instances. Typically, a large amount of data needs to be processed. Therefore, you must consider performance and cost issues when you process access logs. SLB can analyze 100 million logs in one second and has obvious cost advantages compared with open source solutions.

• Real-time: Scenarios such as DevOps, monitoring, and alerting require real-time log data. Traditional methods cannot meet this requirement. For example, performing ETL in Hive requires a long time. During this process, a lot of work is required for data integration. The access log feature of SLB combines the powerful big data computing capability of Alibaba Cloud Log Service to analyze and process real-time data in seconds.

• Elastic: You can enable or disable the access log feature based on SLB instances. You can set the storage period as needed and dynamically scale the Logstore capacity to meet business growth requirements.

Configure Log Service to collect SLB Layer 7 access logs

Prerequisites

1. SLB and Log Service are activated. An SLB instance, a Log Service project, and a Logstore are created. The Log Service project and the Logstore must belong to the same region. For more information about how to create an SLB instance, see #unique_129.

   Note:
   Only Layer 7 load balancing supports the access log feature. This feature is now available in all regions.

2. If you log on to the console as a RAM user, you must first use your Alibaba Cloud account to authorize the RAM user. For more information, see #unique_130.

Procedure

1. Log on to the Log Service console, and then click the target project name.

2. Click the Import Data button. On the Import Data page that appears, select SLB Layer-7 Access Logs.

3. Select a Logstore.

   Select an existing Logstore, or create a new project or Logstore.
4. Create a dispatch rule as prompted. Click Configure SLB to go to the SLB console.
   a. In the left-side navigation pane, choose Logs > Access Logs.
   b. Find the target SLB instance and click Configure Logging next to the instance.

   ![Note: Make sure that the project and the SLB instance are in the same region.]

   Figure 4-11: Log settings

   c. Specify the project and Logstore and then click OK.
   d. After the configuration is complete, close the dialog box. Return to the data import wizard and click Next.

5. Configure query and analysis.

   Log Service provides several preset indexes for querying SLB access logs. For more information about the related fields, see Field description. Confirm the settings and click Next.

   ![Note: By default, the system creates two dashboards whose names start with Logstore names: {LOGSTORE}-slb_layer7_access_center and {LOGSTORE}-slb_layer7_operation_center. After the configuration is complete, you can view the dashboards on the Dashboard page.]

What to do next

- Real-time log query

You can use any keywords to query logs. Both exact query and fuzzy matching are supported. Log query can be used for troubleshooting or statistical query.
- Preset analysis reports

Log Service predefines two dashboards for your SLB logs: access_center and operation_center. access_center displays access details and operation_center displays the overview information. The displayed information includes top clients that send the most access requests, distribution of request status codes, top URIs that send the most access requests, trends of request packet traffic, and statistics on backend server response time.

- Custom analysis charts

You can perform ad-hoc queries based on any log data and save the results as charts to meet your statistical and business requirements.

- Log monitoring alert

You can perform custom analysis on SLB request logs and save the result as a query. You can also set an alert based on the query. When real-time log computing results exceed the defined threshold, the system sends an alert.

Field description

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>body_bytes_sent</td>
<td>The number of bytes of the HTTP message body sent to the client.</td>
</tr>
<tr>
<td>client_ip</td>
<td>The IP address of the request client.</td>
</tr>
<tr>
<td>host</td>
<td>The hostname. Obtain the value from the request parameters first. If no value is obtained, obtain it from the host header. If the value still cannot be obtained, use the IP address of the backend server that processes the request as the hostname.</td>
</tr>
<tr>
<td>http_host</td>
<td>The host header in the request message.</td>
</tr>
<tr>
<td>http_referer</td>
<td>The HTTP referer header in the request message received by the proxy.</td>
</tr>
<tr>
<td>http_user_agent</td>
<td>The HTTP user-agent header in the request message received by the proxy.</td>
</tr>
<tr>
<td>http_x_forwarded_for</td>
<td>The x-forwarded-for content in the request message received by the proxy.</td>
</tr>
<tr>
<td>http_x_real_ip</td>
<td>The actual IP address of the client.</td>
</tr>
<tr>
<td>Field</td>
<td>Description</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>read_request_time</td>
<td>The time when the proxy reads the request message. Unit: ms.</td>
</tr>
<tr>
<td>request_length</td>
<td>The length of the request message, including the start-line, HTTP headers, and HTTP body.</td>
</tr>
<tr>
<td>request_method</td>
<td>The request method.</td>
</tr>
<tr>
<td>request_time</td>
<td>The interval between the time when the proxy receives the first request message and the time when the proxy returns a response message. Unit: seconds.</td>
</tr>
<tr>
<td>request_uri</td>
<td>The URI of the request message received by the proxy.</td>
</tr>
<tr>
<td>scheme</td>
<td>The request schema. Valid values: &quot;http&quot; and &quot;https&quot;.</td>
</tr>
<tr>
<td>server_protocol</td>
<td>The HTTP version received by the proxy, such as &quot;HTTP/1.0&quot; or &quot;HTTP/1.1&quot;.</td>
</tr>
<tr>
<td>slb_vport</td>
<td>The listening port of the SLB instance.</td>
</tr>
<tr>
<td>slbid</td>
<td>The ID of the SLB instance.</td>
</tr>
<tr>
<td>ssl_cipher</td>
<td>The cipher suite used, such as ECDHE-RSA-AES128-GCM-SHA256.</td>
</tr>
<tr>
<td>ssl_protocol</td>
<td>The protocol used to establish an SSL connection, such as TLSv1.2.</td>
</tr>
<tr>
<td>status</td>
<td>The status of the proxy response message.</td>
</tr>
<tr>
<td>tcpinfo_rtt</td>
<td>The TCP RTT of the client. Unit: microseconds.</td>
</tr>
<tr>
<td>time</td>
<td>The time when the log entry was created.</td>
</tr>
<tr>
<td>upstream_addr</td>
<td>The IP address and port number of the backend server.</td>
</tr>
<tr>
<td>upstream_response_time</td>
<td>The total time taken by the SLB instance to establish a connection to the backend server, receive data, and then close the connection. Unit: seconds.</td>
</tr>
<tr>
<td>upstream_status</td>
<td>The response status code of the backend server received by the proxy.</td>
</tr>
</tbody>
</table>
4.6 DDoS log collection

4.6.1 Overview

Log Service supports real-time collection of Alibaba Cloud Anti-DDoS Pro website access logs, CC attack logs, and supports real-time query and analysis of collected log data. The results of the query are displayed in the form of dashboards.

Functional advantages

- Simple configuration: Easily configure to capture real-time protected logs.
- Real-time analysis: Relying on Log Service, it provides real-time log analysis and out-of-box report center, that gives information about CC attack status and customer access details.
- Real-time alarms: Supports custom monitoring and alarms based on specific indicators in real time to provide timely response to critical business exceptions.
- Ecosystem: Supports the docking of other ecosystems, such as stream computing, cloud storage, and visualization solutions for the further data value exploration.
- FreeTier quota: Provides a free data import quota, and three days free log storage, query and real-time analysis. You can freely expand your storage time for compliance management, tracing, and filing. Support unlimited storage time, and the storage cost is 0.35 USD/GB per month.

Limits and instructions

- Exclusive Logstores do not support writing additional data.

Exclusive Logstore is used to store Anti-DDoS Pro website logs, so writing other data is not supported. There are no restrictions on other functions such as query, statistics, alarms, and streaming consumption.

- Pay-As-You-Go billing method If DDoS log collection protection is not enabled for any website, no charge appears.

DDoS log collection function is billed according to the charge item of Log Service. If DDoS log collection function is not enabled for any website, no charge
Log Service supports Pay-As-You-Go billing method, and provides FreeTier quota. For more information, see #unique_133.

Scenarios

- **Troubleshoot website access exceptions**

  Log Service has been configured to collect DDoS logs, you can query and analyze the collected logs in real time. Using SQL statement to analyze the DDoS access log, you can quickly check and analyze the website access exceptions, and view information such as read and write delays and operator distribution.

  For example, view the DDoS access log by using the following statement:

  ```sql
  __topic__: ddos_access_log
  ```

- **Track CC attack source**

  The distribution and source of CC attacks are recorded in the DDoS access log. By performing real-time query and analysis on the DDoS access log, you can conduct source tracking, trace CC attacks, and provide a reference for response strategy.

  For example, analyze the CC attack country distribution recorded in the DDoS access log by the following statement:

  ```sql
  __topic__: ddos_access_log and cc_blocks > 0| SELECT ip_to_country (if(real_client_ip='-', remote_addr, real_client_ip)) as country, count(1) as "number of attacks" group by country
  ```

- **For example, view the PV access by the following statement:**

  ```sql
  __topic__: ddos_access_log | select count(1) as PV
  ```

- **Website operation analysis**

  DDoS access log records the website access data in real time. You can perform SQL query analysis of the collected access log data to obtain real-time access status, such as determining the website popularity, the source and channel of the access, the client distribution, and assist in website operation analysis.

  For example, view the visitor traffic distribution from different network clouds:

  ```sql
  __topic__: ddos_access_log | select ip_to_provider(if(real_client_ip='-', remote_addr, real_client_ip)) as provider, round(sum(request_length)/1024.0/1024.0, 3) as mb_in group by provider having
  ```
4.6.2 Collection procedure

In the Anti-DDoS Pro console, you can enable DDoS log collection function for the website.

Prerequisites

1. Enable Anti-DDoS Pro function, purchase Anti-DDoS Pro instances, and Online configuration.
3. Activate Log Service.

Context

Log Service supports real-time collection of Alibaba Cloud Anti-DDoS Pro website access logs, CC attack logs, and supports real-time query and analysis of collected log data. The results of the query are displayed in the form of dashboards, and logs are used to analyze the access and attack behavior in real time, and assist the security department to formulate a protection strategy.

Procedure

1. Log on to the Anti-DDoS Pro console and select Log > Full Log in the left-side navigation pane. Enter the Full Log page.
2. If you are configuring DDoS log collection for the first time, follow the instructions on the page.
   DDoS has permission to distribute DDoS logs to your Logstore after authorization.
3. Select the website for which you want to enable DDoS log collection function and make sure the Status is on.

Figure 4-12: Enable the function
At this point, you have successfully enabled DDoS log collection for the current website. Log Service automatically creates a Logstore under your account. DDoS imports all the logs of the website that have this feature enabled into this Logstore. For Logstore default configurations, see Default configuration.

Table 4-9: Default configuration

<table>
<thead>
<tr>
<th>Default configuration item</th>
<th>Configuration content</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Project</strong></td>
<td>By default, ddos-pro-logstore project is created.</td>
</tr>
</tbody>
</table>
| **Logstore**               | By default, Logstore is created. Logstore name is determined by the domain of the DDoS you purchased.  
  - Other DDoS instances: ddos-pro-project-Alibaba Cloud Account ID-ap-southeast-1  
  All logs generated by the DDoS log collection function are saved in this Logstore. |
| **Region**                 |  
  - If the DDoS region is in mainland China, the default project is saved in China East 1.  
  - If the DDoS region is outside mainland China, the default project is saved in Asia Pacific SE 1. |
| **Shard**                  | By default, two shards are created and the Auto split shard feature is turned on. |
| **Log storage time**       | The default storage time is three days, within the free quota. After three days logs are automatically deleted.  
  For longer storage time, you can customize the configurations. For more information, see the How to modify the storage time of the website log section in unique_133. |
### Default configuration item

<table>
<thead>
<tr>
<th>Default configuration item</th>
<th>Configuration content</th>
</tr>
</thead>
</table>
| Dashboard                 | By default, two dashboards are created:  
  - ddos-pro-logstore_ddos_operation_center: Operation center  
  - ddos-pro-logstore_ddos_access_center: Access center  
For more information about dashboards, see Log Report. |

You can query and analyze the collected logs in real time on the current Full Log page. See the following figure for a log field description. In addition, Log Service creates two DDoS Operation center and Access center dashboards. You can also customize the dashboard configurations.

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>topic</strong></td>
<td>The topic of the log is fixed to ddos_access_log.</td>
<td>-</td>
</tr>
<tr>
<td>body_bytes_sent</td>
<td>Request to send the size of the Body. The unit is byte.</td>
<td>2</td>
</tr>
<tr>
<td>content_type</td>
<td>Content type.</td>
<td>application/x-www-form-urlencoded</td>
</tr>
<tr>
<td>host</td>
<td>Source website.</td>
<td>api.zhihu.com</td>
</tr>
<tr>
<td>http_cookie</td>
<td>Request cookie.</td>
<td>k1=v1;k2=v2</td>
</tr>
<tr>
<td>http_referer</td>
<td>Request referer. If none, the - is displayed.</td>
<td><a href="http://xyz.com">http://xyz.com</a></td>
</tr>
<tr>
<td>http_user_agent</td>
<td>User agent request.</td>
<td>Dalvik/2.1.0 (Linux; U; Android 7.0; EDI-AL10 Build/HUAWEIEDISON-AL10)</td>
</tr>
<tr>
<td>http_x_forwarded_for</td>
<td>The upstream user IP that is redirected by the proxy.</td>
<td>-</td>
</tr>
<tr>
<td>Field</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>https</td>
<td>Whether the request is an HTTPS request, wherein:</td>
<td>true</td>
</tr>
<tr>
<td></td>
<td>• true: the request is an HTTPS request.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• false: the request is an HTTP request.</td>
<td></td>
</tr>
<tr>
<td>matched_host</td>
<td>The source website of the matching configuration may be a pan-domain name.</td>
<td>*.zhihu.com</td>
</tr>
<tr>
<td></td>
<td>If not matching, the - is displayed.</td>
<td></td>
</tr>
<tr>
<td>real_client_ip</td>
<td>Access the customer real IP. If not available, the - is displayed.</td>
<td>1.2.3.4</td>
</tr>
<tr>
<td>isp_line</td>
<td>Line information, such as BGP, telecommunication, Unicom.</td>
<td>Telecommunication</td>
</tr>
<tr>
<td>remote_addr</td>
<td>Request client IP connection.</td>
<td>1.2.3.4</td>
</tr>
<tr>
<td>remote_port</td>
<td>Request client port connection.</td>
<td>23713</td>
</tr>
<tr>
<td>request_length</td>
<td>The length of the request. The unit is byte.</td>
<td>123</td>
</tr>
<tr>
<td>request_method</td>
<td>The HTTP request method.</td>
<td>GET</td>
</tr>
<tr>
<td>request_time_msec</td>
<td>Request time. The unit is microsecond.</td>
<td>44</td>
</tr>
<tr>
<td>request_uri</td>
<td>Request path.</td>
<td>/answers/377971214/banner</td>
</tr>
<tr>
<td>server_name</td>
<td>The matching host name. If not matching, the default is displayed.</td>
<td>api.abc.com</td>
</tr>
<tr>
<td>status</td>
<td>HTTP status code.</td>
<td>200</td>
</tr>
<tr>
<td>time</td>
<td>Time.</td>
<td>2018-05-02T16:03:59+08:00</td>
</tr>
<tr>
<td>Field</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>cc_action</td>
<td>CC protection policy, such as none, challenge, pass, close, captcha, wait, logon, n.</td>
<td>close</td>
</tr>
<tr>
<td>cc_blocks</td>
<td>Indicates whether CC protection is blocked, wherein:</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>• 1: Blocked.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Other codes: Passed.</td>
<td></td>
</tr>
<tr>
<td>cc_phase</td>
<td>CC protection policy, including seccookie, server_ip_blacklist, static_white_list, server_header_blacklist, server_cookie_blacklist, server_args_blacklist, qps_overmax.</td>
<td>server_ip_blacklist</td>
</tr>
<tr>
<td>ua_browser</td>
<td>Browser.</td>
<td>ie9</td>
</tr>
<tr>
<td>ua_browser_family</td>
<td>Browser series.</td>
<td>Internet explorer</td>
</tr>
<tr>
<td>ua_browser_type</td>
<td>Browser type.</td>
<td>web_browser</td>
</tr>
<tr>
<td>ua_browser_version</td>
<td>Browser version.</td>
<td>9.0</td>
</tr>
<tr>
<td>ua_device_type</td>
<td>Client device type.</td>
<td>computer</td>
</tr>
<tr>
<td>ua_os</td>
<td>Client operating system.</td>
<td>windows_7</td>
</tr>
<tr>
<td>ua_os_family</td>
<td>Client operating system series.</td>
<td>windows</td>
</tr>
<tr>
<td>upstream_addr</td>
<td>Return source address list, the format is IP:Port. Multiple addresses are separated by commas.</td>
<td>1.2.3.4:443</td>
</tr>
<tr>
<td>upstream_ip</td>
<td>The actual return source address IP.</td>
<td>1.2.3.4</td>
</tr>
<tr>
<td>upstream_response_time</td>
<td>The response time of the source. The unit is second.</td>
<td>0.044</td>
</tr>
<tr>
<td>upstream_status</td>
<td>Return source request HTTP status.</td>
<td>200</td>
</tr>
</tbody>
</table>
### Field Description Example

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>user_id</td>
<td>Alibaba Cloud user ID.</td>
<td>12345678</td>
</tr>
</tbody>
</table>

What's next

- Click Log Analysis, *Query Analysis* on the collected log data.
- Click Log Report to view the built-in *dashboard*.
- Click Advanced Management to go to Log Service console to query and collect statistics, stream consumption, and set alarms for the collected log data.

#### 4.6.3 Log analysis

Anti-DDoS Pro is embedded in the Full log page of Log Service in the Log analysis and Log report. After you have enabled the DDoS log protection function for a specific website, you can query and analyze the collected log data in real time on the current page, view or edit the dashboard, and set monitoring alarms.

**Procedure**

1. Log on to the Anti-DDoS Pro console, and select Log > Full log in the left-side navigation pane.
2. Select the website for which you want enable DDoS log collection protection, then confirm the Status is on.
3. Click Log analysis.

The current page is embedded in the Query analysis page of Log Service, and the system automatically enters the query statement for you, such as `matched_host: www.aliyun.com`, to view the log data based on the selected website.

Figure 4-13: Log analysis

4. Enter query analysis statement, select the log time range and click Query.

Note:
The default storage time of DDoS logs is three days. After three days, the log data is deleted. By default, you can only query log data for the past three days. To modify the log storage time, see Modify log storage time.

Figure 4-14: Log query

On the Query and Analysis page, you can also perform the following operations.

- **Custom query and analysis**

  Log Service provides different query and analysis syntaxes to support log queries in various complex scenarios. For more information, see Custom query and analysis.

- **View the log time distribution**

  Under the search box, the time distribution of the log matching the query time and the query statement is displayed. Time distribution is displayed in the form of a histogram with the horizontal and vertical axis. The total number of queried logs is displayed.

Note:
You can slide the histogram to select a smaller range of time zones, and the time picker automatically updates the selected time range and refresh the results.

Figure 4-15: Log time distribution

- View the raw logs
  
  In the Raw log, the details of each log are displayed in pagination, including time and content of these fields. You can sort the columns, download the current query results, or click the gear to select specific fields to be displayed.

  Click on the value or part of the corresponding field in the page to automatically enter the appropriate search criteria in the search box. For example, click the
value GET in request_method: GET, the following statement is automatically added to the search box:

```
Raw search statement and request_method: GET
```

Figure 4-16: Raw logs
• View analysis charts

Log Service supports graphical presentation of the analysis results, you can select different chart types on the Statistics Chart page. For more information, see Analysis charts.

Figure 4-17: Statistic chart
• Quick analysis

Quick analysis feature provides one-click interactive query that helps you quickly analyze the distribution of a field over a period of time and reduce the time cost of indexing critical data. For more information, see Quick analysis.

Figure 4-18: Quick analysis

Custom query analysis

Log query statement consists of two parts: query syntax (Search) and analysis syntax (Analytics), which are divided by |:

$Search | $Analytics

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query (Search)</td>
<td>The query conditions can be generated by keywords, fuzzy, numerical values, interval range and combination conditions. If left empty or *, all data is displayed.</td>
</tr>
<tr>
<td>Analysis (Analytics)</td>
<td>Calculate and count the query results or the full amount of data.</td>
</tr>
</tbody>
</table>

Note:
Both Search and Analytics are optional. If Search is empty, all the data in the specified period is not filtered and the results are counted directly. If Analytics is empty, the query results are returned and no statistics are collected.

Query syntax

Log Service query syntax supports Full-text query and Field query. Query box supports line break display, syntax highlighting, and other functions.

- Full-text query

You do not need to specify a field to enter the keyword query directly. You can wrap a keyword in double quotation marks (""), separated by a space or by and between multiple keywords.

Example

- Multiple keywords query

Search for logs containing www.aliyun.com and error. For example:

www.aliyun.com error

or

www.aliyun.com and error

- Conditional query

Search for logs containing www.aliyun.com and including error or 404. For example:

www.aliyun.com and (error or 404)

- Prefix query

Search for all keywords that contain www.aliyun.com and start with failed_. For example:

www.aliyun.com and failed_*

Note:
Query only supports suffix plus *, does not support prefix *, such as *_error.
Field query

Log Service supports more accurate queries based on fields.

A comparison of numeric type fields can be implemented in the format field :value or field>=value, using and, or. It can also be combined with full-text search, also by using the combination of and and or.

DDoS website access log and attack log can also base on field query. For the meaning, type, format and other information of each field, see DDoS log field.

Example

- Multiple fields query

Search for logs containing www.aliyun.com attacked by CC:

matched_host: www.aliyun.com and cc_blocks: 1

Search the access logs containing the error 404 of a client 1.2.3.4 on the website www.aliyun.com:

real_client_ip: 1.2.3.4 and matched_host: www.aliyun.com and status: 404

Note:
**Fields used in the examples** matched_host, cc_blocks, real_client_ip, and status are fields of DDoS access and attack logs. For more information about fields, see *DDoS log fields*.

- **Numeric field query**

  Search for all slow request logs with a response time of more than 5 seconds:

  \[ \text{request_time_msec} > 5000 \]

  Interval queries are also supported, querying logs with a response time greater than 5 seconds and less than or equal to 10 seconds:

  \[ \text{request_time_msec} \in [5000 \ 10000] \]

  The query can also be performed by the following statement:

  \[ \text{request_time_msec} > 5000 \text{ and request_time_msec} \leq 10000 \]

- **Check whether Japanese characters are used.**

  Query for the presence of specific fields:

  - Query logs in the **ua_browser** field: \[ \text{ua_browser}: * \]
  
  - Query logs that do not belong to the **ua_browser** field: \[ \text{not ua_browser}: * \]

  For more information about query syntax, see *Index and query*.

**Analysis syntax**

You can use the SQL/92 syntax for log data analysis and statistics. For more information about the syntax and functions supported by Log Service, see *#unique_142*.

**Note:**

- The **from table name** statement in the SQL standard syntax can be omitted from the analysis statement, that is, **from log**.

- Log data returns the first 100 entries by default, and you can modify the return range by *#unique_143*.

**Time-based log query analysis**

Each DDoS log has a **time** field, in the format **year-month-day T hour: minute : second + time zone**. For example, **2018-05-31T20:11:58+08:00**, where the
time zone is UTC+8, that is Beijing time. At the same time, each log has a built-in field: __time__, which also indicates the time of this log, so that time-based calculations can be performed in statistics. The format is Unix timestamp. The essence is a cumulative number of seconds since the 1970-1 0:0:0 UTC time. Therefore, in actual use, after calculation, time must be formatted before it can be displayed.

- **Select and show time**

  Over a specific period of time, select the latest 10 logs of the website www.aliyun.com attacked by CC, show the time, source IP and access client, using the time field directly:

  ```sql
  matched_host: www.aliyun.com and cc_blocks: 1
  | select time, real_client_ip, http_user_agent
  | order by time desc
  | limit 10
  ```

- **Calculation time**

  To query the number of days after the CC attack, use __time__ to calculate:

  ```sql
  matched_host: www.aliyun.com and cc_blocks: 1
  | select time,
  | round((to_unixtime(now()) - __time__) / 86400, 1) as "days_passed",
  | real_client_ip, http_user_agent
  | order by time desc
  | limit 10
  ```

  **Note:**

  Use `round((to_unixtime(now()) - __time__) / 86400, 1)`, first part `to_unixtime` the time obtained by `now()`, is converted to a Unix timestamp, and subtracted from the built-in time field __time__ to get the number of seconds that have passed. Finally, divide by 86400, which is the total number of seconds in a day, and then round it to the decimal with the function `round(data, 1)`. One-digit value indicates that each attack log has passed a few days.

- **Group statistics based on specific time**

  If you want to know how a website is being attacked by CC every day for a specific time frame, use the following SQL:

  ```sql
  matched_host: www.aliyun.com and cc_blocks: 1
  | select date_trunc('day', __time__) as dt,
  | count(1) as PV
  | group by dt
  ```
Log Service

Data Collection / 4 Cloud product collection

---

**Note:**

This example uses the built-in time field `__time__` to pass the function `date_trunc('day', ..)` to the time alignment. Each log is grouped into the partition of the day it belongs to for the total number of statistics (count(1)) and sorted by partition time block. The first argument of the function `date_trunc` provides alignment for other units, including `second`, `minute`, `hour`, `week`, `month`, `year`. For more information about function, see #unique_144.

- **Time-based group statistics**

For more flexible grouping time rules, for example, to know the trend of a website being attacked by CC every five minutes the math calculations are required. Run the following SQL:

```sql
matched_host: www.aliyun.com and cc_blocks: 1
| select from_unixtime(__time__ - __time__% 300) as dt,
    count(1) as PV
from logs
group by dt
order by dt
limit 1000
```

**Note:**

Use the built-in time field to calculate `__time__ - __time__% 300` and format it using the function `from_unixtime`. Each log is grouped into a 5 minute (300 seconds) partition for the total number of statistics (count(1)), and sorted by partition time block to obtain the first 1000 logs, which is equivalent to the first 83 hours of data in the selection time.

More time-resolved functions, such as converting a time format, require using `date_parse` and `date_format`. For more information, see #unique_144.

**Client IP-based query analysis**

DDoS log has a field `real_client_ip`. However, if the user cannot obtain the real IP by the proxy and the IP address in the header is incorrect, you can use the `remote_addr` field to directly connected to the client IP.

- **Country attack distribution**

Distribution of source countries of CC attacks on a website:

```sql
matched_host: www.aliyun.com and cc_blocks: 1
```

---

Issue: 20200323
| SELECT ip_to_country(if(real_client_ip='-', remote_addr, real_client_ip)) as country, count(1) as "number of attacks" group by country

**Note:**

*Use the function* \( \text{if}(\text{condition}, \text{option1}, \text{option2}) \) **to select the field** \( \text{real_client_ip} \) or \( \text{real_client_ip} \) (when \( \text{real_client_ip} \) is -). Pass the obtained IP to the function \( \text{ip_to_country} \) **to get the country information corresponding to this IP.**

- **Access distribution**

To get more detailed province-based distribution, use the \( \text{ip_to_province} \) function, for example:

```sql
SELECT ip_to_province(if(real_client_ip='-', remote_addr, real_client_ip)) as province, count (1) as "number of attacks" group by province
```

**Note:**

Another IP function \( \text{ip_to_province} \) **to get a province of IP. If IP address is outside of China, system still tries to convert to the province (state).**

- **Attackers heat distribution**

To get an attackers heat map, use the \( \text{ip_to_geo} \) function, for example:

```sql
SELECT ip_to_geo(if(real_client_ip='-', remote_addr, real_client_ip)) as geo, count (1) as "number of attacks" group by geo limit 10000
```

**Note:**

*Use another IP function* \( \text{ip_to_geo} \) **to get the latitude and longitude of an IP and get the first 10,000.**

More IP-based parsing functions, such as obtaining the IP operator \( \text{ip_to_provider} \), determining whether the IP is Internet or Intranet \( \text{ip_to_domain} \), see #unique_145.
4.6.4 Log Report

Log Reports page is embedded in the dashboard of the Log Service. This page displays the default dashboard. You can view dashboard data under various filter conditions by modifying the time range and adding filters.

View reports

1. Log on to the Anti-DDoS Pro console and select Log > Full Log in the left-side navigation pane. Enter the Full Log page.
2. Select the website for which you want to enable DDoS log collection function and make sure the Status is on.
3. Click Log Reports.

   Dashboard page of Log Service is embedded in the current page, and the filter condition is automatically added. For example, use `matched_host: www.aliyun.com` to view log reports based on selected website.

Figure 4-19: View reports

After the DDoS log collection function is enabled for the website, Log Service automatically creates two default instruments for reporting: operation center and access center. For more information about the default dashboard, see Default dashboard.

<table>
<thead>
<tr>
<th>Dashboard</th>
<th>Dashboard name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ddos-pro-logstore_ddos_operation_center</td>
<td>DDoS operation center</td>
<td>Displays the current overall operational status of DDoS protected websites, including valid request status, traffic, trends, attack distributions, and traffic volumes and peaks attacked by CC.</td>
</tr>
</tbody>
</table>
### Dashboard

<table>
<thead>
<tr>
<th>Dashboard name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ddos-pro-logstore_ddos_access_center</td>
<td>Displays the current overall operational status of DDoS protected websites, including PV/UV trends and bandwidth peaks, visitors, traffic, client type, request, and visited websites distribution.</td>
</tr>
</tbody>
</table>

Figure 4-20: Default dashboard

Besides viewing the report, the following operations can be performed:

- **Select time range**
- **Add or edit filter condition**
- **View charts**

**Time picker**

All charts on the dashboard page are based on statistical results for different time periods. For example, the default time range for visits is one day and the access...
trend is 30 days. To set all charts on the current page to be displayed in the same time range, you can configure the time picker.

1. Click Select.
2. Configure the settings in the dialog box. You can select relative time, entire point time, or set a custom time.

Note:

- When the time range is modified, the time of all charts is changed to this time range.
- Time picker only provides a temporary view of the chart on the current page, and the system does not save the setting. The next time you view the report, the system will display the default time range.

Filter conditions

Select the website and click Log Reports to enter the dashboard page. System automatically adds filter condition, such as matched_host: www.aliyun.com to view log reports based on selected website.

You can modify the data display range of the report by setting filter condition.
· View overall reports for all websites

Clear the filter condition to display the overall reports library `ddos-pro-logstore`.

· Add more filter conditions

You can filter the report data by setting key and value. AND relationship between multiple filters is supported.

For example, view the overall situation of access requests by telecommunications lines.

Figure 4-22: Add filter conditions

---

**Note:**

The `isp_line` is the field of the DDoS log, indicating the operator network connecting to the port. For more information about fields, see [DDoS log fields](#unique_146).

**Chart type**

The report display area shows multiple reports according to a predefined layout, including the following types. For more information about chart types, see [#unique_146](#unique_146).

<table>
<thead>
<tr>
<th>Chart type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>Displays important indicators, such as effective request rate, and attack peaks.</td>
</tr>
<tr>
<td>Line/area map</td>
<td>Displays trend graphs for certain important indicators within a specific time period, such as inbound bandwidth trends and attack interception rates.</td>
</tr>
<tr>
<td>Map</td>
<td>Displays the geographical distribution of visitors and attackers, such as CC attack country, access hotspot.</td>
</tr>
<tr>
<td>Pie chart</td>
<td>Displays the distribution of the information, such as the top 10 of the websites being attacked, client type distribution.</td>
</tr>
<tr>
<td>Chart type</td>
<td>Description</td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Table</td>
<td>Displays information such as the list of attackers, typically divided into multiple columns.</td>
</tr>
<tr>
<td>Maps</td>
<td>Displays the geographical distribution of the data.</td>
</tr>
</tbody>
</table>

Default dashboards

- Operation center

  Operations center displays the current overall operational status of DDoS protected websites, including valid request status, traffic, trends, attacker distributions, and traffic volumes and peaks attacked by CC.

<table>
<thead>
<tr>
<th>Chart</th>
<th>Type</th>
<th>Default time range</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid request package rate</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>A valid request, that is, the number of non-CC attacks or 400 error requests in the total number of all requests.</td>
<td>95%</td>
</tr>
<tr>
<td>Valid request flow rate</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>Valid request percentage of the total flow of all requests.</td>
<td>95%</td>
</tr>
<tr>
<td>Received traffic</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>The sum of valid request inflows. The unit is MB.</td>
<td>300 MB</td>
</tr>
<tr>
<td>Attack traffic</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>The sum of inbound traffic of CC attacks. The unit is MB.</td>
<td>30 MB</td>
</tr>
<tr>
<td>Chart</td>
<td>Type</td>
<td>Default time range</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>------------------</td>
<td>--------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Outbound traffic</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>The sum of valid request outbound traffic. The unit is MB.</td>
<td>300 MB</td>
</tr>
<tr>
<td>Network in bandwidth peak.</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>The highest peak of incoming traffic rate requested by the website. The unit is bytes/s.</td>
<td>100 Bytes/s</td>
</tr>
<tr>
<td>Network out bandwidth peak.</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>The highest peak of outbound traffic rate requested by the website. The unit is bytes/s.</td>
<td>100 Bytes/s</td>
</tr>
<tr>
<td>Received data packets</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>The number of incoming requests for valid requests (non-CC attacks), measured in units.</td>
<td>30,000</td>
</tr>
<tr>
<td>Attack data packets</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>The sum of the number of requests for the CC attack, measured in units.</td>
<td>100</td>
</tr>
<tr>
<td>Chart</td>
<td>Type</td>
<td>Default time range</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>----------------------------</td>
<td>--------------------</td>
<td>--------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Attack peak</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>The highest peak of CC attack. The unit is number per minute.</td>
<td>100 per minute</td>
</tr>
<tr>
<td>Inbound bandwidth and attack trends</td>
<td>Two-line diagram</td>
<td>1 hour (entire point)</td>
<td>Trend chart of valid requests per minute and traffic bandwidth for attack requests. The unit is KB/s.</td>
<td>-</td>
</tr>
<tr>
<td>Request and interception trends</td>
<td>Two-line diagram</td>
<td>1 hour (entire point)</td>
<td>Trend chart of the total number of requests and intercepted CC attack requests per minute. The unit is number per minute.</td>
<td>-</td>
</tr>
<tr>
<td>Valid request rate trend</td>
<td>Two-line diagram</td>
<td>1 hour (entire point)</td>
<td>Trend chart of the number of valid requests per minute (non-CC attacks or 400 error requests) in the total number of all requests.</td>
<td>-</td>
</tr>
<tr>
<td>Chart</td>
<td>Type</td>
<td>Default time range</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>------------------------------</td>
<td>------------------</td>
<td>--------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Access status distribution trend</td>
<td>Flow chart</td>
<td>1 hour (entire point)</td>
<td>Trend chart of various request processing statuses (400, 304, 20) per minute. The unit is number per minute.</td>
<td>-</td>
</tr>
<tr>
<td>CC attacks distribution</td>
<td>World map</td>
<td>1 hour (relative)</td>
<td>The sum of the number of CC attacks in the source country.</td>
<td>-</td>
</tr>
<tr>
<td>CC attack distribution</td>
<td>Map of China</td>
<td>1 hour (relative)</td>
<td>The sum of the number of CC attacks in the source province (China).</td>
<td>-</td>
</tr>
<tr>
<td>List of attacks</td>
<td>Table</td>
<td>1 hour (relative)</td>
<td>The attacker information of the first 100 attacks, including IP, city, network, number of attacks, and total traffic.</td>
<td>-</td>
</tr>
<tr>
<td>Attack access line distribution</td>
<td>Pie chart</td>
<td>1 hour (relative)</td>
<td>CC attack source access DDoS protection line distribution, such as telecommunications, Unicom, and BGP.</td>
<td>-</td>
</tr>
</tbody>
</table>
### Access center

Access center displays the current overall operational status of DDoS protected websites, including PV/UV trends and bandwidth peaks, visitors, traffic, client type, request, and visited websites distribution.

<table>
<thead>
<tr>
<th>Chart</th>
<th>Type</th>
<th>Default time range</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Page view</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>The total number of requests.</td>
<td>100,000</td>
</tr>
<tr>
<td>Unique visitors</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>Total number of independent access clients.</td>
<td>100,000</td>
</tr>
<tr>
<td>Inbound traffic</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>The sum of inbound traffic of the website. The unit is MB.</td>
<td>300 MB</td>
</tr>
<tr>
<td>Network in bandwidth peak.</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>The highest peak of inbound traffic rate requested by the website. The unit is bytes/s.</td>
<td>100 Bytes/s</td>
</tr>
<tr>
<td>Chart</td>
<td>Type</td>
<td>Default time range</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------------</td>
<td>--------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Network out bandwidth peak.</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>The highest peak of inbound traffic rate requested by the website. The unit is bytes/s.</td>
<td>100 Bytes/s</td>
</tr>
<tr>
<td>Traffic bandwidth trend</td>
<td>Two-line diagram</td>
<td>1 hour (entire point)</td>
<td>Trend chart of website inbound and outbound traffic per minute. The unit is KB/s.</td>
<td>-</td>
</tr>
<tr>
<td>Request and interception trends</td>
<td>Two-line diagram</td>
<td>1 hour (entire point)</td>
<td>Trend chart of the total number of requests and intercepted CC attack requests per minute. The unit is number per minute.</td>
<td>-</td>
</tr>
<tr>
<td>PV/UV access trends</td>
<td>Two-line diagram</td>
<td>1 hour (entire point)</td>
<td>Trend chart of PV and UV per minute. Measured in units.</td>
<td>-</td>
</tr>
<tr>
<td>Visitor distribution</td>
<td>World map</td>
<td>1 hour (relative)</td>
<td>The distribution of visitors PV (page view) in the source country.</td>
<td>-</td>
</tr>
<tr>
<td>Visitor heat map</td>
<td>Amap</td>
<td>1 hour (relative)</td>
<td>Visitor geographic access heat map.</td>
<td>-</td>
</tr>
<tr>
<td>Chart</td>
<td>Type</td>
<td>Default time range</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>------------------------------------</td>
<td>---------------------</td>
<td>--------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Inbound traffic distribution</td>
<td>World map</td>
<td>1 hour (relative)</td>
<td>Sum of inbound traffic distribution in the source country. The Unit is MB.</td>
<td>-</td>
</tr>
<tr>
<td>Inbound traffic distribution</td>
<td>Map of China</td>
<td>1 hour (relative)</td>
<td>Sum of inbound traffic distribution in the source province. The Unit is MB.</td>
<td>-</td>
</tr>
<tr>
<td>Access line distribution</td>
<td>Donut chart</td>
<td>1 hour (relative)</td>
<td>Source-based access DDoS protection line distribution, such as telecommunications, Unicom, and BGP.</td>
<td>-</td>
</tr>
<tr>
<td>Inbound traffic network provider</td>
<td>Donut chart</td>
<td>1 hour (relative)</td>
<td>The distribution of inbound traffic that visitors access by network operators. For example, telecommunications, Unicom, mobile connections, education network. The Unit is MB.</td>
<td>-</td>
</tr>
<tr>
<td>Chart</td>
<td>Type</td>
<td>Default time range</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>----------------------------</td>
<td>--------------</td>
<td>--------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Most visited clients</td>
<td>Table</td>
<td>1 hour (relative)</td>
<td>The top 100 most visited clients, including IP, city, network, request method distribution, incoming traffic, number of incorrect accesses, number of intercepted CC attacks.</td>
<td>-</td>
</tr>
<tr>
<td>Access domain name</td>
<td>Donut chart</td>
<td>1 hour (relative)</td>
<td>The top 20 most visited domain names.</td>
<td>-</td>
</tr>
<tr>
<td>Referer</td>
<td>Table</td>
<td>1 hour (relative)</td>
<td>The top 100 most redirected referer URLs, hosts, and frequency.</td>
<td>-</td>
</tr>
<tr>
<td>Client type distribution</td>
<td>Donut chart</td>
<td>1 hour (relative)</td>
<td>The top 20 most visited user agents, such as iPhone, iPad, Windows IE, Chrome.</td>
<td>-</td>
</tr>
</tbody>
</table>
4.7 Logs of BGP-line Anti-DDoS Pro

4.7.1 Overview

Alibaba Cloud Anti-DDoS Pro provides BGP bandwidth resources that are exclusively available in Chinese Mainland to mitigate volumetric DDoS attacks. The service can scrub Terabits of attack traffic per second based on eight ISPs. Compared with earlier versions, Anti-DDoS Pro supports more reliable networks with less latency and provides quicker disaster recovery.

Background

Security is always a challenge facing the Internet. Network threats represented by DDoS attacks have a serious impact on network security.

DDoS attacks are becoming more large-scale, mobile, and global. According to recent survey reports, the frequency of DDoS attacks is on the rise. Attackers are difficult to detect and can manipulate a large number of cloud service providers with poor security measures, IDCs, and even cameras to launch attacks. The attackers have formed a mature black industry chain and become increasingly organized. At the same time, the attack methods develop towards polarization. The proportion of slow and hybrid attacks, especially HTTP flood attacks, is increasing, which makes detection and defense more difficult. On the one hand, attacks peaking at 1 Tbit/s or higher become common, and the number of 100 GB attacks is multiplied. On the other hand, application layer attacks are also doubled.

According to Kaspersky Lab DDoS Q1 2018 Intelligence Report, China remains the main source and target of DDoS attacks. The main industries under attack include the Internet, games, software companies, and financial companies. More than 80% of

<table>
<thead>
<tr>
<th>Chart</th>
<th>Type</th>
<th>Default time range</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Request content type distribution</td>
<td>Donut chart</td>
<td>1 hour (relative)</td>
<td>The top 20 most requested content types, such as HTML, Form, JSON, streaming data.</td>
<td>-</td>
</tr>
</tbody>
</table>
DDoS attacks mix with HTTP attacks, and HTTP flood attacks are especially difficult to detect. Therefore, it is particularly important to analyze and study access and attack activities through logs and set protection policies accordingly.

Log Service can collect website access logs and HTTP flood attack logs of Alibaba Cloud Anti-DDoS Pro in real time. Log Service also supports real-time retrieval and analysis of the collected log data and displays the query results in the form of dashboards.

Benefits

• Simple configuration: Real-time anti-DDoS logs can be collected with simple configuration. Log collection is automatically enabled for new websites after they are added.

• Real-time analysis: Relying on Log Service, this function provides real-time log analysis and an out-of-the-box report center. This helps you understand the HTTP attack status and customer access details.

• Real-time alerts: Supports real-time monitoring and alerts based on customized metrics to ensure timely response to critical business exceptions.

• Collaboration: This function can be integrated with real-time computing, cloud storage, visualization, and other data solutions to discover more data value.

• Free quota: Provides a free data import quota of 3 TB and also allows you to use the log storage, query, and real-time analysis functions for 30 days for free.

Restrictions and guidelines

• Additional data cannot be written to the exclusive Logstore.

The exclusive Logstore is used to store Anti-DDoS Pro website logs. Writing other data is not supported. Other functions such as query, statistics, alerts, and streaming consumption are not restricted.

• The data TTL and total storage capacity of the exclusive Logstore cannot be modified.

Purchase the storage capacity of the exclusive Logstore based on your business requirements. Up to 1,000 TB of storage capacity is supported, and logs can be stored for up to 180 days.
Scenarios

- Troubleshoot website access exceptions

After you configure Log Service to collect Anti-DDoS Pro logs, you can query and analyze the collected logs in real time. By using SQL statements to analyze the Anti-DDoS Pro access logs, you can quickly troubleshoot and analyze the website access exceptions. You can also query information such as read and write delays and exception distribution by ISP.

For example, use the following statement to query the access log entries of Anti-DDoS Pro:

```
__topic__: DDoS_access_log
```

The query results are displayed, as shown in the following figure.

Figure 4-23: Access log entries of Anti-DDoS Pro

- Track HTTP flood attack sources

Anti-DDoS Pro access logs record information about the sources and distribution of HTTP flood attacks. You can query and analyze access logs in real time.
to identify the attackers, and use this information to select the most effective protection policy.

For example, use the following statement to analyze the geographical distribution of HTTP flood attacks:

```sql
SELECT ip_to_country((if(real_client_ip='-', remote_addr, real_client_ip)) as country, count(1) as "number" group by country
```

The analysis results are displayed in a dashboard as follows:

Figure 4-24: HTTP flood attacks
- For example, use the following statement to view PVs:

```
__topic__/ DDoS_access_log | select count(1) as PV
```

The analysis results are displayed in a dashboard, as shown in the following figure.

Figure 4-25: PV access

- Analyze website operations

Anti-DDoS Pro access logs record information about website traffic in real time. You can use SQL queries to analyze log data to better understand your visitors and analyze website operations. For example, you can identify the most visited Web pages, the browsers that initiated the requests, and the clients, source IP addresses, and ISPs of the requests.

For example, use the following statement to view the visitor distribution by ISP:

```
__topic__/ ddos_access_log | select ip_to_provider(if(real_client_ip='-', remote_addr, real_client_ip)) as provider, round(sum(request_length)/1024.0/1024.0, 3) as mb_in group by provider having
```

Issue: 20200323
The analysis results are displayed in a dashboard, as shown in the following figure.

Figure 4-26: Visitor distribution

4.7.2 Enable or disable log collection

When you purchase the Anti-DDoS Pro service, the log collection function is automatically enabled. You can disable or re-enable the log collection function of the specified website.

Context

Log Service can collect website access logs and HTTP flood attack logs of Alibaba Cloud Anti-DDoS Pro in real time. Log Service also supports real-time retrieval and analysis of the collected log data and displays the query results in the form of
Log Service

By analyzing the access and attack activities in real time through logs, Log Service helps the security department set protection policies.

Procedure

1. Log on to the Anti-DDoS Pro console, and choose Statistics > Full Log in the left-side navigation pane.

2. Select the website for which you want to enable log collection, turn on or off the Status switch.

By default, log collection is enabled when you purchase Anti-DDoS Pro. This function is also enabled for newly added websites.

Log Service automatically creates an exclusive project and an exclusive Logstore under your account. The Anti-DDoS Pro logs of all websites that have log collection enabled are imported to this exclusive Logstore. For more information about the default configurations of the exclusive project and Logstore, see the following table.

Table 4-10: Default configuration

<table>
<thead>
<tr>
<th>Default configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project</td>
<td>By default, a project is created. The name of the project is doscoo-project-Alibaba Cloud account ID-cn-hangzhou.</td>
</tr>
<tr>
<td>Default configuration item</td>
<td>Description</td>
</tr>
<tr>
<td>----------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td><strong>Logstore</strong></td>
<td>By default, a Logstore is created. The name of the Logstore is <code>ddoscoo-logstore</code>. All log entries generated by the Anti-DDoS Pro log collection function are stored in this Logstore.</td>
</tr>
<tr>
<td><strong>Region</strong></td>
<td>By default, the project is created under the China (Hangzhou) region.</td>
</tr>
<tr>
<td><strong>Shard</strong></td>
<td>By default, two shards are created and the <em>automatic shard splitting</em> function is enabled.</td>
</tr>
<tr>
<td><strong>Log TTL</strong></td>
<td>A log entry can be stored for 30 days. After 30 days, the log entry is automatically deleted.</td>
</tr>
<tr>
<td><strong>Log storage capacity</strong></td>
<td>You can purchase the storage capacity of the exclusive Logstore based on your business requirements. The maximum storage capacity is 1,000 TB, and log entries can be stored for up to 180 days.</td>
</tr>
</tbody>
</table>
| **Dashboard**              | By default, the following dashboards are created:  
  - Access center: displays website access metrics, client distribution, traffic, and performance data.  
  - Operations and maintenance center: displays attack status and operational metrics such as PV, UV, and success rate.  
For more information about dashboards, see *Log reports*. |

You can query and analyze collected log entries in real time on the Log Service page. For more information about log fields, see *Log fields*. In addition, apart from the operations and maintenance center and access center that are created by Log Service, you can also customize a dashboard.

**What's next**

- Click Search & Analysis to *query and analyze* the collected log data.
- Click Log Reports to view built-in *Log reports*. 
Click Advanced Management to go to the Log Service console. You can query log data, collect statistics, consume streaming data, and set alerts for the collected log data.

4.7.3 Manage log storage space

After you enable the Anti-DDoS Pro log collection function, log data is pushed to the specified Logstore in real time. You can view the usage of the log storage space in the Anti-DDoS Pro console.

View log storage usage

You can view the usage of the log storage for Anti-DDoS Pro Log Service at any time.

Note:
The log storage information in the console is not updated in real time. It takes up to two hours to synchronize the actual storage information to the console. Therefore, we recommend that you expand the log storage space before it is exhausted.

1. Log on to the Anti-DDoS Pro console.
2. In the left-side navigation pane, choose Statistics > Full Log.
3. In the upper-right corner of the page, view the log storage usage.
4.7.4 Log fields

This topic describes the supported fields of Anti-DDoS Pro log entries.

You can go to the Log Service page to query and analyze collected logs in real time.

For more information about log fields, see the following figure.

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>topic</strong></td>
<td>The topic of the log entry. The value of this field is fixed to ddos_access_log.</td>
<td>N/A</td>
</tr>
<tr>
<td>body_bytes_sent</td>
<td>The size of the body in the access request, in bytes.</td>
<td>2</td>
</tr>
<tr>
<td>content_type</td>
<td>The content type.</td>
<td>application/x-www-form-urlencoded</td>
</tr>
<tr>
<td>host</td>
<td>The source website.</td>
<td>api.zhihu.com</td>
</tr>
<tr>
<td>http_cookie</td>
<td>The request cookie.</td>
<td>k1=v1;k2=v2</td>
</tr>
<tr>
<td>http_referer</td>
<td>The request referer. If no referer exists, a hyphen (−) is displayed.</td>
<td><a href="http://xyz.com">http://xyz.com</a></td>
</tr>
<tr>
<td>http_user_agent</td>
<td>The User-Agent of the request.</td>
<td>Dalvik/2.1.0 (Linux; U; Android 7.0; EDI-AL10 Build/HUAWEIEDISON-AL10)</td>
</tr>
<tr>
<td>http_x_forwarded_for</td>
<td>The IP address of the upstream user redirected by proxy.</td>
<td>N/A</td>
</tr>
<tr>
<td>https</td>
<td>Indicates whether the request is an HTTPS request.</td>
<td>true</td>
</tr>
<tr>
<td></td>
<td>• true: The request is an HTTPS request.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• false: The request is an HTTP request.</td>
<td></td>
</tr>
<tr>
<td>matched_host</td>
<td>The matching source site, which may be a wildcard domain name. If no match is found, a hyphen (−) is displayed.</td>
<td>*.zhihu.com</td>
</tr>
<tr>
<td>Field</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>real_client_ip</td>
<td>The real IP of the visitor. If no real IP is returned, a hyphen (−) is returned.</td>
<td>1.2.3.4</td>
</tr>
<tr>
<td>isp_line</td>
<td>Line information, such as BGP, China Telecom, and China Unicom.</td>
<td>China Telecom</td>
</tr>
<tr>
<td>remote_addr</td>
<td>The IP address of the client that initiates the connection request.</td>
<td>1.2.3.4</td>
</tr>
<tr>
<td>remote_port</td>
<td>The port number of the client that initiates the connection request.</td>
<td>23713</td>
</tr>
<tr>
<td>request_length</td>
<td>The size of the request in bytes.</td>
<td>123</td>
</tr>
<tr>
<td>request_method</td>
<td>The HTTP method of the request.</td>
<td>GET</td>
</tr>
<tr>
<td>request_time_msec</td>
<td>The request time in ms.</td>
<td>44</td>
</tr>
<tr>
<td>request_uri</td>
<td>The request URI.</td>
<td>/answers/377971214/banner</td>
</tr>
<tr>
<td>server_name</td>
<td>The name of the matching host. If no match is found, the value is default.</td>
<td>api.abc.com</td>
</tr>
<tr>
<td>status</td>
<td>The HTTP status code.</td>
<td>200</td>
</tr>
<tr>
<td>time</td>
<td>The time when the log entry was generated.</td>
<td>2018-05-02T16:03:59+08:00</td>
</tr>
<tr>
<td>cc_action</td>
<td>The HTTP flood protection action. Valid values include none, challenge, pass, close, captcha, wait, and login.</td>
<td>close</td>
</tr>
<tr>
<td>cc_blocks</td>
<td>Indicates whether HTTP flood attacks are blocked. Valid values:</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>• 1: block</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Other values: pass</td>
<td></td>
</tr>
<tr>
<td>Field</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------------------------------------------</td>
</tr>
<tr>
<td>cc_phase</td>
<td>The HTTP flood protection policy. Valid values include seccookie, server_ip_blacklist, static_whitelist, server_header_blacklist, server_cookie_blacklist, server_args_blacklist, and qps_overmax.</td>
<td>server_ip_blacklist</td>
</tr>
<tr>
<td>ua_browser</td>
<td>The browser.</td>
<td>ie9</td>
</tr>
<tr>
<td>ua_browser_family</td>
<td>The browser series.</td>
<td>internet explorer</td>
</tr>
<tr>
<td>ua_browser_type</td>
<td>The browser type.</td>
<td>web_browser</td>
</tr>
<tr>
<td>ua_browser_version</td>
<td>The browser version.</td>
<td>9.0</td>
</tr>
<tr>
<td>ua_device_type</td>
<td>The type of the client device.</td>
<td>computer</td>
</tr>
<tr>
<td>ua_os</td>
<td>The operating system of the client.</td>
<td>windows_7</td>
</tr>
<tr>
<td>ua_os_family</td>
<td>The operating system series of the client.</td>
<td>windows</td>
</tr>
<tr>
<td>upstream_addr</td>
<td>The list of origin addresses that are separated with commas (,). Each address is in the format of IP:Port.</td>
<td>1.2.3.4:443</td>
</tr>
<tr>
<td>upstream_ip</td>
<td>The real origin IP address.</td>
<td>1.2.3.4</td>
</tr>
<tr>
<td>upstream_response_time</td>
<td>The response time in seconds for the back-to-origin process.</td>
<td>0.044</td>
</tr>
<tr>
<td>upstream_status</td>
<td>The HTTP status of the back-to-origin request.</td>
<td>200</td>
</tr>
<tr>
<td>user_id</td>
<td>The user ID of the Alibaba Cloud account.</td>
<td>12345678</td>
</tr>
<tr>
<td>querystring</td>
<td>The request string.</td>
<td>token=bbcd&amp;abc=123</td>
</tr>
</tbody>
</table>
4.7.5 Log analysis

The Search & Analysis page of Log Service is embedded in the Full Log page of the new Anti-DDoS Pro console. You can switch between the Full Log and Log Reports pages. After you enable the log analysis feature of Anti-DDoS Pro, you can query and analyze collected logs in real time, view and edit dashboards, and set alert rules.

Procedure

1. Log on to the new Anti-DDoS Pro console. In the left-side navigation pane, choose Statistics > Full Log.
2. Select the website domain for which you want to view log reports, and ensure that the Status switch is turned on.
3. Click Full Log.

The Search & Analysis page of Log Service is embedded in the Full Log page. The system displays logs of the target website based on a default filtering condition, such as matched_host:"0523.yuanya.aliyun.com".

Figure 4-28: Full log

4. Enter a statement, select a time range, and then click Search & Analysis.

Note:
Logs of Anti-DDoS Pro are stored for 180 days. By default, you can only query logs of the past 180 days.

Figure 4-29: Query logs

On the Search & Analysis page, you can customize query and analysis.

- Search and process logs by using built-in syntax

  Log Service supports query syntax and analysis syntax for log query in complex scenarios. For more information, see Search and process logs by using built-in syntax in this topic.

- View the distribution of log entries by time

  The histogram under the search box displays the distribution of log entries that match both the statement and the time range. The horizontal axis indicates the time and the vertical axis indicates the number of log entries. The total number of queried log entries is displayed below the histogram.

  Note:
You can drag the mouse pointer in the histogram to narrow down the time range. The time picker automatically updates the time range, and the query results are automatically updated.

Figure 4-30: Distribution of log entries by time

- **View raw logs**

  On the Raw Logs tab page, each log entry is detailed on an individual page, which displays the time when the log is generated and the log content. The log content contains fields and their values. You can sort log entries, download logs, and click Column Settings to select column items to be displayed.

  You can click a field value or part of the field value in log content, and then a corresponding condition is automatically specified in the search box. For
example, if you click GET in the request_method: GET field, the following statement is automatically generated in the search box:

```
<The original search statement> and request_method: GET
```

Figure 4-31: Raw logs

- **View analysis charts**

Log Service supports displaying analysis results in charts. You can select a chart type as needed on the Graph tab page. For more information, see *Charts*.

Figure 4-32: Analysis charts
· Quick analysis

The quick analysis feature provides you with an easy-to-use interactive experience. It enables you to analyze the distribution of a field in a specified time range. This feature can reduce the time used for indexing critical data. For more information, see Quick analysis.

Figure 4-33: Quick analysis

**Search and process logs by using built-in syntax**

A statement consists of a query clause (Search) and an analysis clause (Analytics), which are separated with a vertical bar (|).

$\text{Search} \mid \text{Analytics}$

<table>
<thead>
<tr>
<th>Clause</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Query</strong></td>
<td>A query clause can contain keywords, strings, numbers, value ranges, or a combination of them. If the clause is not specified or only contains an asterisk (*), the search result includes all log entries.</td>
</tr>
<tr>
<td><strong>Analysis</strong></td>
<td>You can use the analysis clause to process query results.</td>
</tr>
</tbody>
</table>

**Note:**
Both query and analysis clauses are optional. If no query clause is specified, the query result includes all log entries in the specified time range, and all log entries are processed based on the analysis clause. If no analysis clause is specified, the query results are returned without being processed.

Query syntax

The query feature of Log Service supports full text search and search by field. Statements in the search box can be displayed in multiple lines and highlighted.

• Full text search

You can enter keywords to search for all log entries without specifying field names. To use multiple keywords, you can enclose each keyword within quotation marks (") and separate them with spaces or and.

Examples:

- Multi-keyword search

You can use the following statements to search for log entries that contain www.aliyun.com and error. Example:

www.aliyun.com error

Alternative:

www.aliyun.com and error

- Conditional search

You can use the following statement to search for log entries that contain www.aliyun.com and error, or log entries that contain www.aliyun.com and 404.

Example:

www.aliyun.com and (error or 404)

- Prefix search

You can use the following statement to search for log entries that contain both www.aliyun.com and keywords starting with failed_. Example:

www.aliyun.com and failed_*

Note:
You can only add an asterisk (\*) as a suffix and you cannot add an asterisk (\*) as a prefix. For example, the statement cannot be \*_error.

- Search by field

Log Service supports accurate queries based on fields.

You can use the comparison of numeric fields in the format of field: value or field>=value and separate filtering conditions with and and or. You can use this feature together with full-text search and separate filtering conditions with and and or.

Website access logs and attack logs of Anti-DDoS Pro also support search by field. For more information about the description, type, and format of each log field, see Log fields.

Examples:

- Search by specifying multiple fields

You can use the following statement to search for all log entries that record CC attacks on www.aliyun.com:

matched_host: www.aliyun.com and cc_blocks: 1

You can search for all log entries that record visits with 404 errors from a specific client to www.aliyun.com. In the following example, the client IP address is 10.2.3.4.

real_client_ip: 10.2.3.4 and matched_host: www.aliyun.com and status: 404

Note:
In the preceding examples, matched_host, cc_blocks, real_client_ip, and status are fields defined in access and attack logs of Anti-DDoS Pro. For more information about log fields, see Log fields.

- Search by specifying numeric fields

You can use the following statement to search for log entries where the response time exceeds 5 seconds:

```
request_time_msec > 5000
```

You can search for log entries by specifying a value range. In the following example, the response time is greater than 5 seconds, and is less than or equal to 10 seconds:

```
request_time_msec in (5000 10000]
```

You can also use the following statement:

```
request_time_msec > 5000 and request_time_msec <= 10000
```

- Check whether a specific field exists

You can use the following statements to check whether a specific field exists:

- Search for log entries that contain the `ua_browser` field: `ua_browser: *`
- Search for log entries that do not contain the `ua_browser` field: `not ua_browser: *`

For more information about query syntax, see Index and query.

Analysis syntax

You can use the SQL-92 syntax for log analysis and statistics. For more information about the syntax and functions supported by Log Service, see #unique_142.

Note:

- In analysis clauses, the `from log part` is similar to the `from <table name> part` in standard SQL statements, and can be omitted.
- The first 100 log entries are returned by default. You can modify the number of returned log entries by using the #unique_143.
Time-based log query and analysis

Each log entry of Anti-DDoS Pro has a time field in the `yyyy-MM-ddTHH:mm:ss<timezone>` format. For example, in `2018-05-31T20:11:58+08:00`, the timezone is UTC+8. Each log entry has a built-in field: `__time__`, which indicates the time when the log entry is generated, so that time-based calculations can be performed. This field is in the Unix timestamp format, and the value of this field indicates the number of seconds that have elapsed since 00:00:00 Coordinated Universal Time (UTC), January 1, 1970. Therefore, after a timestamp is calculated, it must be formatted before it is displayed.

- Select and display the time

The following example shows how to search for the latest 10 log entries that record CC attacks on www.aliyun.com over a specific period of time. The query result includes the time, real_client_ip, and http_user_agent fields, and the log entries are sorted based on the time field.

```
matched_host: www.aliyun.com and cc_blocks: 1
| select time, real_client_ip, http_user_agent
order by time desc
limit 10
```

Figure 4-34: Select and display the time
Calculate the time

To calculate the number of days since a CC attack, you can use the `__time__` field.

Example:

```sql
matched_host: www.aliyun.com and cc_blocks: 1
| select time,
    round((to_unixtime(now()) - __time__) / 86400, 1) as "days_passed",
    real_client_ip, http_user_agent
order by time desc
limit 10
```

Note:
In the preceding example, `round((to_unixtime(now()) - __time__) / 86400, 1)` is used for calculation. The `to_unixtime` function is used to convert the time obtained by `now()` into a Unix timestamp. The build-in `__time__` field subtracted from the calculated value is the number of seconds that have elapsed. The number of days since each CC attack equals the number of seconds divided by 86,400 and then rounded to the decimal by using the `round(data, 1)` function. 86,400 is the total number of seconds in a day.

Figure 4-35: Query results
- Group log entries by a built-in time

To query CC attacks on a website every day in a specific time range, you can use the following SQL statements:

```sql
| select date_trunc('day', __time__) as dt,
   count(1) as PV
| group by dt
| order by dt
```

**Note:**

In the preceding example, the built-in `__time__` field is used in the `date_trunc('day', ...)` function to specify the time range of log entries as a day. Each log entry is assigned to a group based on the day when the log entry is generated. The total number of log entries in each group is counted by using `count(1)`. These log entries are grouped and ordered by using the `dt` field. You can use other values for the first parameter of the `date_trunc` function to group log
 entries based on other time units, such as second, minute, hour, week, month, and year. For more information about time-related functions, see #unique_144.

Figure 4-36: Analysis results

<table>
<thead>
<tr>
<th>dt</th>
<th>PV</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018-05-28 00:00:00.000</td>
<td>1319828</td>
</tr>
<tr>
<td>2018-05-29 00:00:00.000</td>
<td>2402020</td>
</tr>
<tr>
<td>2018-05-30 00:00:00.000</td>
<td>2473332</td>
</tr>
<tr>
<td>2018-05-31 00:00:00.000</td>
<td>8381076</td>
</tr>
<tr>
<td>2018-06-01 00:00:00.000</td>
<td>11293643</td>
</tr>
</tbody>
</table>

The analysis results can be displayed in a line chart.

Figure 4-37: Line chart

• Group log entries by a self-defined time

If you want to group log entries by a self-defined time, complex calculations are required. To query log entries of CC attacks on a website within every 5 minutes, you can use the following SQL statements:

```sql
matched_host: www.aliyun.com and cc_blocks: 1
| select from_unixtime(__time__ - __time__% 300) as dt, count(1) as PV
group by dt
```
```sql
order by dt
limit 1000
```

**Note:**
In the preceding example, the `__time__ - __time__ % 300` expression contains the built-in time field, and the expression result is formatted by using the `from_unixtime` function. Each log entry is assigned to a group that indicates a time range of 5 minutes (300 seconds). The total number of log entries in each group is counted by using `count(1)`. These log entries are grouped and ordered by using the `dt` field. The first 1,000 log entries are equivalent to the first 83 hours of log entries in the selected time range.

Figure 4-38: Analysis results

<table>
<thead>
<tr>
<th>dt</th>
<th>PV</th>
</tr>
</thead>
<tbody>
<tr>
<td>2019-06-20 17:35:00.000</td>
<td>1430</td>
</tr>
<tr>
<td>2019-06-20 17:40:00.000</td>
<td>6893</td>
</tr>
<tr>
<td>2019-06-20 17:45:00.000</td>
<td>6902</td>
</tr>
<tr>
<td>2019-06-20 17:50:00.000</td>
<td>5230</td>
</tr>
</tbody>
</table>

The analysis results can be displayed in a line chart.

Figure 4-39: Line chart
For more information about time-related functions, see §unique_144. For example, the `date_parse` and `date_format` functions can convert a time format to another format.

Client IP address-based log query and analysis

In log entries of Anti-DDoS Pro, the `real_client_ip` field represents the real client IP address. However, if you cannot obtain the real IP address because a user use a proxy or the IP address in the header is incorrect, you can use the `remote_addr` field.

- Distribution of attackers by country

You can use the following statements to analyze the source countries of CC attacks on a website:

```sql
matched_host: www.aliyun.com and cc_blocks: 1
| SELECT ip_to_country(if(real_client_ip='-', remote_addr, real_client_ip)) as country,
        count(1) as "number of attacks"
  group by country
```

Note:

In the preceding example, the `if(condition, option1, option2)` function returns the real client IP address. If `real_client_ip` is -, the function returns the value of `remote_addr`. Otherwise, the function returns `real_client_ip`
The `ip_to_country` function is used to retrieve the country information corresponding to the client IP address.

The analysis results can be displayed in a world map.

Figure 4-40: World map

- **Distribution of visitors by province**

  The following example shows how to use the `ip_to_province` function to obtain the distribution of visitors by province.

  ```sql
  matched_host: www.aliyun.com and cc_blocks: 1
  | SELECT ip_to_province(if(real_client_ip='-', remote_addr, real_client_ip)) as province,
  | count(1) as "number of attacks"
  | group by province
  ```

  **Note:**

In the preceding example, the `ip_to_province` function is used to retrieve the origin (province) of the IP address. If the IP address is not in China, the system attempts to obtain the province or state where this IP address is located.

- **Distribution of attackers in heatmap**

  The following example shows how to use the `ip_to_geo` function to obtain the heatmap that indicates the distribution of attackers.

  ```sql
  matched_host: www.aliyun.com and cc_blocks: 1
  | SELECT ip_to_geo(if(real_client_ip='-', remote_addr, real_client_ip)) as geo,
  | count(1) as "number of attacks"
  | group by geo
  ```
Note:
In the preceding example, the `ip_to_geo` function is used to retrieve the latitude and longitude of the IP address. The limit is set at 10,000 to retrieve the first 10,000 records.

Figure 4-41: Analysis results: distribution of attackers in heatmap

The analysis results can be displayed in a Amap.

Figure 4-42: Amap

For more information about IP-based functions, see #unique_145. For example, you can use the `ip_to_provider` function to obtain the provider of IP addresses. You can use the `ip_to_domain` function to determine whether an IP address is public or private.

4.7.6 Log reports
The Dashboard page of Log Service is embedded in the Log Reports page. The Log Reports page displays the default dashboards. You can adjust dashboard data by modifying the time range and adding filtering conditions.

View log reports

1. Log on to the new Anti-DDoS Pro console. In the left-side navigation pane, choose Statistics > Full Log.
2. Select the website domain for which you want to view log reports, and ensure that the Status switch is turned on.

3. Click Log Reports.

The Dashboard page of Log Service is embedded in the current page. The system displays log reports of the target website based on a default filtering condition, such as `matched_host:"0523.yuanya.aliyun.com"`.

Figure 4-43: View log reports

After you enable the log collection feature of Anti-DDoS Pro for your website, Log Service automatically creates two default dashboards: Operation Center and Access Center.

<table>
<thead>
<tr>
<th>Dashboard name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operation Center</td>
<td>Displays the current operational status of the website protected by Anti-DDoS Pro. This includes the traffic volume incurred by valid requests and by CC attacks, the valid request ratio, the maximum bandwidth occupied by CC attacks, and the distribution of CC attackers.</td>
</tr>
</tbody>
</table>
In addition to viewing log reports, you can perform the following operations:

- Specify a time range.
- Add or edit filtering conditions.
- View charts.

Time picker

Each chart on a dashboard is generated based on the statistical data of a separate time range. For example, the default time range is one day for the PV chart and 30 days for the chart showing PV and UV trends. To set all charts on the dashboard to be displayed in the same time range, you can configure the time picker.

1. Click Please Select.
2. Specify a time range in the right-side pane that appears. You can select a relative time, a time frame, or set a custom time.
Note:

- The specified time range applies to all charts on the dashboard.
- The settings of time pickers apply to a temporary view of charts on a dashboard, and the system does not save these settings. The next time you view log reports, the system still uses the default time range.

Figure 4-45: Specify a time range

Chart types

The dashboard displays multiple types of charts based on a predefined layout. For more information about chart types, see #unique_146.

<table>
<thead>
<tr>
<th>Chart type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single value chart</td>
<td>Displays key indicators, such as valid request ratio and attack peaks.</td>
</tr>
<tr>
<td>Line chart and area chart</td>
<td>Displays trends of key indicators within a time range, such as inbound traffic, attacks, and interception.</td>
</tr>
<tr>
<td>Map</td>
<td>Displays the geographical distribution of visitors and attackers, such as CC attack source and access heatmap.</td>
</tr>
</tbody>
</table>
## Chart type | Description
--- | ---
Pie chart | Displays the distribution of information such as top 10 attacked websites and different types of clients.
Table | Displays information such as attacker list, which typically contains multiple columns.
Map | Displays the geographical distribution of data.

### Default dashboards

- **Operation Center**

  Operation Center displays the current operational status of the website protected by Anti-DDoS Pro. This includes the traffic volume incurred by valid requests and by CC attacks, the valid request ratio, the maximum bandwidth occupied by CC attacks, and the distribution of CC attackers.

<table>
<thead>
<tr>
<th>Chart name</th>
<th>Type</th>
<th>Default time range</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid request ratio</td>
<td>Single value chart</td>
<td>1 hour (relative)</td>
<td>The ratio of valid requests to all requests. Valid requests are requests except CC attack requests and 400 bad requests.</td>
<td>95%</td>
</tr>
<tr>
<td>Valid request traffic ratio</td>
<td>Single value chart</td>
<td>1 hour (relative)</td>
<td>The ratio of the traffic incurred by valid requests to the traffic incurred by all requests.</td>
<td>95%</td>
</tr>
<tr>
<td>Chart name</td>
<td>Type</td>
<td>Default time range</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>----------------------</td>
<td>---------------------</td>
<td>--------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Traffic received</td>
<td>Single value chart</td>
<td>1 hour (relative)</td>
<td>The total inbound traffic incurred by valid requests. Unit: MB.</td>
<td>300 MB</td>
</tr>
<tr>
<td>Attack traffic</td>
<td>Single value chart</td>
<td>1 hour (relative)</td>
<td>The total inbound traffic incurred by CC attacks. Unit: MB.</td>
<td>30 MB</td>
</tr>
<tr>
<td>Traffic out</td>
<td>Single value chart</td>
<td>1 hour (relative)</td>
<td>The total outbound traffic that is generated by valid requests. Unit: MB.</td>
<td>300 MB</td>
</tr>
<tr>
<td>Peak network in</td>
<td>Single value chart</td>
<td>1 hour (relative)</td>
<td>The maximum inbound throughput of the website's requests. Unit: Bytes/s.</td>
<td>100 Bytes/s</td>
</tr>
<tr>
<td>Peak network out</td>
<td>Single value chart</td>
<td>1 hour (relative)</td>
<td>The maximum outbound throughput of the website's requests. Unit: Bytes/s.</td>
<td>100 Bytes/s</td>
</tr>
<tr>
<td>Received requests</td>
<td>Single value chart</td>
<td>1 hour (relative)</td>
<td>The total number of received valid requests (not CC attacks).</td>
<td>30,000</td>
</tr>
<tr>
<td>Chart name</td>
<td>Type</td>
<td>Default time range</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>------------------</td>
<td>--------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Attack count</td>
<td>Single value chart</td>
<td>1 hour (relative)</td>
<td>The total number of requests initiated by CC attacks.</td>
<td>100</td>
</tr>
<tr>
<td>Peak attack size</td>
<td>Single value chart</td>
<td>1 hour (relative)</td>
<td>The maximum number of CC attack requests per minute.</td>
<td>100 per minute</td>
</tr>
<tr>
<td>Network traffic in and attack</td>
<td>Two-line chart</td>
<td>1 hour (time frame)</td>
<td>The traffic incurred by valid requests and attack requests per minute. Unit: KB/s.</td>
<td>N/A</td>
</tr>
<tr>
<td>Request and interception</td>
<td>Two-line chart</td>
<td>1 hour (time frame)</td>
<td>The total number of requests and intercepted CC attack requests per minute.</td>
<td>N/A</td>
</tr>
<tr>
<td>Valid request ratio</td>
<td>Two-line chart</td>
<td>1 hour (time frame)</td>
<td>The ratio of valid requests (excluding CC attack requests and 400 bad requests) to all requests every minute.</td>
<td>N/A</td>
</tr>
<tr>
<td>Chart name</td>
<td>Type</td>
<td>Default time range</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-----------------</td>
<td>--------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Access status distribution</td>
<td>Flow chart</td>
<td>1 hour (time frame)</td>
<td>The distribution of requests with different status codes (such as 400, 304, and 200) per minute.</td>
<td>N/A</td>
</tr>
<tr>
<td>Attack source (world)</td>
<td>World map</td>
<td>1 hour (relative)</td>
<td>The distribution of CC attacks in origin countries.</td>
<td>N/A</td>
</tr>
<tr>
<td>Attack source (China)</td>
<td>China map</td>
<td>1 hour (relative)</td>
<td>The distribution of CC attacks that originate in the provinces of China.</td>
<td>N/A</td>
</tr>
<tr>
<td>Attacker list</td>
<td>Table</td>
<td>1 hour (relative)</td>
<td>The information about top 100 attackers, including IP addresses, countries, cities, network, attack count, and attack throughput.</td>
<td>N/A</td>
</tr>
<tr>
<td>Attack access line distribution</td>
<td>Pie chart</td>
<td>1 hour (relative)</td>
<td>The distribution of ISP lines accessed by CC attacks, such as lines of China Telecom, China Unicom, and BGP.</td>
<td>N/A</td>
</tr>
</tbody>
</table>
**Chart name** | **Type** | **Default time range** | **Description** | **Example**
---|---|---|---|---
Top 10 attacked websites | Doughnut chart | 1 hour (relative) | The top 10 most attacked websites. | N/A

**Access Center**

Access Center displays the current access status of the website protected by Anti-DDoS Pro. This includes statistics on PV, UV, peak throughput, visitor locations, lines, client types, request types, visited websites.

<table>
<thead>
<tr>
<th>Chart name</th>
<th>Type</th>
<th>Default time range</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>PV</td>
<td>Single value chart</td>
<td>1 hour (relative)</td>
<td>The total number of page views (PVs).</td>
<td>100,000</td>
</tr>
<tr>
<td>UV</td>
<td>Single value chart</td>
<td>1 hour (relative)</td>
<td>The total number of unique visitors (UVs).</td>
<td>100,000</td>
</tr>
<tr>
<td>Traffic in</td>
<td>Single value chart</td>
<td>1 hour (relative)</td>
<td>The total inbound traffic. Unit: MB.</td>
<td>300 MB</td>
</tr>
<tr>
<td>Peak network in traffic</td>
<td>Single value chart</td>
<td>1 hour (relative)</td>
<td>The maximum inbound throughput of the website's requests. Unit: Bytes/s.</td>
<td>100 Bytes/s</td>
</tr>
<tr>
<td>Peak network out traffic</td>
<td>Single value chart</td>
<td>1 hour (relative)</td>
<td>The maximum outbound throughput of the website's requests. Unit: Bytes/s.</td>
<td>100 Bytes/s</td>
</tr>
<tr>
<td>Chart name</td>
<td>Type</td>
<td>Default time range</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>---------------</td>
<td>--------------------</td>
<td>------------------------------------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Traffic network trend</td>
<td>Two-line chart</td>
<td>1 hour (time frame)</td>
<td>The trend of inbound and outbound website traffic per minute. Unit: Bytes/s.</td>
<td>N/A</td>
</tr>
<tr>
<td>Request and interception</td>
<td>Two-line chart</td>
<td>1 hour (time frame)</td>
<td>The total number of requests and intercepted CC attack requests per minute.</td>
<td>N/A</td>
</tr>
<tr>
<td>PV/UV trends</td>
<td>Two-line chart</td>
<td>1 hour (time frame)</td>
<td>The trends of PV and UV per minute.</td>
<td>N/A</td>
</tr>
<tr>
<td>Access source</td>
<td>World map</td>
<td>1 hour (relative)</td>
<td>The distribution of visitors in origin countries.</td>
<td>N/A</td>
</tr>
<tr>
<td>Access heatmap</td>
<td>Amap</td>
<td>1 hour (relative)</td>
<td>The heatmap that represents the geographical locations of visitors.</td>
<td>N/A</td>
</tr>
<tr>
<td>Traffic in source (world)</td>
<td>World map</td>
<td>1 hour (relative)</td>
<td>The distribution of inbound traffic in origin countries. Unit: MB.</td>
<td>N/A</td>
</tr>
<tr>
<td>Traffic in source (China)</td>
<td>China map</td>
<td>1 hour (relative)</td>
<td>The distribution of inbound traffic that originates in the provinces of China. Unit: MB.</td>
<td>N/A</td>
</tr>
<tr>
<td>Chart name</td>
<td>Type</td>
<td>Default time range</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------------</td>
<td>--------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Access line distribution</td>
<td>Doughnut chart</td>
<td>1 hour (relative)</td>
<td>The distribution of ISP lines accessed by visitors, such as lines of China Telecom, China Unicom, and BGP.</td>
<td>N/A</td>
</tr>
<tr>
<td>Network provider source</td>
<td>Doughnut chart</td>
<td>1 hour (relative)</td>
<td>The proportion of inbound traffic that is carried by the line of each ISP, such as China Telecom, China Unicom, China Mobile, and CERNET. Unit: MB.</td>
<td>N/A</td>
</tr>
<tr>
<td>Chart name</td>
<td>Type</td>
<td>Default time range</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>---------------------</td>
<td>--------------</td>
<td>--------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Top clients</td>
<td>Table</td>
<td>1 hour (relative)</td>
<td>The information about the top 100 most visited clients. The information includes IP addresses, countries, cities, network, request method distribution, inbound traffic, the number of invalid requests, and the number of intercepted CC attacks.</td>
<td>N/A</td>
</tr>
<tr>
<td>Accessed websites</td>
<td>Doughnut chart</td>
<td>1 hour (relative)</td>
<td>The domain names of the top 20 most visited websites.</td>
<td>N/A</td>
</tr>
<tr>
<td>Referer</td>
<td>Table</td>
<td>1 hour (relative)</td>
<td>The top 100 most used referer URLs, redirection target hosts, and the number of redirections.</td>
<td>N/A</td>
</tr>
<tr>
<td>Chart name</td>
<td>Type</td>
<td>Default time range</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-------------------</td>
<td>----------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>PC client distribution</td>
<td>Doughnut chart</td>
<td>1 hour (relative)</td>
<td>The top 20 most used user agents, such as iPhone, iPad, Windows IE, and Chrome.</td>
<td>N/A</td>
</tr>
<tr>
<td>Request content type distribution</td>
<td>Doughnut chart</td>
<td>1 hour (relative)</td>
<td>The top 20 most requested content types, such as HTML, Form, JSON, and streaming data.</td>
<td>N/A</td>
</tr>
</tbody>
</table>

4.7.7 Advanced settings

The Anti-DDoS Pro log collection function supports advanced management. You can click Advanced Management to go to the Logstores page in the Log Service console. You can perform advanced operations on Anti-DDoS Pro logs, including real-time subscription and consumption, data shipping, and other visualization operations.

In the upper-right corner of the Log Service page, click Advanced Management to go to the Log Service console. You can perform advanced operations including exporting log data and configuring log consumption.
Export log data

1. After you enable log collection, click the Download button on the right of the Raw Logs tab page.

2. In the Log Download dialog box that appears, click Download Log in Current Page to export the log entries displayed on this page into a file in CSV format.
3. You can also click Download All Logs Using Command Line Tool to download all log entries.

Log Download

- Download Log in Current Page
- Download All Logs Using Command Line Tool

1. Install the command line tool

For information about the command line tool installation, see: Documentation

2. View the AccessKeyId and AccessKeySecret of the current user

Address: Security information management

3. Use the command line tool

```
aliyunlog log get_log_all --project="ddoscoo-project-1289654106023090-cn-hangzhou" --logstore="ddoscoo-logstore" --query="matched host:"reportauto.\q.com\"" --from_time="2019-06-20 09:46:47+08:00" --to_time="2019-06-20 10:1:47+08:00" --region-endpoint="cn-hangzhou.log.aliyuncs.com" --format-output=no_escape --jmes-filter="join(\n', map(\n to_string(\n', )))" --access-id =""AccessKeyId obtained in step 2"" --access-key=""AccessKeySecret obtained in step 2"" >> ./downloaded_data.txt
```

Switch to Internal Endpoint  Copy Command

4. Modify the AccessKeyId and AccessKeySecret in the command

After the command is executed, the search result is automatically downloaded to download_data.txt under the current directory where the command was executed. Click OK to view the detailed information about the command line tool usage.

- a. Click Documentation to see User Guide for Alibaba Cloud CLI.
- b. Install the command line tool.
- c. Click Security information management to view and copy the AccessKey ID and AccessKey Secret of the current user.
d. Click Copy Command and replace AccessKeyId obtained in step 2 and AccessKeySecret obtained in step 2 with the AccessKeyId and AccessKey Secret of the current user.

e. Run the command in the command line tool.

After you run the command, log entries are automatically downloaded and saved to the download_data.txt file in the directory where the command was run.

Other advanced operations

- Alerts and notifications
- Real-time subscription and consumption
- Data shipping
- Integration with other visualization tools

4.8 TDS logs

Alibaba Cloud Threat Detection Service (TDS) provides a log analysis function to collect, analyze, query, store, and distribute risk and threat data in real time. This frees you from the need to manually collect, query, and analyze data, improving your overall O&M efficiency.

Overview

Alibaba Cloud TDS is fully integrated with Log Service and provides TDS log collection and analysis functions, which can help you better understand and more effectively address server security risks and manage your assets on the cloud. TDS is suitable for the following enterprise-level scenarios:

- Large-scale enterprises and organizations, such as finance companies and government agencies, which require strict storage compliance for hosts, networks, and security logs, among other assets on the cloud
- Large-scale real-estate, e-commerce, or finance companies, along with government agencies, which possess on-premises security operations centers (SOCs) and require centralization collection and management of security and alarm logs
- Enterprises with advanced technologies, such as companies in IT, gaming, or finance, which require in-depth analysis of logs collected from various cloud assets and automated alarm handling
Benefits

- Quick analysis capabilities: The analysis of security and host logs can be completed in seconds, and analysis of network logs within an hour.
- Comprehensive support: A total of 14 log types are provided, including network, host, and security logs.
- Fully integrated: TDS is fully integrated with the open-source streaming and big data system solutions on Alibaba Cloud and is publicly open to our partners.
- Flexible to various applications: With support for WYSIWYG analysis capabilities, you can customize service views as needed.

Limits

- TDS-dedicated Logstores cannot store non-TDS data.

  TDS logs are stored in dedicated Logstores. These Logstores cannot store non-TDS data that is written through APIs or SDKs. Dedicated Logstores have no limits on queries, statistics, alarms, and stream consumption.

- Basic settings, such as the storage period of dedicated Logstores, cannot be modified.

- Dedicated Logstores do not incur charges.

  Dedicated Logstores do not incur charges on the condition that the Log Service functions normally.

  Note:
  The TDS log analysis function is unavailable in the case that Log Service charges are overdue. In such a case, you must first pay your overdue payments before you can gain access to this function.

Scenarios

- Track host and network logs and trace the source of security threats.

  You can retrieve the `__topic__` field in logs and view the time distribution of different types of logs to track host and network logs in real time.

- View host and network operations in real time to gain insight into security status and trends.

  You can view host and network operations in the Web access center dashboard to assess the security of your assets in a timely manner.
· Understand security operating efficiency and handle issues and threats in a prompt manner.

You can view your current security operating efficiency in the vulnerability center dashboard.

4.9 WAF logs

4.9.1 Real-time log analysis

Integrated with Log Service, WAF provides access logs and attack logs, and allows you to analyze logs in real time.

The real-time log analysis feature in WAF collects and stores access logs in real time and provides the following capabilities based on Log Service: log querying, analysis, reporting, alerting, forwarding, and computing. The service makes it easy to search log data so that you can focus on log analysis.

Target users

· Large enterprises and institutions that need to meet compliance requirements regarding the use of cloud hosts, networks, and the storage of security logs, such as financial companies and government agencies.

· Enterprises that have private security operations centers (SOCs) and need to collect security logs for centralized operations and management, such as large real estate, e-commerce, financial companies, and government agencies.

· Enterprises that have strong technical capabilities and need to perform in-depth analysis on logs of cloud resources, such as IT, gaming, and financial companies.

· Small and medium-sized enterprises and institutions that need to meet compliance requirements regarding their business on the cloud or need to generate business reports on a regular basis, such as monthly, quarterly, and annual reports.

Benefits

· Compliance: Stores the website's access logs for more than six months to help the website meet the compliance requirements.

· Simple configuration: You can easily configure the service to collect access logs and attack logs on your site.
· Real-time analysis: Integrated with Log Service, the service supports real-time log analysis and offers a ready-to-use report center. You can easily gain information about the details of attacks, and visits to your site.

· Real-time alarms: Near real-time monitoring and custom alarms based on specific metrics are available to ensure a timely response to critical service failures.

· Collaboration: The service can be integrated with real-time computing, cloud storage, visualization, and other data solutions to help you gain valuable insights into your data.

Prerequisites and limits

To use the real-time analysis feature in WAF, you must meet the following prerequisites:

· You have activated Log Service.

· You have activated WAF Enterprise Edition and enabled the log analysis module.

All log data in WAF is stored in an exclusive logstore that has the following limits:

· Users cannot use APIs or SDKs to write data to the logstore or change attributes of the logstore, such as the storage period.

Note:
The logstore supports common features, including log querying, reporting, alerting, and stream computing.

· The logstore is free of charge on condition that Log Service is available and your account has no overdue payments.

· The system reports may be updated at irregular intervals.

Scenarios

· Analyze log data to track attacks and identify threats.

· Monitor Web requests in real time to predict traffic trends.

· Quickly learn about the efficiency of security operations and obtain timely feedback.

· Transfer network logs to user-created data centers or computing centers.
4.9.2 Activate WAF Log Service

After purchasing a Web Application Firewall instance, you can activate the real-time log query and analysis service for your websites on the App Management page in the console.

Scope

With WAF Log Service, you can collect multiple log entries in real time from your websites that are protected by WAF. You can also perform real-time log query and analysis and display results in dashboards. WAF Log Service fully meets the business protection needs and operational requirements of your websites. You can select the log storage period and the log storage size as needed when enabling WAF Log Service.

Note:
At the moment, WAF Log Service is only available to WAF subscription instances (Pro, Business, or Enterprise edition).

Enable WAF Log Service

1. Log on to the Web Application Firewall console.
2. Choose App Market > App Management, and select the region where your WAF instance is located.
3. Click Upgrade in Real-time Log Query and Analysis Service.
4. On the page that is displayed, enable Log Service, select the log storage period and the log storage size, and then click Buy Now.

Note:
For more information about the billing of WAF Log Service, see WAF Log Service Billing methods.

5. Return to the WAF console and choose App Market > App Management, and then click Authorize in Real-time Log Query and Analysis Service.
6. Click Agree to authorize WAF to write log entries to your exclusive logstore.

WAF Log Service is then enabled and authorized.
7. Return to the WAF console and choose App Market > App Management and then, click Configure in Real-time Log Query and Analysis Service.
8. On the Log Service page, select the domain name of your website that is protected by WAF, and turn on the Status switch on the right to enable WAF Log Service.

Log Service collects all web log recorded by WAF in real time. These log entries can be queried and analyzed in real time.

4.9.3 Billing method

Web Application Firewall (WAF) Log Service is billed based on the log storage period and the log storage size of your choice.

WAF Log Service is activated on a subscription basis.

Note:
To activate WAF Log Service, you must buy a WAF subscription.

In the WAF purchase page, enable Activate Log Service and select the log storage period and the log storage size. Then, the price is automatically calculated based on the log store specification of your choice and the validity of the WAF instance.

Log storage specification

The detailed pricing for each log storage specification for WAF Log Service is shown in the following table.

<table>
<thead>
<tr>
<th>Log storage period</th>
<th>Log storage size</th>
<th>Recommended scenarios</th>
<th>For International region instances</th>
<th>For Mainland China region instances</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Monthly subscription</td>
<td>Yearly subscription</td>
</tr>
<tr>
<td>180 days</td>
<td>3 TB</td>
<td>Average daily QPS is up to 80.</td>
<td>USD 450</td>
<td>USD 5,400</td>
</tr>
<tr>
<td></td>
<td>5 TB</td>
<td>Average daily QPS is up to 120.</td>
<td>USD 750</td>
<td>USD 9,000</td>
</tr>
<tr>
<td></td>
<td>10 TB</td>
<td>Average daily QPS is up to 260.</td>
<td>USD 1,500</td>
<td>USD 18,000</td>
</tr>
<tr>
<td>Log storage period</td>
<td>Log storage size</td>
<td>Recommended scenarios</td>
<td>For International region instances</td>
<td>For Mainland China region instances</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------------</td>
<td>----------------------</td>
<td>----------------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Monthly subscription</td>
<td>Yearly subscription</td>
</tr>
<tr>
<td></td>
<td>20 TB</td>
<td>Average daily QPS is up to 500.</td>
<td>USD 3,000</td>
<td>USD 36,000</td>
</tr>
<tr>
<td></td>
<td>50 TB</td>
<td>Average daily QPS is up to 1,200.</td>
<td>USD 7,500</td>
<td>USD 90,000</td>
</tr>
<tr>
<td></td>
<td>100 TB</td>
<td>Average daily QPS is up to 2,600.</td>
<td>USD 15,000</td>
<td>USD 180,000</td>
</tr>
<tr>
<td></td>
<td>360 days</td>
<td>5 TB</td>
<td>Average daily QPS is up to 60.</td>
<td>USD 750</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10 TB</td>
<td>Average daily QPS is up to 120.</td>
<td>USD 1,500</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20 TB</td>
<td>Average daily QPS is up to 260.</td>
<td>USD 3,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50 TB</td>
<td>Average daily QPS is up to 600.</td>
<td>USD 7,500</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100 TB</td>
<td>Average daily QPS is up to 1,200.</td>
<td>USD 15,000</td>
</tr>
</tbody>
</table>

Upgrade storage capacity

If you have no log storage left, a notification appears to remind you to expand the storage size. You can expand the log storage size at any time.

⚠️ Notice:
If log storage is full, WAF stops writing new log entries to the exclusive logstore in Log Service. A log entry stored in the logstore is deleted based on the specified...
period. If the WAF Log Service instance expires and you do not renew it within seven days, all log entries in the logstore are deleted.

Validity

The validity of the WAF Log Service instance is based on your WAF subscription.

- **Buy**: When you buy a WAF subscription and enable Log Service, the price of Log Service is calculated based on the validity of the subscription.
- **Upgrade**: When you enable Log Service by upgrading an existing WAF subscription, the price of Log Service is calculated based on the remaining validity of the existing WAF instance. The remaining validity is accurate to minutes.

Service expiration

If your WAF instance expires, WAF Log Service expires at the same time.

- When the service expires, WAF stops writing log entries to the exclusive logstore in Log Service.
- The log entries recorded by WAF Log Service are retained within seven days after the service expires. If you renew the service within seven days after the service expires, you can continue to use WAF Log Service. Otherwise, all stored WAF log entries are deleted.

4.9.4 Log collection

You can enable the Web Application Firewall (WAF) log collection feature for a specified domain in the WAF console.

Prerequisites

- Buy a WAF instance and protect the domain using WAF.
- Enable Log Service.

Context

Log Service collects log entries that record visits to and attacks on websites that are protected by Alibaba Cloud WAF, and supports real-time log query and analysis. The query results are displayed in dashboards. You can timely perform analytical investigation on visits to and attacks on your websites and help security engineers to develop protection strategies.

Procedure
1. Log on to the **Web Application Firewall console**.


   **Note:**
   If you are configuring the WAF log collection feature for the first time, click Authorize and follow the instructions on the authorization page to authorize WAF to write all log entries to your exclusive logstore.

3. Select the domain and turn on the Status switch on the right to enable the log collection feature.

   The WAF log collection feature has now been enabled for the domain. Log Service automatically creates an exclusive logstore for your account. WAF
automatically writes log entries to the exclusive logstore. The following Default configuration table describes the default configuration of the exclusive logstore.

Table 4-11: Default configuration

<table>
<thead>
<tr>
<th>Default configuration item</th>
<th>Description</th>
</tr>
</thead>
</table>
| Project                    | A project is created by default. The project name format is determined by the region of your WAF instance.  
  - If the WAF instance is created in Mainland China, the project name is waf-project-Your Alibaba Cloud account ID-cn-hangzhou.  
  - If the WAF instance is created in other regions, the project name is waf-project-Your Alibaba Cloud account ID-ap-southeast-1. |
| Logstore                   | A logstore waf-logstore is created by default.  
  All log entries collected by the WAF log collection feature are saved in this logstore. |
| Region                     |  
  - If the WAF instance is created in Mainland China, the project is saved in the Hangzhou region by default.  
  - If the WAF instance is created in other regions, the project is saved in the Singapore region by default. |
| Shard                      | Two shards are created by default with the Automatic shard splitting feature enabled. |
### Default configuration item

<table>
<thead>
<tr>
<th>Default configuration item</th>
<th>Description</th>
</tr>
</thead>
</table>
| Dashboard                 | Three dashboards are created:  
  • Access Center  
  • Operation Center  
  • Security Center  
  For more information about dashboards, see *WAF Log Service—Log Reports.* |

### Limits and instructions

- Other data cannot be written to the exclusive logstore.

  Log entries generated by WAF are stored in the exclusive logstore. You cannot write other data to this logstore by using API, SDK or other methods.

  **Note:**
  The exclusive logstore has no special limits in query, statistics, alerts, streaming consumption and other functions.

- Basic configurations, such as the storage period of log entries, cannot be modified.

- The exclusive logstore is not billed.

  To use the exclusive logstore, you must enable Log Service for your account. The exclusive logstore is not billed.

  **Note:**
  When your Log Service is overdue, the WAF log collection feature is suspended until you pay the bills in a timely manner.

- Do not delete or modify the configurations of the project, logstore, index, and dashboards, which are created by Log Service by default. Log Service updates the WAF log query and analysis service on an irregular basis. The index of the exclusive logstore and the default reports are also updated automatically.

- If you want to use the WAF log query and analysis service with a RAM user, you must grant the required Log Service permissions to the RAM user. For more information about how to grant permissions, see *Grant log query and analysis permissions to a RAM user.*
4.9.5 Log Analyses

The Real-time Log Query and Analysis Service page in the Web Application Firewall (WAF) console is integrated with the Log Analyses feature and the Log reports feature. After enabling the WAF log collection feature for a domain, you can perform real-time query and analysis, view or edit dashboards, and set up monitoring and alarms in the Real-time Log Query and Analysis Service page.

Procedure

1. Log on to the Web Application Firewall console, and choose App Market > App Management.
2. Click on the Real-time Log Query and Analysis Service area to open the Log Service page.
3. Select the domain and check that the Status switch on the right is turned on.
4. Click Log Analyses.

The current page is integrated with the Querying and analyzing page. A query statement is automatically inserted. For example, matched_host: "www.aliyun.com" is used to query all log entries that is related to the domain in the statement.
5. Enter a query and analysis statement, select a log time range, and then click Search & Analysis.

More operations

The following operations are available in the Log Analyses page.

- Customize query and analysis

Log Service provides rich query and analysis syntax for querying log entries in a variety of complex scenarios. For more information, see the Custom query and analysis in this topic.

- View the distribution of log entries by time period

Under the query box, you can view the distribution of log entries that are filtered by time period and query statement. A histogram is used to indicate the distribution, where the horizontal axis indicates the time period, and the vertical axis indicates the number of log entries. The total number of the log entries in the query results is also displayed.

Note:
You can hold down the left mouse button and drag the histogram to select a shorter period. The **time picker** automatically updates the time period, and the query results are also updated based on the updated time period.

- **View raw log entries**

  In the Raw Logs tab, each log entry is detailed in a single page, which includes the time when the log entry is generated, the content, and the properties in the log entry. You can click Display Content Column to configure the display mode (Full Line or New Line) for long strings in the Content column. You can click Column Settings to display specific fields, or click the Download Log button to download the query results.

  Additionally, you can click a value or a property name to add a query criterion to the query box. For example, if you click the value **GET** in the `request_method:` **GET** filed, the query statement in the query box is updated to:

  ```
  <The original query statement> and request_method: GET
  ```
· View analysis graphs

Log Service enables you to display the analysis results in graphs. You can select the graph type as needed in the Graph tab. For more information, see *Analysis graph*.

· Perform quick analysis

The Quick Analysis feature in the Raw Logs tab provides you with an one-click interactive experience, which gives you a quick access to the distribution of log entries by a single property within a specified time period. This feature can
reduce the time used for indexing key data. For more information, see *Quick analysis* in the following section.

Customize query and analysis

The log query statement consists of the query (Search) and the analysis (Analytics). These two parts are divided by a vertical bar (|):

$Search | $Analytics

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query (Search)</td>
<td>A keyword, a fuzzy string, a numerical value, a range, or other criteria can be used in the query criteria. A combined condition can also be used. If the statement is empty or only contains an asterisk (*), all log entries are displayed.</td>
</tr>
<tr>
<td>Analysis (Analytics)</td>
<td>Performs computing and statistics to the query results or all log entries.</td>
</tr>
</tbody>
</table>
Note:
Both the query part and the analysis part are optional.

- When the query part is empty, all log entries within the time period are displayed. Then, the query results are used for statistics.
- When the analysis part is empty, only the query results are returned without statistics.

Query syntax
The query syntax of Log Service supports full-text index and field search. You can enable the New Line display mode, syntax highlighting, and other features in the query box.

- Full text index
  You can enter keywords without specifying properties to perform the query by using the full-text index. You can enter the keyword with double quotation marks (""") surrounded to query log entries that contain the keyword. You can also add a space or and to separate keywords.

Examples
- Multiple-keywords query
  The following statements can be used to query all log entries that contain www.aliyun.com and error.

  www.aliyun.com error or www.aliyun.com and error.

- Criteria query
  The following statement can be used to search for all log entries that contain www.aliyun.com, error or 404.

  www.aliyun.com and (error or 404)

- Prefix query
  The following statement can be used to query all log entries that contain www.aliyun.com and start with failed_.

  www.aliyun.com and failed_*
An asterisk (*) can be added as a suffix, but it cannot be added as a prefix. For example, the statement cannot be *_error.

- **Field search**

You can perform a more accurate query based on specified fields.

The field search supports comparison queries for fields of numeric type. The format is field name:value or field name>=value. Moreover, you can perform combination queries using and or or, which can be used in combination with the full text index.

**Note:**
The log entries that record access, operation, and attack on the domain name in WAF Log Service can also be queried by fields. For more information about the meaning, type, format, and other information of the fields, see *Fields in the WAF log entries*.

**Examples**

- **Multiple-fields query**

  The following statement can be used to query all log entries that record the HTTP flood attack on the www.aliyun.com domain and are intercepted by WAF.

  ```
  matched_host: www.aliyun.com and cc_blocks: 1
  ```

  If you want to query all log entries that record access from a specific client whose IP address is 1.2.3.4 to www.aliyun.com, and access is blocked by the 404 error, you can use the following statement.

  ```
  real_client_ip: 1.2.3.4 and matched_host: www.aliyun.com and status: 404
  ```
In this example, the matched_host, cc_blocks, real_client_ip, and status fields are the fields defined in the WAF log.

- Numeric fields query

The following statement can be used to query all log entries where the response time exceeds five seconds.

request_time_msec > 5000

Range query is also supported. For example, you can query all log entries where the response time exceeds five seconds and is no more than 10 seconds.

request_time_msec in (5000 10000]

Note:
The following query statement has the same function.

request_time_msec > 5000 and request_time_msec <= 10000

- Field existence query

You can perform a query based on the existence of a field.

■ The following statement can be used to search for all log entries where the ua_browser field exists.

ua_browser: *

■ The following statement can be used to search for all log entries where the ua_browser field does not exist.

not ua_browser: *

For more information about the query syntax that is supported by Log Service, see Index and query.

Syntax for analysis

You can use the SQL/92 syntax for log analysis and statistics.

For more information about the syntax and functions supported by Log Service, see Syntax description.
The `from` table name part that follows the SQL standard syntax can be omitted from the analysis statement. In WAF Log Service, `from log` can be omitted.

The first 100 results are returned by default, and you can modify the number of results that are returned by using the `LIMIT` syntax.

Examples of query and analysis

**Time-based log query and analysis**

Each WAF log entry has a `time` field, which is used to represent the time when the log entry is generated. The format of the value in this field is `<year>-<month>-<day> T<hour>:<minute>:<second>+<time zone>`. For example, `2018-05-31T20:11:58+08:00` is 20:11:58 UTC+8 (Beijing Time), May 15, 2018.

In addition, each log entry has a built-in field `__time__`, which is also used to indicate the time when the log entry is generated. This field is used for calculation when performing statistics. The format of this field is a *Unix timestamp*, and the value of this field indicates the number of seconds that have elapsed since 00:00:00 Coordinated Universal Time (UTC), January 1, 1970. Therefore, if you want to display a calculated result, you must convert the format first.

- Select and display the time

  You can query the log based on the `time` field. For example, you can search for the last 10 log entries that record the HTTP flood attacks on `www.aliyun.com` and are intercepted by WAF. Then, you can display the time field, the source IP field, and the client field.

```sql
matched_host: www.aliyun.com and cc_blocks: 1
| select time, real_client_ip, http_user_agent
order by time desc
```
```sql
matched_host: [redacted].com and cc_blocks: 1
| select time, real_client_ip, http_user_agent
| order by time desc
| limit 10
```
• Calculate using time.

You can use the `__time__` field to calculate using time. For example, you can calculate the number of days that have elapsed since the domain suffered a HTTP flood attack.

```sql
matched_host: www.aliyun.com and cc_blocks: 1
round((to_unixtime(now()) - __time__/86400, 1) as "days_passed",
real_client_ip, http_user_agent
order by time desc
limit 10
```

**Note:**

In this example, `round((to_unixtime(now()) - __time__/86400, 1)` is used to calculate the number of days that have elapsed since the domain had a HTTP flood attack. First, use `now()` to get the current time, and convert the current time into a Unix timestamp using `to_unixtime`. Then, subtract the converted time with the value of the built-in field `__time__` to get the number of seconds that have elapsed. Finally, divide it by 86400 (the total number of seconds in a day) and apply the `round(data, 1)` function to keep one decimal place. The result is the number of days that have elapsed since each attack log entry is generated.

• Perform group statistics based on a specific time

You can query the log based on the trend of HTTP flood attacks on the domain within a specified time period.

```sql
matched_host: www.aliyun.com and cc_blocks: 1
| select date_trunc('day', __time__) as dt, count(1) as PV
| group by dt
```
log service

data collection / 4 cloud product collection

order by dt

Note:
In this example, the built-in field __time__ is used by the date_trunc('day', ...) function to align the time of the entries by day. Each log entry is assigned to a group based on the day when the log entry is generated. The total number of log entries in each group is counted using count(1). Then, these entries are ordered by the group. You can use other values for the first parameter of the date_trunc function to group the log entries based on other time units, such as second, minute, hour, week, month, and year. For more information about this function, see Date and time functions.

<table>
<thead>
<tr>
<th>dt</th>
<th>PV</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018-12-03 00:00:00.000</td>
<td>3</td>
</tr>
</tbody>
</table>

Note:
You can also display the results with a line chart.

- Perform group statistics based on time.

If you want to analyze the log based on time using more flexible groupings, complex calculations are required. For example, you can query the log based on the trend of HTTP flood attacks on the domain within every five minutes.

matched_host: www.aliyun.com and cc_blocks: 1

| select from_unixtime(__time__ - __time__% 300) as dt, count(1) as PV group by dt |
Log Service

Order by dt
limit 1000

Note:
In this example, the built-in field is used for aligning the time by using the formula `__time__ - __time__% 300`, and the `from_unixtime` function converts the format of the result. Then, each entry is assigned to a group that indicates a time period of five minutes (300 seconds), and the total number of log entries in each group is counted using `count(1)`. Finally, the query results are ordered by group and the first 1,000 results are returned, which include the log entries that are generated within 83 hours before the specified time period.

<table>
<thead>
<tr>
<th>dt</th>
<th>PV</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018-05-31 21:30:00,000</td>
<td>134795</td>
</tr>
<tr>
<td>2018-05-31 21:35:00,000</td>
<td>137891</td>
</tr>
<tr>
<td>2018-05-31 21:40:00,000</td>
<td>140171</td>
</tr>
<tr>
<td>2018-05-31 21:45:00,000</td>
<td>142037</td>
</tr>
<tr>
<td>2018-05-31 21:50:00,000</td>
<td>139958</td>
</tr>
<tr>
<td>2018-05-31 21:55:00,000</td>
<td>142906</td>
</tr>
<tr>
<td>2018-05-31 22:00:00,000</td>
<td>145093</td>
</tr>
<tr>
<td>2018-05-31 22:05:00,000</td>
<td>147474</td>
</tr>
</tbody>
</table>

Note:
You can also display the results with a line graph.
The `date_parse` and `date_format` functions are used to convert the time format. For more information about the functions that can be used to parse the time field, see Date and time functions.

**Client IP address-based log query and analysis**

The WAF log contains the field `real_client_ip`, which reflects the real client IP address. In cases where the user accesses your website through a proxy server, or the IP address in the request header is wrong, you cannot get the real IP address of the user. However, the `remote_addr` field forms a direct connection to the client, which can be used to get the real IP address.

- **Classify attackers by country**

  You can query the log based on the distribution of HTTP flood attackers by country.

  ```sql
  matched_host: www.aliyun.com and cc_blocks: 1
  | SELECT ip_to_country(if(real_client_ip='-', remote_addr,
  real_client_ip)) as country,
  count(1) as "number of attacks"
  group by country
  ```

  **Note:**

  In this example, the function `if(condition, option1, option2)` returns the real client IP address. If `real_client_ip` is `-`, the function returns the value of
remote_addr. Otherwise, the function returns real_client_ip. Then, use the ip_to_country to get the country information from the IP address of the client.

Note:
You can also display the results with a world map.

- Distribution of visitors by province

If you want to get the distribution of visitors by province, you can use the ip_to_province function to get the province information from the IP addresses.

matched_host: www.aliyun.com and cc_blocks: 1

\[
\text{SELECT ip_to_province(if(real_client_ip='-', remote_addr, real_client_ip)) as province, count(1) as "number of attacks" group by province}
\]

Note:
In this example, the ip_to_province function is used to get the country information from the real IP address of the client. If the IP address is not in the Mainland of China, the function returns the province or state of the IP address in the country field. However, if you choose to display the results with a map of China, IP addresses that are not in the Mainland of China are not displayed.

Note:
You can also display the results with a map of China.
Heat map that indicates the distribution of attackers

You can use the `ip_to_geo` function to get the geographic information (the latitude and the longitude) from the real IP addresses of the clients. This information can be used to generate a heat map to indicate the density of attacks.

```
matched_host: www.aliyun.com and cc_blocks: 1
| SELECT ip_to_geo(if(real_client_ip='-', remote_addr, real_client_ip)) as geo,
    count (1) as "number of attacks"
group by geo
limit 10000
```

Note:
In this example, the `ip_to_geo` function is use to get the latitude and the longitude from the real IP addresses of the clients. The first 10,000 results are returned.

Select Amap and click Show Heat Map.

The `ip_to_provider` function can be used to get the IP provider name, and the `ip_to_domain` function can be used to determine whether the IP is a public IP or a private IP. For more information about the functions that can be used to resolve IP addresses, see IP functions.
4.9.6 Log Reports

The Log Reports page is integrated with the Dashboard page of Log Service. On this page, you can view default dashboards. You can filter business and security data about your website by modifying the time range or adding filters.

View reports

1. Log on to the Web Application Firewall console, and choose App Market > App Management.
2. Click the Real-time Log Query and Analysis Service area to open the Log Service page.
3. [DO NOT TRANSLATE]
4. Select a domain and check that the Status switch on the right is turned on.
5. Click Log Reports.

The page that appears is integrated with the Dashboard page of Log Service. A filter is automatically added to display all log entries that are recorded for the domain you selected. In this example, the filter is matched_host: www.aliyun.com.

After you enable the WAF log collection feature, Log Service creates three dashboards by default: the Operation Center, Access Center, and Security Center.

Note:
For more information about the default dashboards, see Default dashboards.
<table>
<thead>
<tr>
<th>Dashboard</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operation Center</td>
<td>Displays operation details such as the proportion of valid requests and the statistics of attacks, traffic details such as the peak of both inbound and outbound throughput and the number of requests received, operation trends, attack overview, and other information.</td>
</tr>
<tr>
<td>Access Center</td>
<td>Displays basic access details such as the number of page views (PV) and the number of unique visitors (UV), the access trend, the distribution of visitors, and other information.</td>
</tr>
<tr>
<td>Security Center</td>
<td>Displays basic index information of attacks, attack types, attack trend, attacker distribution, and other information.</td>
</tr>
</tbody>
</table>

**Note:**
Dashboards displays various reports using the layout that is predefined in WAF Log Service. The following table describes the graph types supported for reports. For more information about the graph types supported by Log Service, see *Graph description.*

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>Graphs of this type display important metrics, such as the valid request ratio and the peak of attacks.</td>
</tr>
<tr>
<td>Type</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Line chart and area chart</td>
<td>Graphs of these types display the trend of important metrics within a specified time period, such as the trend of inbound throughput and the trend of attack interceptions.</td>
</tr>
<tr>
<td>Map</td>
<td>Graphs of this type display the geographical distribution of visitors and attackers, for example, by country. Heat maps are also supported to illustrate the distribution of attackers.</td>
</tr>
<tr>
<td>Pie chart</td>
<td>Graphs of this type display a distribution, such as the distribution of attackers and the distribution of client types.</td>
</tr>
<tr>
<td>Table</td>
<td>Graphs of this type display a table that contains information, such as information of attackers.</td>
</tr>
<tr>
<td>Map</td>
<td>Graphs of this type display the geographical distribution of data.</td>
</tr>
</tbody>
</table>

**Time selector**

The data in all graphs on the dashboard page are generated based on different time ranges. If you want to unify the time ranges, configure the time selector.

1. On the Log Reports page, click Please Select and
2. select a time range in the pane that appears. You can select a relative time, a time frame, or customize a time range.

**Note:**

- After you set a time range, the time range is applied to all reports.
- If you set a time range, a temporary view is generated on the current page. When you view reports next time, the default time range is used.
- To change the time range for a single report in the dashboard, click in the upper-right corner.
Time

Relative

1Minute 5Minutes 15Minutes
1Hour 4Hours 1Day Today
1Week 30Days Custom

Time Frame

1Minute 15Minutes 1Hour
4Hours 1Day 1Week 30Days
Today Yesterday
The Day before Yesterday This Week
Previous Week This Month This Quarter
Custom
Data drilldown

The drilldown operation is enabled for some graphs on the dashboard page, which provides you a quick access to the detailed data.

The drilldown operation is available for graphs marked with a icon in the upper-right corner. You can click a number with an underline to view the detailed underlying data. For example, to quickly find the domains that are attacked and the number of attacks, click the number in the Attacked Hosts graph of the Security Center report.

Note:
Alternatively, switch to the Raw Log tab to find the relevant log entries.

Description of values in default dashboards

- Operation Center: Displays operation details such as the proportion of valid requests and the statistics of attacks, traffic details such as the peak of both inbound and outbound throughput and the number of requests received, the operation trend, the attack overview, and other information.

<table>
<thead>
<tr>
<th>Graph</th>
<th>Type</th>
<th>Default time range</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid Request Ratio</td>
<td>Single value</td>
<td>Today (time frame)</td>
<td>Displays the percentage of valid requests in all requests. A valid request is a request that is neither an attack nor a request that is blocked by a 400 error.</td>
<td>95%</td>
</tr>
<tr>
<td>Graph</td>
<td>Type</td>
<td>Default time range</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>----------------------------</td>
<td>---------------</td>
<td>--------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Valid Request Traffic Ratio</td>
<td>Single value</td>
<td>Today (time frame)</td>
<td>Displays the percentage of the traffic generated by valid requests in the traffic generated by all requests.</td>
<td>95%</td>
</tr>
<tr>
<td>Peak Attack Size</td>
<td>Single value</td>
<td>Today (time frame)</td>
<td>Displays the peak of attack traffic, which is measured in Bps.</td>
<td>100 B/s</td>
</tr>
<tr>
<td>Attack Traffic</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>Displays the total attack traffic, which is measured in B.</td>
<td>30 B</td>
</tr>
<tr>
<td>Attack Count</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>The total number of attacks.</td>
<td>100</td>
</tr>
<tr>
<td>Peak Network In</td>
<td>Single value</td>
<td>Today (time frame)</td>
<td>Displays the peak inbound throughput, which is measured in KB/s.</td>
<td>100 KB/s</td>
</tr>
<tr>
<td>Peak Network Out</td>
<td>Single value</td>
<td>Today (time frame)</td>
<td>Displays the peak outbound throughput, which is measured in KB/s.</td>
<td>100 KB/s</td>
</tr>
<tr>
<td>Received Requests</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>Displays the total number of valid requests.</td>
<td>7,800</td>
</tr>
<tr>
<td>Received traffic</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>Displays the total inbound traffic that is generated by valid requests, which is measured in MB.</td>
<td>1.4 MB</td>
</tr>
<tr>
<td>Traffic Out</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>Displays the total outbound traffic that is generated by valid requests, which is measured in MB.</td>
<td>3.8 MB</td>
</tr>
<tr>
<td>Graph</td>
<td>Type</td>
<td>Default time range</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------</td>
<td>------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Network Traffic In And Attack</td>
<td>Area chart</td>
<td>Today (time frame)</td>
<td>Displays the trends of throughput generated by valid requests and attacks, which is measured in Kbit/s.</td>
<td></td>
</tr>
<tr>
<td>Request And Interception</td>
<td>Line chart</td>
<td>Today (time frame)</td>
<td>Displays the trends of valid requests and requests that are intercepted, which is measure in Kbit/h.</td>
<td></td>
</tr>
<tr>
<td>Access Status Distribution</td>
<td>Flow chart</td>
<td>Today (time frame)</td>
<td>Displays the trends of requests with different status codes (404, 304, 200, and other status codes), which is measured in Kbit/h.</td>
<td></td>
</tr>
<tr>
<td>Attack Source (World)</td>
<td>World map</td>
<td>1 hour (relative)</td>
<td>Displays the distribution of attackers by country.</td>
<td></td>
</tr>
<tr>
<td>Attack Source (China)</td>
<td>Map of China</td>
<td>1 Hour (Relative)</td>
<td>Displays the distribution of attackers in China by province.</td>
<td></td>
</tr>
<tr>
<td>Attack Type</td>
<td>Pie chart</td>
<td>1 hour (relative)</td>
<td>Displays the distribution of attacks by attack type.</td>
<td></td>
</tr>
<tr>
<td>Attacked Hosts</td>
<td>Tree map</td>
<td>1 hour (relative)</td>
<td>Displays the domains that are attacked and the number of attacks.</td>
<td></td>
</tr>
</tbody>
</table>
- Access center: Displays basic access details such as the number of PV and the number of UV, the access trend, the distribution of visitors, and other information.

<table>
<thead>
<tr>
<th>Graph</th>
<th>Type</th>
<th>Default time range</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>PV</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>Displays the total number of PV.</td>
<td>100,000</td>
</tr>
<tr>
<td>UV</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>Displays the total number of UV.</td>
<td>100</td>
</tr>
<tr>
<td>Traffic In</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>Displays the total inbound traffic, which is measured in MB.</td>
<td>300 MB</td>
</tr>
<tr>
<td>Peak Network In Traffic</td>
<td>Single value</td>
<td>Today (time frame)</td>
<td>Displays the peak inbound throughput, which is measured in KB/s.</td>
<td>0.5 KB/s</td>
</tr>
<tr>
<td>Peak Network Out Traffic</td>
<td>Single value</td>
<td>Today (time frame)</td>
<td>Displays the peak outbound throughput, which is measured in KB/s.</td>
<td>1.3 KB/s</td>
</tr>
<tr>
<td>Traffic Network Trend</td>
<td>Area chart</td>
<td>Today (time frame)</td>
<td>Displays the trends of inbound and outbound throughput, which are measured in KB/s.</td>
<td>-</td>
</tr>
<tr>
<td>PV/UV Trends</td>
<td>Line chart</td>
<td>Today (time frame)</td>
<td>Displays the trends of PV and UV, which is measured in Kbit/h.</td>
<td>-</td>
</tr>
<tr>
<td>Access Status Distribution</td>
<td>Flow chart</td>
<td>Today (time frame)</td>
<td>Displays the trends of requests with different status codes (404, 304, 200, and other status code), which is measured in Kbit/h.</td>
<td>-</td>
</tr>
<tr>
<td>Graph</td>
<td>Type</td>
<td>Default time range</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------</td>
<td>--------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Access Source</td>
<td>World map</td>
<td>1 hour (relative)</td>
<td>Displays the distribution of attackers by country.</td>
<td>-</td>
</tr>
<tr>
<td>Traffic In Source (World)</td>
<td>World map</td>
<td>1 hour (relative)</td>
<td>Displays the distribution (by country) of inbound traffic from requests.</td>
<td>-</td>
</tr>
<tr>
<td>Traffic In Source (China)</td>
<td>Map of China</td>
<td>1 hour (relative)</td>
<td>Displays the distribution (by province) of inbound traffic from requests in China.</td>
<td>-</td>
</tr>
<tr>
<td>Access Heatmap</td>
<td>Amap</td>
<td>1 hour (relative)</td>
<td>Displays the heatmap that indicates the source distribution of requests by geographical position.</td>
<td>-</td>
</tr>
<tr>
<td>Network Provider Source</td>
<td>Pie chart</td>
<td>1 hour (relative)</td>
<td>Displays the source distribution of requests by Internet service provider that provides network for the source, such as China Telecom, China Unicom, China Mobile, and universities.</td>
<td>-</td>
</tr>
<tr>
<td>Referer</td>
<td>Table</td>
<td>1 hour (relative)</td>
<td>Displays the first 100 referer URLs which the hosts are most often redirected from, and displays the information of hosts and redirection frequency.</td>
<td>-</td>
</tr>
<tr>
<td>Graph</td>
<td>Type</td>
<td>Default time range</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>------------------------------</td>
<td>------------</td>
<td>--------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Mobile Client Distribution</td>
<td>Pie chart</td>
<td>1 hour (relative)</td>
<td>Displays the distribution of requests from mobile clients, by client type.</td>
<td>-</td>
</tr>
<tr>
<td>PC Client Distribution</td>
<td>Pie chart</td>
<td>1 hour (relative)</td>
<td>Displays the distribution of requests from PC clients, by client type.</td>
<td>-</td>
</tr>
<tr>
<td>Request Content Type</td>
<td>Pie chart</td>
<td>1 hour (relative)</td>
<td>Displays the distribution of request sources by content type, such as HTML, form, JSON, and streaming data.</td>
<td>-</td>
</tr>
<tr>
<td>Accessed Sites</td>
<td>Tree map</td>
<td>1 Hour (Relative)</td>
<td>Displays the addresses of 30 domains that are visited most.</td>
<td>-</td>
</tr>
<tr>
<td>Top Clients</td>
<td>Table</td>
<td>1 hour (relative)</td>
<td>Displays the information of 100 clients that visit your domains most. The information includes the client IP address, the region and city, network information, the request method, inbound traffic, the number of incorrect accesses, the number of attacks, and other information.</td>
<td>-</td>
</tr>
<tr>
<td>Graph</td>
<td>Type</td>
<td>Default time range</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>---------</td>
<td>--------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>URL With Slowest Response</td>
<td>Table</td>
<td>1 hour (relative)</td>
<td>Displays the information of 100 URLs that have the longest response times. The information includes the website address, the URL, the average response time, the number of accesses, and other information.</td>
<td></td>
</tr>
</tbody>
</table>

- Security Center: Displays basic details of attacks, attack types, the attack trend, the distribution of attackers, and other information.

<table>
<thead>
<tr>
<th>Chart</th>
<th>Type</th>
<th>Default time range</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak Attack Size</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>Displays the peak of the throughput when your website is suffering attacks, which is measured in Bps.</td>
<td>100 B/s</td>
</tr>
<tr>
<td>Attacked Hosts</td>
<td>Single value</td>
<td>Today (time frame)</td>
<td>Displays the number of domains that are attacked.</td>
<td>3</td>
</tr>
<tr>
<td>Source Country Of Attack</td>
<td>Single value</td>
<td>Today (time frame)</td>
<td>Displays the number of countries that are attack sources.</td>
<td>2</td>
</tr>
<tr>
<td>Attack Traffic</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>Displays the total amount of traffic that is generated by attacks, which is measured in B.</td>
<td>1 B</td>
</tr>
<tr>
<td>Chart</td>
<td>Type</td>
<td>Default time range</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------</td>
<td>--------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Attacker UV</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>Displays the number of unique clients that are attack sources.</td>
<td>40</td>
</tr>
<tr>
<td>Attack type distribution</td>
<td>Flow chart</td>
<td>Today (time frame)</td>
<td>Displays the distribution of attacks by attack type.</td>
<td>-</td>
</tr>
<tr>
<td>Intercepted Attack</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>Displays the number of attacks that are intercepted by WAF.</td>
<td>100</td>
</tr>
<tr>
<td>HTTP flood attack Interception</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>Displays the number of HTTP flood attacks that are intercepted by WAF.</td>
<td>10</td>
</tr>
<tr>
<td>Web Attack Interception</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>Displays the number of Web application attacks that are intercepted by WAF.</td>
<td>80</td>
</tr>
<tr>
<td>Access Control Event</td>
<td>Single value</td>
<td>1 hour (relative)</td>
<td>Displays the number of requests that are intercepted by the HTTP ACL policies of WAF.</td>
<td>10</td>
</tr>
<tr>
<td>HTTP flood attack (World)</td>
<td>World map</td>
<td>1 hour (relative)</td>
<td>Displays the distribution of HTTP flood attackers by country.</td>
<td>-</td>
</tr>
<tr>
<td>HTTP flood attack (China)</td>
<td>China map</td>
<td>1 hour (relative)</td>
<td>Displays the distribution of HTTP flood attackers by province in China.</td>
<td>-</td>
</tr>
<tr>
<td>Web Attack (World)</td>
<td>World map</td>
<td>1 Hour (Relative)</td>
<td>Displays the distribution of Web application attacks by country.</td>
<td>-</td>
</tr>
<tr>
<td>Chart</td>
<td>Type</td>
<td>Default time range</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>---------------------</td>
<td>--------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Web Attack (China)</td>
<td>Map of China</td>
<td>1 hour (relative)</td>
<td>Displays the distribution of Web application attacks by province in China.</td>
<td>-</td>
</tr>
<tr>
<td>Access Control Attack (World)</td>
<td>World Map</td>
<td>1 hour (relative)</td>
<td>Displays the distribution by country of requests that are intercepted by the HTTP ACL policies of WAF.</td>
<td>-</td>
</tr>
<tr>
<td>Access Control Attack (China)</td>
<td>Map of China</td>
<td>1 Hour (Relative)</td>
<td>Displays the distribution by province in China of requests that are intercepted by the HTTP ACL policy of WAF.</td>
<td>-</td>
</tr>
<tr>
<td>Attacked Hosts</td>
<td>Tree map</td>
<td>1 hour (relative)</td>
<td>Displays the websites that are attacked most.</td>
<td>-</td>
</tr>
<tr>
<td>HTTP flood attack Strategy Distribution</td>
<td>Pie chart</td>
<td>1 hour (relative)</td>
<td>Displays the distribution of security policies being activated for HTTP flood attacks.</td>
<td>-</td>
</tr>
<tr>
<td>Web Attack Type Distribution</td>
<td>Pie chart</td>
<td>1 hour (relative)</td>
<td>Displays the distribution of Web attacks by attack type.</td>
<td>-</td>
</tr>
</tbody>
</table>
### Chart  | Type  | Default time range | Description | Example
--- | --- | --- | --- | ---
Top Attackers | Table | 1 hour (relative) | Displays IP addresses, provinces, and network providers of the first 100 clients that launch the recent attacks, and displays the number of attacks and the amount of traffic generated by these attacks. | -
Attacker Referer | Table | 1 Hour (Relative) | Displays the information in referers of attack requests, which includes referer URLs, referer hosts, and the number of attacks. | -

#### 4.9.7 Fields in the log entry

WAF keeps detailed log entries for your domains, including access requests and attack logs. Each log entry contains dozens of fields. You can perform query and analysis based on specific fields.

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>topic</strong></td>
<td>The topic of the log entry. The value of this field is waf_access_log, which cannot be changed.</td>
<td>waf_access_log</td>
</tr>
<tr>
<td>acl_action</td>
<td>The action generated by the WAF HTTP ACL policy to the request, such as pass, drop, and captcha.</td>
<td>pass</td>
</tr>
</tbody>
</table>

**Note:**
If the value is null or -, it indicates that the action is pass.
<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>acl_blocks</td>
<td>Indicates whether the request is blocked by the HTTP ACL policy.</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>• If the value is 1, the request is blocked.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• If the value is not 1, the request is passed.</td>
<td></td>
</tr>
<tr>
<td>antibot</td>
<td>The type of the Anti-Bot Service protection strategy that applies, which includes:</td>
<td>ratelimit</td>
</tr>
<tr>
<td></td>
<td>• ratelimit: Frequency control</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• sdk: APP protection</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• intelligence: Algorithmic model</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• acl: HTTP ACL policy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• blacklist: Blacklist</td>
<td></td>
</tr>
<tr>
<td>antibot_action</td>
<td>The action performed by the Anti-Bot Service protection strategy, which includes:</td>
<td>challenge</td>
</tr>
<tr>
<td></td>
<td>• challenge: Verifying using an embedded JavaScript script</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• drop: Blocking</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• report: Logging the access event</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• captcha: Verifying using a slider captcha</td>
<td></td>
</tr>
<tr>
<td>Field</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>block_action</td>
<td>The type of the WAF protection that is activated, which includes:</td>
<td>tmd</td>
</tr>
<tr>
<td></td>
<td>• tmd: Protection against HTTP flood attacks</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• waf: Protection against Web application attacks</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• acl: HTTP ACL policy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• geo: Blocking regions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• antifraud: Risk control for data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• antibot: Blocking Web crawlers</td>
<td></td>
</tr>
<tr>
<td>body_bytes_sent</td>
<td>The size of the body in the access request, which is measured in Bytes.</td>
<td>2</td>
</tr>
<tr>
<td>cc_action</td>
<td>Protection strategies against HTTP flood attacks, such as none, challenge,</td>
<td>close</td>
</tr>
<tr>
<td></td>
<td>pass, close, captcha, wait, login, and n.</td>
<td></td>
</tr>
<tr>
<td>cc_blocks</td>
<td>Indicates whether the request is blocked by the CC protection.</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>• If the value is 1, the request is blocked.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• If the value is not 1, the request is passed.</td>
<td></td>
</tr>
<tr>
<td>cc_phase</td>
<td>The CC protection strategy that is activated, which can be seccookie,</td>
<td>server_ip_blacklist</td>
</tr>
<tr>
<td></td>
<td>server_ip_blacklist, static_whitelist, server_header_blacklist, server_</td>
<td></td>
</tr>
<tr>
<td></td>
<td>cookie_blacklist, server_cookie_blacklist, server_args_blacklist, server_</td>
<td></td>
</tr>
<tr>
<td></td>
<td>argus_blacklist, or qps_overmax.</td>
<td></td>
</tr>
<tr>
<td>content_type</td>
<td>The content type of the access request.</td>
<td></td>
</tr>
<tr>
<td>host</td>
<td>The source website.</td>
<td>api.aliyun.com</td>
</tr>
<tr>
<td>http_cookie</td>
<td>The client-side cookie, which is included in the request header.</td>
<td>k1=v1;k2=v2</td>
</tr>
<tr>
<td>Field</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>---------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>http_referer</td>
<td>The URL information of the request source, which is included in the request header. - indicates no URL information.</td>
<td><a href="http://xyz.com">http://xyz.com</a></td>
</tr>
<tr>
<td>http_user_agent</td>
<td>The User Agent field in the request header, which contains information such as the client browser and the operating system.</td>
<td>Dalvik/2.1.0 (Linux; U; Android 7.0; EDI-AL10 Build/HUAWEIEDISON-AL10)</td>
</tr>
<tr>
<td>http_x_forwarded_for</td>
<td>The X-Forwarded-For (XFF) information in the request header, which identifies the original IP address of the client that connects to the Web server using a HTTP proxy or load balancing.</td>
<td>-</td>
</tr>
<tr>
<td>https</td>
<td>Indicates whether the request is an HTTPS request.</td>
<td>true</td>
</tr>
<tr>
<td></td>
<td>• true: the request is an HTTPS request.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• false: the request is an HTTP request.</td>
<td></td>
</tr>
<tr>
<td>matched_host</td>
<td>The matched domain name (extensive domain name) that is protected by WAF. If no domain has been matched, the value is -.</td>
<td>*.aliyun.com</td>
</tr>
<tr>
<td>querystring</td>
<td>The query string in the request.</td>
<td>title=tm_content%3Darticle&amp;pid=123</td>
</tr>
<tr>
<td>real_client_ip</td>
<td>The real IP address of the client. If the system cannot get the real IP address, the value is -.</td>
<td>1.2.3.4</td>
</tr>
<tr>
<td>region</td>
<td>The information of the region where the WAF instance is located.</td>
<td>cn</td>
</tr>
<tr>
<td>remote_addr</td>
<td>The IP address of the client that sends the access request.</td>
<td>1.2.3.4</td>
</tr>
<tr>
<td>Field</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>remote_port</td>
<td>The port of the client that sends the access request.</td>
<td>3242</td>
</tr>
<tr>
<td>request_length</td>
<td>The size of the request, measured in Bytes.</td>
<td>123</td>
</tr>
<tr>
<td>request_method</td>
<td>The HTTP request method used in the access request.</td>
<td>GET</td>
</tr>
<tr>
<td>request_path</td>
<td>The relative path of the request. The query string is not included.</td>
<td>/news/search.php</td>
</tr>
<tr>
<td>request_time_msec</td>
<td>The request time, which is measured in microseconds.</td>
<td>44</td>
</tr>
<tr>
<td>request_traceid</td>
<td>The unique ID of the access request that is recorded by WAF.</td>
<td>7837b********************ea1f0</td>
</tr>
<tr>
<td>server_protocol</td>
<td>The response protocol and the version number of the origin server.</td>
<td>HTTP/1.1</td>
</tr>
<tr>
<td>status</td>
<td>The status of the HTTP response to the client returned by WAF.</td>
<td>200</td>
</tr>
<tr>
<td>time</td>
<td>The time when the access request occurs.</td>
<td>2018-05-02T16:03:59+08:00</td>
</tr>
<tr>
<td>ua_browser</td>
<td>The information of the browser that sends the request.</td>
<td>ie9</td>
</tr>
<tr>
<td>ua_browser_family</td>
<td>The family of the browser that sent the request.</td>
<td>internet explorer</td>
</tr>
<tr>
<td>ua_browser_type</td>
<td>The type of the browser that sent the request.</td>
<td>web_browser</td>
</tr>
<tr>
<td>ua_browser_version</td>
<td>The version of the browser that sends the request.</td>
<td>9.0</td>
</tr>
<tr>
<td>ua_device_type</td>
<td>The type of the client device that sends the request.</td>
<td>computer</td>
</tr>
<tr>
<td>ua_os</td>
<td>The operating system used by the client that sends the request.</td>
<td>windows_7</td>
</tr>
<tr>
<td>Field</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>---------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>ua_os_family</td>
<td>The family of the operating system used by the client.</td>
<td>windows</td>
</tr>
<tr>
<td>upstream_addr</td>
<td>A list of origin addresses, separated by commas. The format of an address is IP:Port.</td>
<td>1.2.3.4:443</td>
</tr>
<tr>
<td>upstream_ip</td>
<td>The origin IP address that corresponds to the access request. For example, if the origin server is an ECS instance, the value of this field is the IP address of the ECS instance.</td>
<td>1.2.3.4</td>
</tr>
<tr>
<td>upstream_response_time</td>
<td>The time that the origin site takes to respond to the WAF request, which is measured in seconds. &quot;-&quot; indicates the timeout of the request.</td>
<td>0.044</td>
</tr>
<tr>
<td>upstream_status</td>
<td>The response status that WAF receives from the origin server. &quot;-&quot; indicates that no response is received. The reason can be the response timeout, or the request being blocked by WAF.</td>
<td>200</td>
</tr>
<tr>
<td>user_id</td>
<td>Alibaba Cloud account ID.</td>
<td>12345678</td>
</tr>
<tr>
<td>waf_action</td>
<td>The action from the Web attack protection policy.</td>
<td>block</td>
</tr>
<tr>
<td></td>
<td>• If the value is block, the attack is blocked.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• If the value is bypass or other values, the attack is ignored.</td>
<td></td>
</tr>
<tr>
<td>web_attack_type</td>
<td>The Web attack type such as xss, code_exec, webshell, sqli, Ifilei, rfiei, and other.</td>
<td>xss</td>
</tr>
<tr>
<td>waf_rule_id</td>
<td>The ID of the WAF rule that is matched.</td>
<td>100</td>
</tr>
</tbody>
</table>

**4.9.8 Advanced settings**

If you click Advanced Settings on the page of WAF log query and analysis service, you will be redirected to the Log Service console. Then you can set advanced...
features for Log Service. For example, you can set alarms and notifications, real-time log collection and consumption, shipping log data, or provide visual representations with other products.

Procedure

1. Log on to the Web Application Firewall console, choose App Market > App Management.
2. Click the Real-time Log Query and Analysis Service area to open the Log Service page.
3. Click Advanced Settings in the upper-right corner.
4. In the dialog box that appears, click Go to open the Log Service console.
5. In the Log Service console, you can set the following advanced features for log projects and logstores:
   - Real-time log collection and consumption
   - Shipping log data to other Alibaba Cloud storage services in real time
   - Providing visual representations with other products

4.9.9 Grant log query and analysis permissions to a RAM user

If you want to use the WAF log query and analysis service with a RAM user, you must grant required permissions to the RAM user using the Alibaba Cloud account.

Context

The following permissions are required for enabling and using the WAF log query and analysis service.

<table>
<thead>
<tr>
<th>Operation</th>
<th>Required account type and permissions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enable Log Service (the service remains enabled after this operation)</td>
<td>Alibaba Cloud account</td>
</tr>
</tbody>
</table>
| Authorize WAF to write log data to the exclusive logstore in Log Service in real-time (the authorization remains valid after this operation) | · Alibaba Cloud account  
· RAM user that has the AliyunLogFullAccess permission  
· RAM user that has specific permissions |
<table>
<thead>
<tr>
<th>Operation</th>
<th>Required account type and permissions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use the log query and analysis service</td>
<td>· Alibaba Cloud account</td>
</tr>
<tr>
<td></td>
<td>· RAM user that has the AliyunLogFullAccess permission</td>
</tr>
<tr>
<td></td>
<td>· RAM user that has specific permissions</td>
</tr>
</tbody>
</table>

Grant permissions to RAM users as required.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Permission</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grant permissions on all Log Service operations to a RAM user.</td>
<td>AliyunLogFullAccess</td>
<td>For more information, see RAM users.</td>
</tr>
<tr>
<td>Grant the log viewing permission to a RAM user after you enable the WAF log query and analysis service and complete the authorization on the Alibaba Cloud account.</td>
<td>AliyunLogReadOnlyAccess</td>
<td>For more information, see RAM users.</td>
</tr>
<tr>
<td>Grant the RAM user permissions on enabling and using the WAF log query and analysis service. This RAM user is not granted other administrative permissions on Log Service.</td>
<td>Custom authorization policy</td>
<td>For more information, see the following procedure.</td>
</tr>
</tbody>
</table>

Procedure

1. Log on to the *RAM console*.

2. On the Policies page, select the Custom Policy tab.

3. In the upper-right corner of the page, click Create Authorization Policy.

4. Click Create Authorization Policy. In the template, specify the Authorization Policy Name, and then enter the following in the Policy Content field.

---

**Note:**
Replace `${Project}` and `${Logstore}` in the following policy content with the names of the exclusive project and logstore in WAF Log Service.

```json
{
  "Version": "1",
  "Statement": [
    {
      "Action": "log:GetProject",
      "Resource": "acs:log:*:*:project/${Project}",
      "Effect": "Allow"
    },
    {
      "Action": "log:CreateProject",
      "Resource": "acs:log:*:*:project/*",
      "Effect": "Allow"
    },
    {
      "Action": "log:ListLogStores",
      "Resource": "acs:log:*:*:project/${Project}/logstore/*",
      "Effect": "Allow"
    },
    {
      "Action": "log:CreateLogStore",
      "Resource": "acs:log:*:*:project/${Project}/logstore/*",
      "Effect": "Allow"
    },
    {
      "Action": "log:GetIndex",
      "Resource": "acs:log:*:*:project/${Project}/logstore/${Logstore}",
      "Effect": "Allow"
    },
    {
      "Action": "log:CreateIndex",
      "Resource": "acs:log:*:*:project/${Project}/logstore/${Logstore}",
      "Effect": "Allow"
    },
    {
      "Action": "log:UpdateIndex",
      "Resource": "acs:log:*:*:project/${Project}/logstore/${Logstore}",
      "Effect": "Allow"
    },
    {
      "Action": "log:CreateDashboard",
      "Resource": "acs:log:*:*:project/${Project}/dashboard/*",
      "Effect": "Allow"
    },
    {
      "Action": "log:UpdateDashboard",
      "Resource": "acs:log:*:*:project/${Project}/dashboard/*",
      "Effect": "Allow"
    },
    {
      "Action": "log:CreateSavedSearch",
      "Resource": "acs:log:*:*:project/${Project}/savedsearch/*",
      "Effect": "Allow"
    }
  ]
}
```
5. Click Create Authorization Policy.

6. Go to the Users page, find the RAM user, and then click Authorize.

7. Add the authorization policy that you created and click OK.

   This RAM user can enable and use the WAF log query and analysis service, and cannot use other features of Log Service.

4.9.10 Manage log storage

After WAF Log Service is activated, log storage is allocated for your WAF Log Service based on the specified log storage size. You can view the usage of the log storage on the Log Service page in the Web Application Firewall console.

View the usage of the log storage

You can view the usage of the log storage that is generated by the WAF log query and analysis service at any time.

Note:

It takes two hours for changes in the storage usage to be updated in the console. You need to upgrade the log storage when only a little log storage space is available.

1. Log on to the Web Application Firewall console.

2. Choose App Market > App Management, select the region where your WAF instance is located, and then click Real-time Log Query and Analysis Service.

3. At the top of the Log Service page, view the usage of log storage.

Upgrade log storage

To upgrade the log storage size, click Upgrade Storage at the top of the Log Service page.

Note:
If log storage is full, new log data cannot be written to the exclusive logstore. We recommend that you upgrade log storage before log storage is full.

Clear log storage

You can delete all log entries in the log storage as needed. For example, you can delete the log entries generated during the test phase to make full use of the log storage by recording only log entries that is generated during the production phase.

Click Clear at the top of the Log Service page, and click Confirm to delete all log entries in the log storage.

Notice:
Log entries that are deleted cannot be restored. Delete log entries with caution.

Note:
You can clear the log storage for only a limited number of times.

4.10 Anti-Bot logs

4.10.1 Enable Log Service for Anti-Bot

Log Service can collect access logs and protection logs from the websites protected by Anti-Bot in real time. Also, it can retrieve and analyze the collected log data in real time.

You can analyze the website access and attack behaviors based on the website logs collected in the Anti-Bot console in real time. This further allows you to assist your security management personnel in developing protection policies.

Procedure

1. Log on to the Anti-Bot console.
2. Choose Reports > Log Service. Select the region where your instance is located.

Note:
If you are using Log Service for Anti-Bot for the first time, click Authorize to authorize Anti-Bot to store all the recorded logs in your logstore as instructed.
3. From the Website Domain drop-down list, select the website domain name for which you want to enable Log Service. Then, click Enable.

Note:
The Website Domain drop-down list displays all the website domain names configured with Anti-Bot.

Now, Log Service has been enabled for the website domain name. Log Service automatically creates a dedicated logstore for your Alibaba Cloud account. Anti-Bot automatically imports the logs of all the website domain names enabled with Log Service to the dedicated logstore in real time.

Then, you can retrieve and analyze the access logs of the website domain names enabled with Log Service.

Restrictions and instructions

- Other data cannot be written to the dedicated logstore.

Note:
The website logs recorded by Anti-Bot are stored in the dedicated logstore, where you cannot write other data through APIs and SDKs.
Currently, the basic settings (such as the storage period) of the dedicated logstore cannot be modified.

Do not delete or modify the default settings created by Log Service, such as the default project, logstore, index, and dashboard.

Log Service updates and upgrades the log query analysis function from time to time. The indexes and default reports of the dedicated logstore will be automatically updated.

If the RAM user requires the log query analysis function, grant the related Log Service permissions to the RAM user through RAM.

### 4.10.2 Log field description

Log Service for Anti-Bot Service (Anti-Bot) records the access logs and attack and defense logs of protected website domain names in detail. A log contains dozens of fields. You can select specific fields for query analysis as needed.

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>topic</strong></td>
<td>The log topic. This field is invariably set to antibot_access_log.</td>
<td>antibot_access_log</td>
</tr>
<tr>
<td>antibot</td>
<td>The type of the triggered Anti-Bot protection policy, including:</td>
<td>ratelimit</td>
</tr>
<tr>
<td></td>
<td>• ratelimit: rate limiting</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• sdk: app protection</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• algorithm: algorithm pattern</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• intelligence: bot intelligence</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• acl: access control list</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• blacklist: blacklist</td>
<td></td>
</tr>
<tr>
<td>antibot_action</td>
<td>The operation specified by the Anti-Bot protection policy, including:</td>
<td>drop</td>
</tr>
<tr>
<td></td>
<td>• challenge: Deliver a JavaScript script for verification</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• drop: Intercept</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• captcha: Verify by dragging a slider</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• report: Monitor only</td>
<td></td>
</tr>
<tr>
<td>Field</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>antibot_rule</td>
<td>The ID of the triggered Anti-Bot protection rule.</td>
<td>5472</td>
</tr>
<tr>
<td>antibot_verify</td>
<td>The result of the verification performed by Anti-Bot.</td>
<td>challenge_fail</td>
</tr>
<tr>
<td></td>
<td>Note: This value is recorded when the antibot_action field is set to challenge or captcha.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• challenge_fail: JavaScript verification fails.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• challenge_pass: JavaScript verification is passed.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• captcha_fail: Slide captcha verification fails.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• captcha_pass: Slide captcha verification is passed.</td>
<td></td>
</tr>
<tr>
<td>block_action</td>
<td>The type of the bot protection that is triggered. The value is invariably set to antibot.</td>
<td>antibot</td>
</tr>
<tr>
<td>body_bytes_sent</td>
<td>The size of HTTP body (in byte) sent to the client.</td>
<td>2</td>
</tr>
<tr>
<td>content_type</td>
<td>The content type of the access request.</td>
<td>application/x-www-form-urlencoded</td>
</tr>
<tr>
<td>host</td>
<td>The source website.</td>
<td>api.aliyun.com</td>
</tr>
<tr>
<td>http_cookie</td>
<td>The cookie information about the access client, which is included in the access request header.</td>
<td>k1=v1;k2=v2</td>
</tr>
<tr>
<td>http_referer</td>
<td>The source URL of the access request, which is included in the access request header. – is displayed if no source URL is available.</td>
<td><a href="http://xyz.com">http://xyz.com</a></td>
</tr>
<tr>
<td>Field</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------</td>
</tr>
<tr>
<td>http_user_agent</td>
<td>The User Agent field in the access request header, which typically includes the web browser identifier and operating system identifier of the source client.</td>
<td>Dalvik/2.1.0 (Linux; U; Android 7.0; EDI-AL10 Build/HUAWEIEDISON-AL10)</td>
</tr>
<tr>
<td>http_x_forwarded_for</td>
<td>The XFF header information in the access request header, which is used to identify the original IP addresses of the clients connected to a web server through the HTTP proxy or SLB.</td>
<td>-</td>
</tr>
<tr>
<td>https</td>
<td>Whether the access request is an HTTPS request. Valid values:</td>
<td>true</td>
</tr>
<tr>
<td></td>
<td>• true: The access request is an HTTPS request.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• false: The access request is an HTTP request.</td>
<td></td>
</tr>
<tr>
<td>matched_host</td>
<td>The matched domain name configured with Anti-Bot, which may be a wildcard domain name. - is displayed if no related domain name configuration is matched.</td>
<td>*.aliyun.com</td>
</tr>
<tr>
<td>real_client_ip</td>
<td>The actual IP address of the access client. - is displayed if no actual IP address is retrieved.</td>
<td>1.2.3.4</td>
</tr>
<tr>
<td>region</td>
<td>The information about the region where the Anti-Bot instance is located.</td>
<td>cn</td>
</tr>
<tr>
<td>remote_addr</td>
<td>The IP address of the client that initiates the access request.</td>
<td>1.2.3.4</td>
</tr>
<tr>
<td>remote_port</td>
<td>The port of the client that initiates the access request.</td>
<td>23713</td>
</tr>
<tr>
<td>request_length</td>
<td>The length of the access request. Unit: bytes.</td>
<td>123</td>
</tr>
<tr>
<td>Field</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>request_method</td>
<td>The HTTP request method of the access request.</td>
<td>GET</td>
</tr>
<tr>
<td>request_path</td>
<td>The relative path of the request (excluding the query string).</td>
<td>/news/search.php</td>
</tr>
<tr>
<td>request_time_msec</td>
<td>The duration of the access request. Unit: milliseconds.</td>
<td>44</td>
</tr>
<tr>
<td>request_traceid</td>
<td>The unique ID of the access request.</td>
<td>7837b11715410386943437009ea1f0</td>
</tr>
<tr>
<td>server_protocol</td>
<td>The protocol and version of the response returned by the origin server.</td>
<td>HTTP/1.1</td>
</tr>
<tr>
<td>status</td>
<td>The status of the HTTP response that Anti-Bot returns to the client.</td>
<td>200</td>
</tr>
<tr>
<td>time</td>
<td>The occurrence time of the access request.</td>
<td>2018-05-02T16:03:59+08:00</td>
</tr>
<tr>
<td>ua_browser</td>
<td>The information about the web browser that initiates the access request.</td>
<td>ie9</td>
</tr>
<tr>
<td>ua_browser_family</td>
<td>The family of the web browser that initiates the access request.</td>
<td>internet explorer</td>
</tr>
<tr>
<td>ua_browser_type</td>
<td>The type of the web browser that initiates the access request.</td>
<td>web_browser</td>
</tr>
<tr>
<td>ua_browser_version</td>
<td>The version of the web browser that initiates the access request.</td>
<td>9.0</td>
</tr>
<tr>
<td>ua_device_type</td>
<td>The device type of the client that initiates the access request.</td>
<td>computer</td>
</tr>
<tr>
<td>ua_os</td>
<td>The operating system of the client that initiates the access request.</td>
<td>windows_7</td>
</tr>
<tr>
<td>ua_os_family</td>
<td>The operating system family of the client that initiates the access request.</td>
<td>windows</td>
</tr>
<tr>
<td>Field</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>upstream_addr</td>
<td>The origin address list of Anti-Bot in the format of IP address:Port. Separate multiple IP addresses with commas (,).</td>
<td>1.2.3.4:443</td>
</tr>
<tr>
<td>upstream_ip</td>
<td>The origin IP address corresponding to the access request. For example, if Anti-Bot forwards the access request to an ECS instance, this parameter returns the IP address of the back-to-origin ECS instance.</td>
<td>1.2.3.4</td>
</tr>
<tr>
<td>upstream_response_time</td>
<td>The time for the origin server to respond to an Anti-Bot request. Unit: seconds. The response times out if &quot;-&quot; is returned.</td>
<td>0.044</td>
</tr>
<tr>
<td>upstream_status</td>
<td>The status of the response that the origin server returns to Anti-Bot. No response is available if &quot;,-&quot; is returned. For example, the request is intercepted by Anti-Bot, or the response returned by the origin server times out.</td>
<td>200</td>
</tr>
<tr>
<td>user_id</td>
<td>AliUID of the Alibaba Cloud account.</td>
<td>12345678</td>
</tr>
<tr>
<td>wxbb_action</td>
<td>If the protection type of Anti-Bot is app protection, the following actions are supported:</td>
<td>close</td>
</tr>
<tr>
<td></td>
<td>· close: intercepts requests. That is, the antibot_action field is set to drop.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>· test: only monitors requests. That is, the antibot_action field is set to report.</td>
<td></td>
</tr>
<tr>
<td><strong>Note:</strong></td>
<td>This field is set to – if SDK protection is not configured.</td>
<td></td>
</tr>
</tbody>
</table>
4.11 ActionTrail access logs

4.11.1 Overview

At present, ActionTrail is in connection with Log Service, which provides functions of log collection and analysis in real time. The operation log data collected by ActionTrail is delivered to Log Service in real time. Log Service provides rich functions such as real-time query and analysis, and dashboard presentation for this part of logs.

Benefits

- Simple configuration: Easily configure to collect real-time logs. For information about configuration steps and log fields, see Procedure.
- Real-time analysis: Relying on Log Service, it provides real-time log analysis, an out-of-the-box report center, and details available for real-time mining with records of operations on important cloud assets.
- Real-time alarms: Supports custom quasi-real-time monitoring and alarming based on specific indicators to ensure timely response to critical business exceptions.
- Ecosystem: Supports dock with other ecosystems such as stream computing, cloud storage, and visualization solutions to further explore data value.
- Free quota: Provides 500 MB free quotas of data import and storage per month. You can expand the storage time for compliance, traceability, and filing. The storage service without time limitation is provided at a low price of 0.0875 USD/GB/month. For information about billing, see #unique_31.

Application scenarios

- Troubleshooting and analysis for abnormal operations
  Monitors cloud resource operations under all names in real time and supports real-time troubleshooting and analysis for abnormal operations. Accidental
deletion, high-risk operations, and other operations can be traced through logging.

For example, to view the Elastic Compute Service (ECS) release operation log:

Figure 4-46: View the ECS release operation log
• Distribution and source tracking of important resource operations

You can track and trace the distribution and source of important resource operations by analyzing the log content, and specify and optimize resolution strategies based on the analysis results.

For example, to view the country distribution of operators who deleted the Relational Database Service (RDS):

Figure 4-47: View the distribution of RDS deletion

• Resource operation distribution view

You can query and analyze the collected ActionTrail operation logs through SQL query statements in real time, and view the distribution and time trends of all resource operations, and other operation and maintenance actions. By doing this, you assist the operation and maintenance personnel to monitor the resource
running status in real time. Operation and maintenance reliability indicators are clear at a glance.

For example, to view trends of failed operations:

Figure 4-48: Trends of failed operations

- Real-time analysis of operation data
  Customize diverse query statements based on operation requirements, customize fast queries and analysis dashboard for different data requirements,
and you can also customize real-time data dashboard for data such as resource usage status and user logon status.

For example, to view the frequency distribution of operators from network operators:

Figure 4-49: Frequency distribution of operators from network operators

4.11.2 Procedure

At present, ActionTrail is in connection with Log Service. Operation log data collected by ActionTrail is delivered to Log Service in real time. This document introduces the log fields and collection procedures of ActionTrail logs.

Prerequisites

1. Enable Log Service
2. Enable ActionTrail service.

Procedure

1. Log on to the ActionTrail console.
2. Click Trail list in the left-side navigation pane to go to the Trail list page.
3. Click Create Trail in the upper-right corner to go to the Create Trail page.
4. Configure trail parameters.

   a. Enter Trail name.

   b. Deliver audit events to an OSS Bucket (optional).

      For more information, see Create trail.

   c. Select an region in Log Service Region.

   d. Enter Log Service Project

      The project is used to store ActionTrail logs. You can enter an existing project name under the selected region or enter a new project name to deliver the logs to the new project.

   e. Enable logging.

      Click Enable logging. After you enable this feature, operation logs of cloud resource recorded by your ActionTrail is delivered to Log Service.

Figure 4-50: Configure trail parameters.
5. Click Submit to complete the configuration.

You have created a trail and you can view the created trail in Trail List.

Note:
If you configure ActionTrail log collection for the first time, please authorize ActionTrail to upon prompts on the page. The authorization enables ActionTrail to distribute ActionTrail logs to your Logstore. Click Submit again after the authorization is complete to end the configuration.

Figure 4-51: Trail List

Limits

* Only one trail can be created for an account.

Trail helps you deliver audit events to an OSS bucket or Log Service Logstore specified by you. Currently, only one trail can be created for an account in all regions. This trail delivers audit events across all regions to both or either of the OSS bucket and Logstore.

* If you have created a trail, you can handle the trail in only the region where the trail was created.

If you have created a trail, you can view, modify, or delete the trail in only the region where the trail was created. For example, if you need to configure a trail of Log Service when you have created a trail of OSS, add Log Service configuration to your created trail of OSS.

* The exclusive Logstore does not support writing additional data.

The exclusive Logstore is used to store only operation logs of Action Trail. Therefore, this Logstore does not support writing other data. Other
functions, such as query, statistics, alarms, and streaming consumption, have no restrictions.

- Pay-As-You-Go.

The ActionTrail log collection feature uses the billing method of Log Service. Log Service supports Pay-As-You-Go billing method, and provides a certain amount of free quota. For more information, see #unique_31.

Query and analysis

To query and analyze collected log data after you complete trail configuration, click Log Analysis and Log Report under Log Service list in the Trail List page.

- Log Analysis: Enter the log query and analysis page.

  Log Service provides log query and analysis. In this page, you can query and analyze collected ActionTrail logs in real time.

  By defining query syntax and analysis syntax, Log Service provides log queries in a variety of complex scenarios. For information about query and analysis syntax, see Query syntax and Analysis syntax.

  To monitor important log data at intervals and set alarm notifications for abnormal conditions, save the current query conditions as quick queries and alarms on the query page. For detailed procedures, see #unique_189.

- Log Report: Enter the dashboard page.

  Log Service shows an overall view of real-time dynamics, such as event types and event sources, by a built-in dashboard exclusive to ActionTrail.

  You can modify the exclusive dashboard, create a custom dashboard, and add custom analysis charts in a variety of scenarios to your dashboard. For more information about dashboards, see #unique_190.

Default configuration

When the configuration is completed, Log Service creates an exclusive project and an exclusive Logstore for you. Operation logs of cloud resource collected by ActionTrail is delivered to the Logstore in real time. In addition, Log Service also creates a dashboard for you to view cloud resource operations in real time. For information about default configurations such as the project and Logstore, see the following table.
Table 4-12: Default configuration

<table>
<thead>
<tr>
<th>Default configuration item</th>
<th>Configuration content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project</td>
<td>A project that you select or customize when you create the trail.</td>
</tr>
<tr>
<td>Logstore</td>
<td>By default, Logstore is created. The Logstore name is <code>actiontrail_Trail name</code>. All logs of ActionTrail are saved in this Logstore.</td>
</tr>
<tr>
<td>Region</td>
<td>A region that you select when you create the trail.</td>
</tr>
<tr>
<td>Shard</td>
<td>By default, two shards are created and the Auto Split Shard feature is enabled.</td>
</tr>
<tr>
<td>Log storage time</td>
<td>By default, logs are saved permanently. You can customize the log storage time to a value in the range of 1 to 3000 days. For detailed procedures, see #unique_191.</td>
</tr>
</tbody>
</table>
| Dashboard                  | By default, a dashboard is created:  
  • Chinese environment: `actiontrail_Trail name_audit_center_cn`  
  • English environment: `actiontrail_Trail name_audit_center_en` |

Log field

<table>
<thead>
<tr>
<th>Field name</th>
<th>Name</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>topic</strong></td>
<td>Log topic.</td>
<td>This field is fixed at <code>actiontrail_audit_event</code></td>
</tr>
<tr>
<td>event</td>
<td>Event body, which is in the JSON format. The content of the event body varies with the event.</td>
<td><code>event example</code></td>
</tr>
<tr>
<td>event.eventId</td>
<td>The ID of the event, which uniquely indicates the event.</td>
<td><code>07F1234-3E1D-4BFF-AC6C-12345678</code></td>
</tr>
<tr>
<td>event.eventName</td>
<td>Event name.</td>
<td>CreateVSwitch</td>
</tr>
<tr>
<td>Field name</td>
<td>Name</td>
<td>Example</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>--------------------------------------------------</td>
<td>----------------------------------------------</td>
</tr>
<tr>
<td>event.eventType</td>
<td>Event type.</td>
<td>ApiCallApiCal</td>
</tr>
<tr>
<td>event.eventVersion</td>
<td>The version of the data format of ActionTrail, which is currently fixed to 1.</td>
<td>1</td>
</tr>
<tr>
<td>event.acsRegion</td>
<td>The region where the event is located.</td>
<td>cn-hangzhou</td>
</tr>
<tr>
<td>event.requestId</td>
<td>The request ID of the cloud service operation.</td>
<td>07F1234-3E1D-4BFF-AC6C-12345678</td>
</tr>
<tr>
<td>event.apiVersion</td>
<td>The version of the related API.</td>
<td>2017-12-04</td>
</tr>
<tr>
<td>event.errorMessage</td>
<td>The error message of an event failure.</td>
<td>unknown confidential</td>
</tr>
<tr>
<td>event.serviceName</td>
<td>The event-related service name.</td>
<td>Ecs</td>
</tr>
<tr>
<td>event.sourceIpAddress</td>
<td>The Source IP associated with the event.</td>
<td>1.2.3.4</td>
</tr>
<tr>
<td>event.userAgent</td>
<td>The event-related client agent.</td>
<td>Mozilla/5.0 (...)</td>
</tr>
<tr>
<td>event.requestParameters.HostId</td>
<td>The host ID in the request-related parameter.</td>
<td>ecs.cn-hangzhou.aliyuncs.com</td>
</tr>
<tr>
<td>event.requestParameters.Name</td>
<td>The name in the request-related parameter.</td>
<td>ecs-test</td>
</tr>
<tr>
<td>event.requestParameters.Region</td>
<td>The domain in the request-related parameter.</td>
<td>cn-hangzhou</td>
</tr>
<tr>
<td>event.userIdentity.accessKeyId</td>
<td>The AccessKey ID used by the request.</td>
<td>25**************</td>
</tr>
<tr>
<td>event.userIdentity.accountId</td>
<td>The ID of the account requested.</td>
<td>123456</td>
</tr>
<tr>
<td>event.userIdentity.principalId</td>
<td>The voucher ID of the account requested.</td>
<td>123456</td>
</tr>
<tr>
<td>event.userIdentity.type</td>
<td>The type of account requested.</td>
<td>root-account</td>
</tr>
</tbody>
</table>
### 4.12 Inner-ActionTrail

#### 4.12.1 Overview

Log Service is integrated with ActionTrail and provides the Inner-ActionTrail feature. This feature allows you to ship operations logs of Alibaba Cloud services from ActionTrail to Log Service in near real time for analysis and auditing.

Using the Inner-ActionTrail feature, you can quickly record and store operations logs of Alibaba Cloud services. Currently, only operations logs of Object Storage Service (OSS) are supported. Then, you can ship these logs to Log Service for analysis and auditing based on the search and analysis, reporting, alerting, and downstream computing and shipping features of Log Service.

**Benefits**

The Inner-ActionTrail feature has the following benefits:

---

<table>
<thead>
<tr>
<th>Field name</th>
<th>Name</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>event.userIdentity.</td>
<td>The name of account requested.</td>
<td>root</td>
</tr>
<tr>
<td>.userName</td>
<td>The name of account requested.</td>
<td>root</td>
</tr>
</tbody>
</table>

---

**event example**

```json
{
    "acsRegion": "cn-hangzhou",
    "additionalEventData": {
        "isMFAChecked": "false",
        "loginAccount": "test1234@aliyun.com"
    },
    "eventId": "7be1e173-1234-44a1-b135-1234",
    "eventName": "ConsoleSignin",
    "eventSource": "http://account.aliyun.com:443/login/login_aliyun.htm",
    "eventTime": "2018-07-12T06:14:50Z",
    "eventType": "ConsoleSignin",
    "requestId": "7be1e173-1234-44a1-b135-1234",
    "serviceName": "AasCustomer",
    "sourceIpAddress": "42.120.75.137",
    "userAgent": "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_13_6) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/67.0.3396.99 Safari/537.36",
    "userIdentity": {
        "accessKeyId": "25***************",
        "accountId": "1234",
        "principalId": "1234",
        "type": "root-account",
        "userName": "root"
    }
}
```
• Compliance audits. This feature allows you to store the operations logs of Alibaba Cloud services that are generated in recent six months to meet the compliance auditing requirements.

• Ease of configurations. You can specify the settings with ease to enable this feature.

• Real-time analysis. Log Service provides the real-time analysis capacity and out-of-the-box statistical charts, which allow you to learn about the distribution and details of operations logs.

• Real-time alerts. You can enable the monitoring and alerting feature for the specified metrics to quickly respond to the changes that occur to your important businesses.

Supported Alibaba Cloud services

OSS

Release status

The Inner-ActionTrail feature is currently in the public preview phase. If you want to join the public preview, contact your sales manager or submit a ticket.

Prerequisites and limits

To use the Inner-ActionTrail feature, you must meet the following prerequisites:

You have enabled the Alibaba Cloud Log Service and your account has no overdue payments.

When you use the Inner-ActionTrail feature, you must note the following limits. Operations logs can only be shipped to a single dedicated Logstore, and then the Logstore is limited in use.

• You cannot write data to the Logstore by using API operations or SDKs.

• The built-in charts of the Logstore may be updated.

Note:

• Except these limits, you can use the Logstore as you use normal Logstores. All other features of Logstores are supported, such as search, statistics, alerting, and stream consumption.

• After Inner-ActionTrail is enabled, the storage period of the Logstore is reset to 180 days. You can modify the storage period on the Logstore modification page.
Scenarios

- Track operations logs of Alibaba Cloud services to find the reasons for asset changes.
- View, audit, and evaluate operations logs in near real time.
- Export logs to on-premises data centers.

4.12.2 Activate Inner-ActionTrail

This topic describes how to activate the Inner-ActionTrail feature in the Log Service console for log analysis and auditing.

Procedure

1. Log on to the Log Service console.

2. In the Import Data section, click the Platform Operation Log (Inner-ActionTrail) icon.

Note:

- You can also log on to the ActionTrail console and then click Turn on Inner-ActionTrail on the Trail List page to open the Inner-ActionTrail activation page.

- If you cannot find the Turn on Inner-ActionTrail button in the ActionTrail console, or the Platform Operation Log (Inner-ActionTrail) icon in the Log Service console, contact your sales manager or submit a ticket to join the public preview.

3. Select a Logstore.

   Select a project and Logstore in the target region.
4. Authorize ActionTrail to ship platform operations logs.

Authorize ActionTrail to ship platform operations logs to your Log Service project. If you have enabled Inner-ActionTrail to ship logs to another Log Service project, you must disable Inner-ActionTrail before you can proceed to the next step.

Note: Platform operations logs can be shipped to only one Logstore.

5. Enable Inner-ActionTrail.

Click Next to enable Inner-ActionTrail.

6. On the Configure Query and Analysis page, check the configuration data.

7. Click Next.

What's next

After you activate the Inner-ActionTrail feature, you can view logs and reports in the Log Service console.
Click the Logstore project that you selected. In the left-side navigation pane, click the search icon of the Logstore to query logs.

In the left-side navigation pane, click Dashboard and then click Inner-ActionTrail Center to view reports.

Note:
To stop shipping platform operations logs to your Logstore, log on to the ActionTrail console, and then click Turn off Inner-ActionTrail in the Trail List. After log shipping stops, the newly generated platform operations logs are no longer shipped to the Logstore. The automatic deletion of shipped logs are based on the Logstore configuration when they expire.
4.12.3 Inner-ActionTrail fields

This topic describes the fields of the Inner-ActionTrail feature.

The following table describes the fields used for the Inner-ActionTrail feature.

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EventID</td>
<td>The unique ID of a log event.</td>
</tr>
<tr>
<td>EventVersion</td>
<td>The version of an event. The current version is 1.0.0.</td>
</tr>
<tr>
<td>EventLocation</td>
<td>The location where an operation is performed by an Alibaba Cloud employee or system. ISO 3166-1 alpha 2 country codes are used.</td>
</tr>
<tr>
<td>EventProduct</td>
<td>The name of the product in which an operation is performed, for example, OSS.</td>
</tr>
<tr>
<td>EventName</td>
<td>The name of the event corresponding to an API operation used by a cloud product, for example, Set Bucket Quota Limit.</td>
</tr>
<tr>
<td>EventDescription</td>
<td>The reason for which an operation is performed. The reason may include the ticket ID, internal O&amp;M or change ticket ID, and security scanning ID.</td>
</tr>
<tr>
<td>Field</td>
<td>Description</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>EventType</td>
<td>The type of an operation.</td>
</tr>
<tr>
<td></td>
<td>• CUSTOMER_INITIATED_SUPPORT</td>
</tr>
<tr>
<td></td>
<td>Technical support initiated by Alibaba Cloud O&amp;M personnel, for example, troubleshooting based on tickets submitted by customers.</td>
</tr>
<tr>
<td></td>
<td>• ALIYUN_INITIATED_SERVICE</td>
</tr>
<tr>
<td></td>
<td>Services initiated by Alibaba Cloud employees or systems according to O&amp;M requirements, for example, bucket migration across clusters after the cluster hardware is out of warranty.</td>
</tr>
<tr>
<td></td>
<td>• ALIYUN_INITIATED_PENALTY</td>
</tr>
<tr>
<td></td>
<td>Operations performed on public user data by Alibaba Cloud employees or systems according to applicable laws and regulations.</td>
</tr>
<tr>
<td>EventMethod</td>
<td>The method by which an operation is performed. Methods can be common read/write operations, read/write operations through internal interfaces, and other operations such as backup and recovery. For example, Internal Write.</td>
</tr>
<tr>
<td>ResourceType</td>
<td>The resource type of an event, for example, Acs::Oss::Bucket.</td>
</tr>
<tr>
<td>ResourceID</td>
<td>The ID of a resource, for example, $bucketID.</td>
</tr>
<tr>
<td>ResourceRegionID</td>
<td>The ID of the region to which an event resource belongs. This field only applies to products that belong to a region, for example, cn-hangzhou.</td>
</tr>
<tr>
<td>ResourceOwnerID</td>
<td>The ID of the account to which a resource belongs.</td>
</tr>
<tr>
<td>EventAdditionalDetail</td>
<td>The additional information about an event.</td>
</tr>
<tr>
<td>EventTime</td>
<td>The time (in UTC) when an operation is performed, for example, 2019-09-18T05:23:37Z.</td>
</tr>
</tbody>
</table>
### Field Description

**EventLevel**
- **NOTICE**: When a NOTICE-level operation is performed, it is recorded in the corresponding log.
- **WARNING**: When a WARNING-level operation is performed, it is recorded in a log and the customer is alerted.

### 4.13 RDS SQL execution logs

#### 4.13.1 Overview

Log Service is integrated with ApsaraDB Relational Database Service (RDS) for MySQL to analyze SQL execution logs in a quasi-real-time manner. It provides drill-down reports, custom audit rules, and alerts upon exceptions. You can also subscribe to reports and use Log Service together with other data solutions.

**Background**

A database is the core of an enterprise's business. You need to record and audit database operations, especially all SQL execution operations. In addition, database SQL execution logs provide information such as performance, execution results, and classification, which play a critical role in troubleshooting and online performance optimization.

Log Service is integrated with RDS to analyze SQL execution logs in a quasi-real-time manner.

**Complete SQL execution logs**

SQL execution logs record all SQL operations performed on a database. RDS collects SQL execution logs through network listening, which has little impact on actual execution performance. SQL execution logs include but are not limited to the following types of SQL operations and information:

- **Database logon and logoff**
- **Data definition language (DDL)**: SQL statements that define the database structure, such as CREATE, ALTER DROP, TRUNCATE, and COMMENT
- **Data manipulation language (DML)**: SQL operation statements, such as SELECT, INSERT, UPDATE, and DELETE
• Other SQL operations performed after SQL statements are run, such as rollback and control
• Failed SQL operations
• SQL execution latency, execution results, and number of affected rows

Benefits

Sending real-time SQL execution logs from RDS to Log Service brings the following benefits:

• Simple: simplifies configuration and allows you to easily collect SQL execution logs and import them to Log Service in quasi-real time.
• Reliable: has no impact on the performance of existing databases. Data is imported in quasi-real time with a latency of several minutes.
• Comprehensive: records not only executed SQL statements but also database logon, failed data execution, latency, and impact results. This allows you to master database execution status, performance, and potential security issues.
• Powerful: provides real-time log analysis, an out-of-the-box report center, and custom configuration.
• Flexible: supports quasi-real-time monitoring, alerting, and report subscription based on specific metrics to ensure timely response to critical business exceptions. Log Service can collaborate with other data solutions such as stream computing, cloud storage, and visualization to dig up more data value. Log Service also provides one-stop services, including machine learning, custom reports, and data processing.
• Cost-effective: charges only about USD 0.028/day to store 10 million SQL execution logs for one month.

Limits

• You must activate Alibaba Cloud Log Service to use the SQL audit and analysis features.
• You must activate Alibaba Cloud RDS and create an RDS for MySQL instance that supports the #unique_198 or #unique_199 feature.
• You must enable the #unique_198 feature of the RDS standard edition before enabling log data import in the Log Service console. If you enable the SQL audit feature of the trial edition, you may fail to enable log data import.
The following table describes the RDS instance type for which logs can be automatically sent to Log Service.

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>MySQL</td>
<td>Other types are not supported.</td>
</tr>
<tr>
<td>Region</td>
<td>China (Beijing), China (Shanghai), China (Hangzhou), China (Shenzhen), China (Hong Kong), and China (Qingdao)</td>
<td>Other regions are not available.</td>
</tr>
<tr>
<td>Version</td>
<td>5.5, 5.6, 5.7, and 8.0</td>
<td>V8.0 is a beta version.</td>
</tr>
<tr>
<td>Edition</td>
<td>High-availability edition and finance edition</td>
<td>N/A</td>
</tr>
<tr>
<td>Storage type</td>
<td>Local SSD</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Currently, RDS can send logs only to Log Service projects in the same region. For example, logs of an RDS instance in China (Beijing) can only be sent to a Log Service Logstore in China (Beijing).

Do not modify the default dashboard configuration. Default dashboards are automatically updated with the upgrade of features.

Target customers

- Compliance: large enterprises and institutions that have higher compliance requirements for storage of SQL execution logs, such as financial companies and government agencies.
- Advanced O&M: enterprises, such as Internet service enterprises and financial companies, that need to use the real-time interactive statistics, query analysis, and alerting features of Log Service to monitor database status in real time and quickly troubleshoot performance, reliability, and stability exceptions in business.
- Security: companies, such as financial, securities, e-commerce, and games companies, that need to use the real-time query and analysis features of Log Service to audit database security and assist in data security protection and post-event troubleshooting.
• Behavior analysis: companies that need to analyze database operations, use the statistics and query results as reference to further optimize and verify performance, features, and experience, and connect to external systems for further in-depth analysis.

Scenarios

• Interactive analysis of SQL execution logs

You can troubleshoot database access exceptions and analyze problems in quasi-real time, view read/write latency, and understand the regional distribution of access clients.

• Security analysis

You can view abnormal deletion of important tables, SQL attacks, and SQL injection on dashboards. You can also customize SQL execution audit and alerting rules based on log query and analysis results for timely processing. Log Service supports multiple alerting modes, such as DingTalk and SMS, and provides custom alert content templates.

• Overall access monitoring

You can view the overall access monitoring status and O&M reliability metrics on dashboards, and use the metrics to build your own dashboard.

• Operational analysis

You can view active databases and tables, and the addition, modification, and deletion of key data.

4.13.2 Procedure

This topic describes how to import real-time SQL execution logs from ApsaraDB for RDS to Log Service for real-time analysis.

Prerequisites

• Alibaba Cloud Log Service is activated.

• RDS is activated and an RDS instance that supports #unique_198 or #unique_199 is created. The RDS instance can be ApsaraDB RDS for MySQL instance, ApsaraDB RDS for PostgreSQL instance, or ApsaraDB RDS for SQL Server instance.

• The #unique_198 feature of the RDS standard edition is enabled for log data import. If you have enabled the SQL audit feature of the trial edition, disable the trial...
edition or upgrade it to the standard edition. Otherwise, you cannot enable log data import in the Log Service console.

- Your account has permission to access the SQL audit feature. You can use only your Alibaba Cloud account to enable and use the SQL audit feature by default. If you use a Resource Access Management (RAM) user to enable and use the SQL audit feature, you must grant relevant Log Service permissions to the RAM user. To grant relevant permissions on RDS to the RAM user, use the following policy:

```
Action:
  rds:ModifySQLCollectorPolicy
  rds:DisableSqlLogDistribution
  rds:EnableSqlLogDistribution
  rds:DescribeSqlLogInstances

resource:
  acs:rds:*::*:dbinstance/*
```

Procedure

1. Log on to the Log Service console.

2. Create a project in the same region as the RDS instance. For more information, see #unique_114.

3. Create a Logstore. For more information, see #unique_27.

4. Expand the Logstore and click the plus sign (+) next to Data Import.

   You can also click Import Data in the upper-right corner of the Overview page, and then select a Logstore in the Import Data wizard.
5. Select a data type. In the Import Data dialog box, select RDS SQL Audit - Cloud Products.

6. Select the Logstore.

   If you start the collection configuration process by clicking the plus sign (+) next to Data Import under the Logstore, the system skips this step.

7. Authorize Log Service to distribute logs.

   If you have not authorized Log Service to distribute logs, click Authorize next to RAM. If you have authorized Log Service to distribute logs, skip this step.

8. Turn on the import switch for the RDS instance.

   Note:
   If no RDS instance appears, it means that the instance does not meet the relevant requirements. For more information, see Limits.
9. Configure log query and analysis.
   An index is created by default. To modify the index, click the Logstore and choose Index Attributes > Modify in the upper-right corner.

4.13.3 Log fields

The following table lists the fields in Relational Database Service (RDS) SQL execution logs.

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>topic</strong></td>
<td>The topic of the log. The value is rds_audit_log.</td>
<td>rds_audit_log</td>
</tr>
<tr>
<td>instance_id</td>
<td>The ID of the RDS instance.</td>
<td>rm-abcdefg</td>
</tr>
<tr>
<td>check_rows</td>
<td>The number of scanned rows.</td>
<td>0</td>
</tr>
<tr>
<td>db</td>
<td>The name of the database.</td>
<td>my_db</td>
</tr>
<tr>
<td>fail</td>
<td>Indicates whether the SQL statement failed to be run.</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Valid values:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• 0: successful</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• 1: failed</td>
<td></td>
</tr>
<tr>
<td>client_ip</td>
<td>The IP address of the client that accessed the RDS instance.</td>
<td>10.20.30.40</td>
</tr>
<tr>
<td>latency</td>
<td>The latency of SQL execution, in microseconds.</td>
<td>15</td>
</tr>
<tr>
<td>origin_time</td>
<td>The duration for which the SQL statement was run, in microseconds.</td>
<td>1525671499014609</td>
</tr>
<tr>
<td>return_rows</td>
<td>The number of rows that the SQL statement returned.</td>
<td>10</td>
</tr>
<tr>
<td>sql</td>
<td>The SQL statement that was run.</td>
<td>SELECT * from tb_abc</td>
</tr>
<tr>
<td>thread_id</td>
<td>The ID of the thread.</td>
<td>17171330</td>
</tr>
</tbody>
</table>
### 4.14 NAS access logs

#### 4.14.1 Overview

Log Service allows you to query and analyze Network Attached Storage (NAS) access logs. It also provides out-of-the-box reports and alerts so that you can monitor the system in real time.

Alibaba Cloud NAS is a file storage service for compute nodes, such as Elastic Compute Service (ECS) instances, Elastic High Performance Computing (E-HPC) clusters, and Container Service clusters. NAS complies with standard file access protocols. Without modifying existing applications, you can have a distributed file system that features unlimited capacity and performance scaling, a single namespace, shared access, high reliability, and high availability.

Each request to NAS generates an access log that records the details of the access request, such as the operation type, target object, and response status of the current user.

Log Service allows you to query and analyze NAS access logs. It also provides out-of-the-box reports and alerts so that you can monitor the system in real time. By querying and analyzing NAS access logs, you can collect access events, audit sensitive operations, and diagnose problems in specific scenarios. You can also manage data kanbans and view access data in real time through default dashboards, and configure alert tasks in various scenarios to monitor access status in real time.

**Benefits**

- Simple configuration: collects real-time logs with simple configuration. For more information about how to collect NAS access logs and relevant log fields, see *Procedure*.  

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>user</td>
<td>The username used to run the SQL statement.</td>
<td>spotuserw</td>
</tr>
<tr>
<td>update_rows</td>
<td>The number of rows updated.</td>
<td>0</td>
</tr>
</tbody>
</table>
· Real-time analysis: provides real-time log analysis and an out-of-the-box report center. This allows you to master all operations on important cloud assets and explore details in real time.

· Real-time alerts: supports quasi-real-time monitoring and alerting based on custom operations to ensure timely response to critical business exceptions.

· Collaboration: collaborates with other data solutions such as stream computing, cloud storage, and visualization to dig up more data value.

· Free default configuration: allows you to prolong the storage duration as needed for compliance, tracing, and filing. Log data is stored for seven days by default and can be permanently stored. If log data is stored more than seven days, the minimum storage cost is about USD 0.049/GB/month. For more information about pricing, see #unique_31.

Limits and instructions

· Dedicated Logstore

A dedicated Logstore is created to store NAS access logs. Other data cannot be written to the dedicated Logstore. The query, statistics, alerting, and streaming consumption features of this Logstore are not limited.

· Network File System (NFS) only

Currently, only NFS NAS access logs can be analyzed. Other protocol types will be supported in the future.

· Pay-as-you-go

The NAS log collection feature is currently in the test phase and is free of charge by default. Your NAS access logs are stored in Log Service for seven days. If you store logs for a longer period, you are charged for the extra storage duration. Log Service supports the pay-as-you-go billing method. For more information about pricing, see #unique_31.

Scenarios

· View the read and write operations of each NAS volume

On the summary dashboard, you can view the number of active volumes, the total amount of read and write traffic, and the number and distribution of creation, deletion, read, and write operations on each volume.
• View NAS access details

On the details dashboard, you can view the distribution of operations within different periods, the details of successful or failed operations, and the causes of failure, for example:

- The trend of data read and write flow, that is, the IP address that writes data, the volume that receives data, and the IP address that reads data.
- The number and proportion of files accessed on each volume.
- The queries per second (QPS) for each volume.
- The total read and write traffic and average traffic of each volume per minute.
- The IP addresses of clients whose read and write traffic ranks top N.
- The IP addresses of clients whose operation errors rank top N. This metric is used to identify the source of operation errors.
- The number of read, write, creation, and deletion operations on hotspot data.
- The inodes of frequently accessed files.
- The number of abnormal operations, such as authentication failure and operation failure.
- The AuthRc and NFSProtocolRc fields that indicate the distribution of operation states.

• Audit sensitive operations

On the audit dashboard, you can view your sensitive operations on NAS, such as creating or deleting directories or files, and the files that are most frequently read and written.

• Search for access logs

- Run the following statement to search for logs of failed authentication attempts to identify operations with permission exceptions:

  AuthRc > 0

- Run the following statement to search for logs of failed operations to check incorrect operation results:

  NFSProtocolRc > 0

4.14.2 Procedure

Network Attached Storage (NAS) access logs can be sent to Log Service in real time. Each access log records the details of an access event, such as the operation type,
target object, and response status of the current user. This topic describes how to collect NAS access logs. It also describes relevant log fields.

Prerequisites

1. Alibaba Cloud Log Service and NAS are activated.
2. A file system and a mount point are created. The file system is mounted to an Elastic Compute Service (ECS) instance.

   For more information, see Create a file system, Add a mount target, and Mount a file system.

Procedure

1. Log on to the NAS console with your Alibaba Cloud account.
2. Click Apply to apply for the user-level monitoring feature as prompted at the top of the console.

   User-level monitoring is now available for NAS so that you can monitor multiple I/O metrics of NAS files in real time. You must apply for this feature to configure these metrics.
3. In the left-side navigation pane, choose Log Analysis > Log Management to go to the Log Management page.
4. Set Region.
5. Authorize NAS to access your cloud resources.

   If you configure NAS log collection for the first time, click Auth now. After authorization, NAS can send NAS access logs to your Logstore.
6. Go back to the Log Management page.
7. Click Create Log Store in the upper-right corner.
8. Set Region and File System ID/Name, and then click OK.

After you complete the preceding configuration, Log Service creates a project and a Logstore for you. Your NAS access logs are sent to this default Logstore of Log Service in real time. You can view the project and Logstore of Log Service corresponding to your NAS file system on the Log Management page. You can also:

   • Click to go to the Log Service console to view logs.
   • Click stop to stop dumping logs to Log Service.
More actions

On the Log Management page, find the target file system and click Click to go.

On the query and analysis page of the Log Service console, you can perform the following operations:

- Query and analyze logs
  
  Log Service supports query syntax and analysis syntax for log query in complex scenarios. For more information, see query syntax and analysis syntax.

- View dashboards
  
  Log Service provides three dashboards for NAS to display real-time status such as event types and event source distribution.

  You can also modify the dedicated dashboards, create a custom dashboard, and add custom analysis charts in various scenarios to your dashboards. For more information about dashboards, see #unique_190.

- Configure alerts
  
  On the query and analysis page of the target Logstore, you can save current query conditions as a saved search and an alert to regularly monitor important log data and send alert notifications when exceptions occur. For more information, see #unique_189.

Default configuration

Table 4-13: Default configuration

<table>
<thead>
<tr>
<th>Default configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project</td>
<td>A project is created by default.</td>
</tr>
<tr>
<td></td>
<td>- <strong>Chinese environment</strong>: nas-AliUid-cn-RegionId</td>
</tr>
<tr>
<td></td>
<td>- <strong>English environment</strong>: nas-AliUid-en-RegionId</td>
</tr>
<tr>
<td>Logstore</td>
<td>The <code>nas-nfs</code> Logstore is created by default. All logs collected by the NAS log collection feature are stored in this Logstore.</td>
</tr>
</tbody>
</table>

**Note:**
Currently, only NFS NAS access logs can be analyzed. Other protocol types will be supported in the future.
### Default configuration item

<table>
<thead>
<tr>
<th>Default configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Region</strong></td>
<td>By default, logs are stored in the region where the current file system resides.</td>
</tr>
<tr>
<td><strong>Shard</strong></td>
<td>By default, two shards are created and the Automatic Sharding feature is enabled.</td>
</tr>
<tr>
<td><strong>Log storage duration</strong></td>
<td>By default, logs are stored for seven days. Logs stored more than seven days are automatically deleted. You can modify the storage duration as needed. If you store logs for a longer period, you are charged for the extra storage duration. For more information about pricing, see #unique_31.</td>
</tr>
<tr>
<td><strong>Dashboard</strong></td>
<td>By default, the following dashboards are created:</td>
</tr>
<tr>
<td></td>
<td>• nas-nfs-nas_audit_dashboard</td>
</tr>
<tr>
<td></td>
<td>• nas-nfs-nas_detail_dashboard</td>
</tr>
<tr>
<td></td>
<td>• nas-nfs-nas_summary_dashboard</td>
</tr>
</tbody>
</table>

### Log fields

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>ArgIno</td>
<td>The inode number of the file system.</td>
<td>226</td>
</tr>
<tr>
<td>AuthRc</td>
<td>The authorization return code.</td>
<td>0</td>
</tr>
<tr>
<td>NFSProtocolRc</td>
<td>The NFS return code.</td>
<td>0</td>
</tr>
<tr>
<td>OpList</td>
<td>The NFSv4 procedure number.</td>
<td>null</td>
</tr>
<tr>
<td>Proc</td>
<td>The NFSv3 procedure number.</td>
<td>1</td>
</tr>
<tr>
<td>RWSize</td>
<td>The size of data that was read and write, in bytes.</td>
<td>-1</td>
</tr>
<tr>
<td>RequestId</td>
<td>The ID of the request.</td>
<td>5ACF5CD506EAC7A508F056DF</td>
</tr>
<tr>
<td>ResIno</td>
<td>The inode number of lookup resources.</td>
<td>null</td>
</tr>
<tr>
<td>SourceIp</td>
<td>The IP address of the client.</td>
<td>127.0.0.1</td>
</tr>
</tbody>
</table>
### Field Description Example

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>User</td>
<td>The ID of the user.</td>
<td>123456789</td>
</tr>
<tr>
<td>Vers</td>
<td>The version number of the NFS protocol.</td>
<td>3</td>
</tr>
<tr>
<td>Vip</td>
<td>The IP address of the server.</td>
<td>172.18.158.178</td>
</tr>
<tr>
<td>Volume</td>
<td>The ID of the file system.</td>
<td>2d2794a330</td>
</tr>
<tr>
<td>microtime</td>
<td>The time when the request was sent, in microseconds.</td>
<td>1523539157201995</td>
</tr>
</tbody>
</table>

### 4.15 Cloud Firewall logs

#### 4.15.1 Overview

The Log Analysis service of Cloud Firewall provides internet traffic logs and real-time log analysis.

The Log Analysis service of Cloud Firewall can automatically collect and store real-time log of both inbound and outbound traffic. It outputs query analysis, reports, alarms, and downstream computing interconnection and provide you with detailed analysis result.

![Cloud Firewall Log Analysis](image)

### Benefits

The Log Analysis service of Cloud Firewall has the following benefits:

- **Classified Protection compliance**: Log Analysis provides log storage duration of six months to help your website meet the requirements of classified protection compliance.
- Easy configuration: Easy configuration allows you to collect Internet traffic logs in real time.
- Real-time analysis: Integrated with the Simple Log Service (SLS), the Log Analysis service provides the real-time log analysis service and report center. With the help of log analysis, you can view all the traffic and user's visits going through Cloud Firewall.
- Real-time alarms: Log Analysis supports you to customize real-time monitoring and alerts based on specific indicators. This ensures you receive real-time alerts when there is any threats detected in the critical business.

Prerequisites

Before you begin to use the service of Log Analysis, the following prerequisite must be available:

You have purchased and activated the Log Analysis service of Cloud Firewall (Log Analysis is available in Pro, Enterprise, and Flagship editions). For details, refer to Enable the log analysis service.

Restrictions

The logstore of Cloud Firewall is an exclusive logstore with the following restrictions:

- You cannot write data into logstore with APIs or SDKs, or modify the attributes of the logstore (such as the storage cycle).

  Note:
  Other general logstore features (such as query, statistics, alarms, and stream consumption) are supported, and there is no difference with the general logstore.

- Alibaba Cloud's Log Service (SLS) does not charge for the exclusive logstore of Cloud Firewall, but SLS itself must be available (not overdue).
- Built-in reports provided by Log Analysis of Cloud Firewall may be updated and upgraded automatically.

Scenarios

- Track Internet traffic logs to trace security threats.
- Allow you to view Internet request activities in real time, and check the security status and trend of your assets.
- Provide you with quick understanding of security operation efficiency and handling the risks in a timely manner.
- Output logs to your self-built data and computing centers.

4.15.2 Log analysis billing method

Cloud Firewall Log Analysis service charges fees based on the selected log storage duration and log storage capacity. Log Analysis is charged by monthly and annual subscription.

Enable Activate Log Service and select the log storage period and the log storage size. Then, the price is automatically calculated based on the log store specification of your choice.

Log storage specification

Different log storage specifications of Cloud Firewall are charged as follows:

<table>
<thead>
<tr>
<th>Log storage duration</th>
<th>Log storage size</th>
<th>Applicable bandwidth</th>
<th>Recommended version</th>
<th>Monthly subscription fee</th>
<th>Annual subscription fee</th>
</tr>
</thead>
<tbody>
<tr>
<td>180 days</td>
<td>1TB</td>
<td>Applicable to business scenarios with monthly bandwidth not higher than 10 Mbps</td>
<td>Pro Edition</td>
<td>To be released.</td>
<td>To be released.</td>
</tr>
<tr>
<td>5TB</td>
<td></td>
<td>Applicable to business scenarios with monthly bandwidth not higher than 50 Mbps</td>
<td>Enterprise Edition</td>
<td>To be released.</td>
<td>To be released.</td>
</tr>
</tbody>
</table>
Log Service

Data Collection / 4 Cloud product collection

<table>
<thead>
<tr>
<th>Log storage duration</th>
<th>Log storage size</th>
<th>Applicable bandwidth</th>
<th>Recommended version</th>
<th>Monthly subscription fee</th>
<th>Annual subscription fee</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20TB</td>
<td>Applicable to business scenarios with monthly bandwidth not higher than Mbps</td>
<td>Flagship Edition</td>
<td>To be released.</td>
<td>To be released.</td>
</tr>
</tbody>
</table>

Note:
To increase the bandwidth, we recommend that you expand with 1TB log storage for every 10 Mbps increase.

Upgrade storage capacity

If you have no log storage left, a notification appears to remind you to expand the storage size. You can click Upgrade on the console to expand the storage size.

Notice:
If you fail to upgrade the log storage capacity when the storage capacity is full, Cloud Firewall will stop writing new log data to the exclusive logstore of log analysis service, the stored log data in the logstore is retained. Log data is deleted automatically if it is stored for more than 180 days or is not renewed after the Log Analysis service expires for 7 days. Once the log data is deleted, it cannot be recovered.

Duration

The purchase duration of Cloud Firewall log service is bound to the subscription instance of the Cloud Firewall you purchased.
- Buy: When you buy a Cloud Firewall subscription and enable Log Analysis, the price of Log Service is calculated based on the validity of the subscription.
- Upgrade: When you enable Log Service by upgrading an existing Cloud Firewall subscription, the price of Log Service is calculated based on the log storage size.

Service expiration

If the purchased Cloud Firewall instance is about to expire, Log Analysis service will also expire.

- When the service expires, Cloud Firewall stops writing log entries to the exclusive logstore in Log Service.
- The log entries recorded by Cloud Firewall Log Analysis are retained within seven days after the service expires. If you renew the service within seven days after the service expires, you can continue to use Log Analysis service. Otherwise, all stored log entries are deleted.

4.15.3 Enable the log analysis service

After you activate Cloud Firewall, you can enable the Log Analysis service in the Cloud Firewall console. Log Analysis provides you with the real-time log search and analysis features.

Features

After the Log Analysis service is activated, real-time logs of the internet traffic through Cloud Firewall can be collected automatically. You can also perform real-time log search and analysis with Log Analysis service, and check the results in log dashboards. You need to set the storage duration and storage capacity when you enable the Cloud Firewall log analysis service.

Note:
The log analysis service is available in the Cloud Firewall Pro, Enterprise, and Flagship editions.

Enable the Log Analysis service of Cloud Firewall

1. Log on to the Cloud Firewall console.
2. In the left-side navigation pane, select Advanced Features > Log Analysis.
3. Click Activate Now on the Log Analysis page.

![Log Analysis](image)

4. Select your log storage capacity, and then click Pay to complete the payment.

   **Note:**
   For more information about log analysis pricing, refer to *Log analysis billing method*.

5. Go back to Log Analysis page in Cloud Firewall console.

6. Click the Status switch on the right side to enable the Log Analysis service.

   ![Log Analysis](image)

Log Analysis service retrieves records of both inbound and outbound Internet traffic flowing through Cloud Firewall. You can use the retrieved records to detect threats in real time.

### 4.15.4 Collect the log

You can enable the log collector function for Cloud Firewall in the Cloud Firewall console.

**Prerequisites**

- You have activated Cloud Firewall.
- You have activated Alibaba Cloud Log Service.

**Context**

The log collector function retrieves log data of inbound and outbound Internet traffic for Alibaba Cloud Firewall in real time. The retrieved log data can be searched and analyzed in real time, and the returned results are displayed in dashboards. Based on the log data, you can analyze visits to and attacks on your websites and help the security engineers develop protection strategies.

After you enable the Cloud Firewall log analysis function, the log analysis function automatically creates a dedicated Logstore named cloudfirewall-logstore under your account. Cloud Firewall automatically imports log entries to this dedicated Logstore in real time. For more information about the default configuration of the dedicated Logstore, see *Default configuration*.

**Procedure**

1. In the left-side navigation pane, locate Log Analysis.
2. Click the Status switch on the right side to enable the log collector function.

Table 4-14: Default log analysis configuration

<table>
<thead>
<tr>
<th>Default configuration item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project</td>
<td>The log analysis project created by Cloud Firewall. The project name is determined according to the region of your Cloud Firewall instance.</td>
</tr>
<tr>
<td></td>
<td>• If the Cloud Firewall instance is deployed in a Mainland China region, the project name is: <code>cloudfirewall-project-Alibaba Cloud account ID-cn-hangzhou</code>.</td>
</tr>
<tr>
<td></td>
<td>• If the Cloud Firewall instance is deployed in the Finance Cloud (Hangzhou) region, the project name is: <code>cloudfirewall-project-Alibaba Cloud account ID-cn-hangzhou-finance</code>.</td>
</tr>
<tr>
<td></td>
<td>• If the Cloud Firewall instance is deployed in other regions, the project name is: <code>cloudfirewall-project-Alibaba Cloud account ID-ap-southeast-1</code>.</td>
</tr>
<tr>
<td>Logstore</td>
<td>The default Logstore is <code>cloudfirewall-logstore</code>. All log data retrieved by Cloud Firewall is stored in this Logstore.</td>
</tr>
<tr>
<td>Region</td>
<td>• If the Cloud Firewall instance is deployed in a Mainland China region, the project is saved in the China (Hangzhou) region by default.</td>
</tr>
<tr>
<td></td>
<td>• If the Cloud Firewall instance is deployed in other regions, the project is saved in the Singapore region by default.</td>
</tr>
<tr>
<td>Shard</td>
<td>By default, two shards are created and the Automatic shard splitting function is enabled.</td>
</tr>
</tbody>
</table>
### Default configuration item | Description
---|---
Dashboards | A dashboard is created by default.

**Note:**
The default log analysis configuration items cannot be modified.

### Restrictions and guidelines

- After you enable the Log Analysis function, the system automatically creates a Logstore named cloudfirewall-logstore in the Log Service console. The Logstore is dedicated to Cloud Firewall and stores all log entries of Cloud Firewall. Do not delete this Logstore.

- Other data cannot be written into the dedicated Logstore.

  Log entries generated by Cloud Firewall are stored in the dedicated Logstore. You cannot write other data into this Logstore by using the API, SDK, or other methods.

**Note:**
The dedicated Logstore has no restrictions in search, statistics, alerts, streaming consumption, and other functions.

- Basic configurations, such as the log storage period, cannot be modified.

- The dedicated Logstore is not billed.

To use the dedicated Logstore, you must activate Log Service for your account.

**Note:**
When your Log Service is overdue, the Cloud Firewall log collector function is suspended until you pay the bills.

- Do not delete or modify the configurations of the default project, Logstore, index, and dashboards created by Log Service. Log Service will update the Cloud Firewall log analysis function. The index of the dedicated Logstore and the default report are also updated.

- If you want to use the Cloud Firewall log analysis function with a RAM user account, you must grant the required Log Service permissions to the RAM
user account. For more information, see Authorize RAM user accounts with Log Analysis function.

4.15.5 Log analysis

Cloud Firewall console supports the Log Analysis function.

Overview

After you enable the Log Analysis function in Cloud Firewall console, you can perform real-time log search and analysis, view or edit dashboards, and set up monitoring and alerts on the Log Analysis page.

Procedure

1. Log on to the Cloud Firewall console.
2. In the left-side navigation pane, select Logs > Log Analysis.
3. Click the Status switch on the right side to enable the Log Analysis function.
4. Enter a search and analysis statement, select a time range, and click Search & Analysis.

More actions

On the Log Analysis page, you can perform the following actions to handle the returned search results:
- **Customize search and analysis**

  The log analysis function provides the search and analysis statements for you to search and analyze log entries in different scenarios. For more information, see **Customize search and analysis**.

- **View the distribution of log entries by time**

  The histogram under the search box shows the distribution of log entries that are filtered by time and search statement. The horizontal axis indicates the time period, and the vertical axis indicates the number of log entries. The total number of the log entries returned is also displayed.

  **Note:**

  You can drag the mouse pointer in the histogram to narrow down the time period. The **time picker** automatically updates the time period, and the search results are also updated accordingly.

- **View raw logs**

  On the Raw Logs tab page, each log entry is detailed on an individual page, which includes the time when the log is generated, the content, and the columns in the log entry. You can click Display Content Column to set the display mode for the long strings in the content column. The display modes include Full Line and New Line. You can click Column Settings to customize the columns to be displayed, or click the Download icon to download the search results.

  Additionally, you can click a value or a property name in the content column to add a search condition to the search box. For example, if you click `log_service`
in the `_source_`: `log_service` field, the following search statement is added to the search box:

"Former Search Statement" and source: `log_service`

- **View analysis graphs**

  The log analysis function enables you to show the analysis results in graphs. You can select the graph type as needed on the Graph tab page. For more information, see *Analysis graphs*.

- **Quick analysis**

  The quick analysis function on the Raw Logs tab provides you with a quick interactive search function. You can view the distribution of a property within a specific time period. This function can reduce the time used for indexing key data. For more information, see *Quick analysis*. 
Customize search and analysis

The log analysis function provides the search and analysis statements. Separate the search and analysis statements with a vertical bar (|):

$Search | $Analytics

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search</td>
<td>A keyword, a fuzzy string, a numerical value, or a range can be used as a search condition. You can also combine these search conditions. If the statement is empty or only contains a wildcard character (*), all log entries are searched.</td>
</tr>
<tr>
<td>Analytics</td>
<td>Performs calculation and statistics to the search results or all log entries.</td>
</tr>
</tbody>
</table>

Note:
Both the search and the analysis statements are optional.

- When the search statement is empty, all log entries within the specified time period are displayed. Then, the search results are used for statistics.
- When the analysis statement is empty, the search results are returned. No statistical analysis is performed.

Search statements

The search statements of Log Service support full text search and field search. You can set the New Line mode, syntax highlighting, and other functions in the search box.

- Full text search

You can enter keywords without specifying fields to perform the search. You can enter a keyword enclosed in quotation marks ("), you can enter a keyword enclosed in quotation marks ("), you can enter a keyword enclosed in quotation marks ("), you can enter a keyword enclosed in quotation marks ("), you can enter a keyword enclosed in quotation marks ("), you can enter a keyword enclosed in quotation marks ("), you can enter a keyword enclosed in quotation marks ("") to query log entries that
contain the entire keyword. You can also use spaces or and to separate multiple keywords.

Examples

- Search by keyword

  The following statements can be used to search for log entries that contain www.aliyun.com and error.

  www.aliyun.com error or www.aliyun.com and error.

- Search by condition

  The following statement can be used to search for log entries that contain www.aliyun.com, error, or 404.

  www.aliyun.com and (error or 404)

- Search by prefix

  The following statement can be used to search for log entries that contain www.aliyun.com and start with failed_.

  www.aliyun.com and failed_*

  • Search by field

  To narrow down the search results, you can search by field.

  You can specify numeric fields. The format is field name: value or field name >= value. Moreover, you can use both the and and or operators in full text search.

  • Note:

  The wildcard character (*) can only be added as a suffix. The wildcard character (**) cannot be added as a prefix. For example, the statement cannot be *_error.
The Cloud Firewall log analysis function supports searching by field. For more information about the definition, type, format, and other information of each field, see *Cloud Firewall log field descriptions*.

**Examples**

- **Search by specifying multiple fields**
  
  If you want to search for log entries about client 1.2.3.4 accessing IP address 1.1.1.1, set the following search conditions:

  ```
  src_ip: 1.2.3.4 and dst_ip: 1.1.1.1
  ```

  **Note:**
  
  In this example, the `src_ip` field and `dst_ip` field are log fields created by Cloud Firewall.

- **Search by specifying numeric fields**
  
  The following statement can be used to search log entries where the response time exceeds five seconds.

  ```
  request_time_msec > 5000
  ```

  Searching by time period is also supported. For example, you can search for log entries where the response time exceeds five seconds and is no greater than ten seconds.

  ```
  request_time_msec in (5000 10000]
  ```
You can get the same result by using the following search statement:

```
request_time_msec > 5000 and request_time_msec <= 10000
```

- **Field search**

You can search whether a field exists as follows:

- Search for log entries that include the `total_pps` field.
  
  `total_pps:*`

- Search for log entries that include the `ua_browser` field.
  
  `not total_pps:*`

For more information about the search statements supported by Log Service, see *Indexes and search*.

**Analysis statements**

You can use the SQL/92 statements for log analysis and statistics.

For more information about the statements and functions supported by Log Service, see *Real-time analysis*.

**Note:**

- The `from table name part (the from log part) in the standard SQL statements can be omitted.```
- The first 100 log entries are returned by default. You can modify the number of the returned log entries by using the `LIMIT statement`.```

**Examples of search and analysis**

**Time-based log search and analysis**

Each Could Firewall log entry has a `time` field, which is used to indicate the time. The format of field is `year-month-dayThour:minute:second+time zone`. For example, in `2018-05-31T20:11:58+08:00`, the `time zone` is `UTC+8`.

Meanwhile, each log has a built-in field `__time__`. This field also indicates the time when the log entry is generated. The field is used for calculation during the time-based statistics process. The format of this field is `Unix timestamp`, and the value of this field indicates the amount of seconds that have elapsed since 00:00:00.
Coordinated Universal Time (UTC), January 1, 1970. Therefore, if you want to display a calculated result, you must convert the format first.

- Select and display the time
- Calculate the time
- Statistical analysis by group based on a specific time

**Note:**
You can also display the results with a line graph.

The `date_parse` and `date_format` functions are used to convert the time format. For more information about the functions that can be used to parse the time field, see **Date and time functions**.

4.15.6 Fields in the log entry

Cloud firewall records the inbound and outbound traffic logs, including multiple log fields. You can perform query and analysis based on specific fields.

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Description</th>
<th>Example</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>__time__</code></td>
<td>Time of the operation in Cloud Firewall</td>
<td>2018-02-27 11:58:15</td>
<td>-</td>
</tr>
<tr>
<td>Field Name</td>
<td>Description</td>
<td>Example</td>
<td>Comments</td>
</tr>
<tr>
<td>------------</td>
<td>---------------------------------</td>
<td>-----------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>topic</strong></td>
<td>Log topic</td>
<td>cloudfirew all_access_log</td>
<td>Log topic is unique, which is cloudfirew for Cloud Firewall.</td>
</tr>
<tr>
<td>Log_type</td>
<td>Log types</td>
<td>Internet_log</td>
<td>Internet_log refers to the Internet Traffic Log.</td>
</tr>
<tr>
<td>aluid</td>
<td>User's Alibaba Cloud UID</td>
<td>12333333333333</td>
<td>-</td>
</tr>
<tr>
<td>app_name</td>
<td>Protocol of the access traffic</td>
<td>HTTPS</td>
<td>Possible values include HTTPS, NTP, SIP, SMB, NFS, and DNS. Unknown values are Unknown.</td>
</tr>
</tbody>
</table>
| direction  | Traffic direction                | in                    | • in: traffic goes to the ENI  
• out: traffic goes from the ENI  |
<p>| domain     | Domain name                      | <a href="http://www.aliyun.com">www.aliyun.com</a>         | -                                                                        |
| dst_ip     | Destination IP                   | 1.1.1.1                | -                                                                        |
| dst_port   | Destination port                 | 443                   | -                                                                        |
| end_time   | Session end time                 | 1555399260            | Unit: Seconds (Unix timestamp)                                          |
| In_bps     | Bps of inbound traffic           | 11428                 | Unit: bps                                                               |
| In_packet_bytes | Total number of bytes of inbound traffic | 2857                  | -                                                                        |
| In_packet_count | Total number of packet of inbound traffic | 18                    | -                                                                        |
| In_pps     | Pps of inbound traffic           | 9                     | Unit: pps                                                               |</p>
<table>
<thead>
<tr>
<th>Field Name</th>
<th>Description</th>
<th>Example</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ip_protocol</td>
<td>IP protocol type</td>
<td>TCP</td>
<td>Protocol name . TCP and UDP protocol are supported.</td>
</tr>
<tr>
<td>Out_bps</td>
<td>Bps of outbound traffic</td>
<td>27488</td>
<td>Unit: bps</td>
</tr>
<tr>
<td>Out_packet_bytes</td>
<td>Total number of bytes of outbound traffic</td>
<td>6872</td>
<td>-</td>
</tr>
<tr>
<td>Out_packet_count</td>
<td>Total number of packet of outbound traffic</td>
<td>15</td>
<td>-</td>
</tr>
<tr>
<td>Out_pps</td>
<td>Pps of outbound traffic</td>
<td>7</td>
<td>Unit: pps</td>
</tr>
<tr>
<td>region_id</td>
<td>Region to which the access traffic belongs</td>
<td>cn-beijing</td>
<td>-</td>
</tr>
<tr>
<td>Rule_result</td>
<td>Result of matching with the rules</td>
<td>pass23</td>
<td>The result of matching with the rules. The values are:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>· Pass: The traffic is allowed to pass through Cloud Firewall.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>· Alert: Cloud Firewall detects threats in the traffic.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>· Discard: The traffic is not allowed to pass through Cloud Firewall.</td>
</tr>
<tr>
<td>src_ip</td>
<td>Source IP</td>
<td>1.1.1.1</td>
<td>-</td>
</tr>
<tr>
<td>src_port</td>
<td>The port of the host from which traffic data is sent</td>
<td>47915</td>
<td>-</td>
</tr>
</tbody>
</table>
Log Analysis of Cloud Firewall provides you with Advanced Settings. You can set advanced features for Log Service with Advanced Settings. For example, you can set alarms and notifications, real-time log collection and consumption, shipping log data, or provide visual representations with other products.

Steps

1. Log on to Cloud firewall console.
2. Go to the left-side navigation pane Advanced Functions > Log Analysis.
3. Click Advanced Settings in the upper-right corner.
4. In the dialog box that appears, click Go to open the Log Service console.
5. In the Log Service console, you can set the following advanced features for log projects and logstores:
   • Alarms and notifications
   • Real-time log collection and consumption
   • Shipping log data to other Alibaba Cloud storage services in real time
   • Providing visual representations with other products

4.15.8 Export log entries

The Log Analysis function of Cloud Firewall allows you to export log entries to your local device.

You can export log entries on the current page to a CSV file, or export all log entries to a TXT file.

Procedure

1. Log on to the Cloud Firewall console.
2. In the left-side navigation pane, select Advanced Features > Log Analysis.
3. On the Raw Logs tab page, click the Download icon on the right side.

Note: The download icon does not appear if there's no search result.

4. In the Log Download dialog box, select Download Log in Current Page or Download all logs by the CLI console.
   • Download logs in current page:
     Click OK to export the raw log entries on the current page to a CSV file.
   • Download all logs by CLI:
     a. For more information about installing the CLI, see CLI guide.
     b. Click Security Information Management Link to view and record the AccessKey ID and AccessKey Secret of the current user.
     c. Click Copy Command and paste the command into CLI, replace the
        AccessID obtained in step 2 and AccessKey Secret obtained in step 2
with the AccessKey ID and AccessKey Secret of the current user, and run the command.

After you run the command, all raw log entries created by Cloud Firewall are automatically exported and saved to the file `download_data.txt`.

4.15.9 Authorize RAM user accounts with Log Analysis function

If you want to use Cloud Firewall's Log Analysis function with a RAM user account, you must first use your Alibaba Cloud account to authorize this RAM user account with the Log Analysis functions of Cloud Firewall.

**Context**

The following permissions are required for enabling and using Cloud Firewall's Log Analysis function.

<table>
<thead>
<tr>
<th>Operations</th>
<th>Required account</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enable the Log Analysis function. You only need to perform this operation once.</td>
<td>Alibaba Cloud account</td>
</tr>
</tbody>
</table>

| Authorize Cloud Firewall to write the log data into the dedicated Logstore of Log Analysis in real time. You only need to perform this operation once. | · Alibaba Cloud account  
· A RAM user account with `AliyunLogFullAccess` permission  
· A RAM user account with the customized permission of log writing |
<table>
<thead>
<tr>
<th>Operations</th>
<th>Required account</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use the Log Analysis function.</td>
<td>• Alibaba Cloud account</td>
</tr>
<tr>
<td></td>
<td>• A RAM user account with <strong>AliyunLogFullAccess</strong> permission</td>
</tr>
<tr>
<td></td>
<td>• A RAM user account with the customized permissions</td>
</tr>
</tbody>
</table>

You can grant permissions to a RAM user account as needed.

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Grant a RAM user account permissions</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grant a RAM user account full permission to Log Service.</td>
<td>The <strong>AliyunLogFullAccess</strong> policy specifies full permission to Log Service.</td>
<td>For more information, see <a href="#">RAM user management</a>.</td>
</tr>
<tr>
<td>After you use your Alibaba Cloud account to enable the Cloud Firewall log analysis function and complete the authorization, grant the RAM user account the permission to view logs.</td>
<td>The <strong>AliyunLogReadOnlyAccess</strong> policy specifies the read-only permission.</td>
<td>For more information, see <a href="#">RAM user management</a>.</td>
</tr>
<tr>
<td>Grant the RAM user account the permission s to enable and use the Cloud Firewall log analysis function. Do not grant other permissions to Log Service.</td>
<td>Create a custom authorization policy, and apply the policy to the RAM user account.</td>
<td>For more information, see the following procedure.</td>
</tr>
</tbody>
</table>

Procedure

1. Log on to the *RAM console*.

2. Open the Create Custom Policy tab page on the Policies page.

3. In the upper-right corner of the page, click Create Authorization Policy.

4. Click Blank Template, enter the Policy Name and the following Policy Content into this template.

   - **Note:**
Replace `${Project}` and `${Logstore}` in the following policy with the Log Service Project name and Logstore name dedicated for Cloud Firewall, respectively.

```json
{
  "Version": "1",
  "Statement": [
    {
      "Action": "log:GetProject",
      "Resource": "acs:log:***:project/${Project}\",
      "Effect": "Allow"
    },
    {
      "Action": "log:CreateProject",
      "Resource": "acs:log:***:project/**",
      "Effect": "Allow"
    },
    {
      "Action": "log:ListLogStores",
      "Resource": "acs:log:***:project/${Project}/logstore/**",
      "Effect": "Allow"
    },
    {
      "Action": "log:CreateLogStore",
      "Resource": "acs:log:***:project/${Project}/logstore/**",
      "Effect": "Allow"
    },
    {
      "Action": "log:GetIndex",
      "Resource": "acs:log:***:project/${Project}/logstore/${Logstore}\",
      "Effect": "Allow"
    },
    {
      "Action": "log:CreateIndex",
      "Resource": "acs:log:***:project/${Project}/logstore/${Logstore}\",
      "Effect": "Allow"
    },
    {
      "Action": "log:UpdateIndex",
      "Resource": "acs:log:***:project/${Project}/logstore/${Logstore}\",
      "Effect": "Allow"
    },
    {
      "Action": "log:CreateDashboard",
      "Resource": "acs:log:***:project/${Project}/dashboard/**",
      "Effect": "Allow"
    },
    {
      "Action": "log:UpdateDashboard",
      "Resource": "acs:log:***:project/${Project}/dashboard/**",
      "Effect": "Allow"
    },
    {
      "Action": "log:CreateSavedSearch",
      "Resource": "acs:log:***:project/${Project}/savedsearch/**",
      "Effect": "Allow"
    }
  ]
}```
5. Click Create Authorization Policy.

6. Go to the Users page, locate the RAM user account, and click Authorize.

7. Select the custom authorization policy that you created, and then click OK. The authorized RAM user account then can enable and use the Log Analysis function. However, this RAM user account is not authorized to use other functions of Log Service.
4.15.10 Manage log storage

After you enable the Log Analysis function of Cloud Firewall, the log storage space is allocated based on your specified log storage size. You can view the usage of the log storage space on the Log Analysis page in the Cloud Firewall console.

Check the log storage

You can view the log storage in the Cloud Firewall console at any time.

Note:
The log storage information in the console is not updated in real time. It takes up to two hours to update the actual storage information to the console. We recommend that you expand the log storage space before it is exhausted.

1. Log on to the Cloud Firewall console.
2. In the left-side navigation pane, select Advanced Features > Log Analysis.
3. View the log storage information in the upper-right corner of the Log Analysis page.

Expand log storage

To expand the log storage, click Upgrade Storage at the top of the Log Analysis page.

Note:
We recommend that you expand the log storage space before it is exhausted. If no storage space is available, then the new log data cannot be written into the dedicated Logstore.
Clear log storage

You can delete all log entries stored in the Logstore. For example, you can delete the log entries generated during the testing phase and use the log storage space to store log entries that are generated during the production phase only.

Click Clear at the top of the Log Analysis page, and confirm to delete all stored log entries.

Notice:
You cannot recover the deleted log entries. This operation is irreversible.

Note:
You only have limited times for clearing the log storage.

4.16 SMS logs

Short Message Service (SMS) log analysis is a solution jointly launched by Alibaba Cloud SMS and Log Service. This solution provides powerful statistical analysis, data query, and real-time SMS log analysis.

The SMS log analysis feature allows you to collect all logs sent by SMS in near real time. Based on this feature, you can monitor SMS details and statuses. Log Service outputs query analysis results, reports, and alerts, and delivers logs to downstream log consumers such as OSS and MaxCompute. This allows you to focus on analysis and eliminate the need for time-consuming data queries and sorting.

Scenarios

The SMS log analysis feature allows you to analyze the exceptions that occur when you send SMS messages and monitor the status of each SMS message.

- You can analyze and query the statuses of sent messages, or analyze the number and success rate of sent messages based on different regions or SMS templates.
- You can locate and query SMS sending records based on the template content, SMS batch number, and signature.
• You can customize SQL statements to query the statuses of sent SMS messages and export large amounts of log data.

Best practices

Use SMS in operational activities.

• For example, you have obtained a batch of mobile phone numbers, some of which are no longer in use. Sending text messages to these invalid mobile phone numbers is a waste of operational costs. You can use the SMS log analysis feature to find these invalid mobile phone numbers and reduce costs.

• When you send a marketing SMS message, you need to analyze the marketing effect from multiple aspects based on different regions and service providers (SPs). You can use the built-in template of the SMS log analysis feature to view the marketing SMS message distribution by region or SP. If the built-in template cannot meet your requirements, you can use SQL statements to analyze the sent data.

Documentation

• Enable the log analysis service
• Fields in the log entry
• Export log entries
• Authorize RAM user accounts with Log Analysis function
• Advanced Settings
• #unique_138
5 Other collection methods

5.1 Web Tracking

Log Service supports collecting logs from HTML, H5, iOS, and Android platforms by using Web Tracking, and customizing dimensions and metrics.

As shown in the preceding figure, you can collect user information from various browsers, iOS apps, and Android apps (apart from iOS/Android SDK) by using Web Tracking. For example:

- Browsers, operating systems, and resolutions used by users.
- Browsing behaviors of users, such as the clicking behaviors and purchasing behaviors on the website.
- The staying time in the app for users and whether the users are active or not.

Note:
Using Web Tracking means that this Logstore enables the anonymous write permission of the Internet, and dirty data may be generated.

Precautions

- Using Web Tracking means that this Logstore enables the anonymous write permission of the Internet without valid authentication, and dirty data may be generated.
- Only Get requests are supported. A request body exceeding 16 KB cannot be uploaded.
Procedure

Step 1 Enable Web Tracking

You can enable Web Tracking in the console or by using Java SDK.

- Enable Web Tracking in the console

  1. On the Logstore List page, click Modify at the right of the Logstore that must enable the Web Tracking function.
  2. Turn on the Web Tracking switch.

- Enable Web Tracking by using Java SDK

  **Log Service Java SDK:**

  ```java
  import com.aliyun.openservices.log.Client;
  import com.aliyun.openservices.log.common.LogStore;
  import com.aliyun.openservices.log.exception.LogException;
  public class WebTracking {
      static private String accessId = "your accesskey id";
      static private String accessKey = "your accesskey";
      static private String project = "your project";
      static private String host = "log service data address";
      static private String logStore = "your logstore";
      static private Client client = new Client(host, accessId, accessKey);
      public static void main(String[] args) {
          try {
              //Enable the Web Tracking function on the created Logstore
              LogStore logSt = client.GetLogStore(project, logStore).GetLogStore();
              client.UpdateLogStore(project, new LogStore(logStore, logSt.GetTtl(), logSt.GetShardCount(), true));
              //Disable the Web Tracking function.
              //client.UpdateLogStore(project, new LogStore(logStore, logSt.GetTtl(), logSt.GetShardCount(), false));
              //Create a Logstore that supports the Web Tracking function.
          } catch (LogException e) {
              e.printStackTrace();
          }
      }
  }
  ```
Step 2 Collect logs

After the Web Tracking function is enabled for the Logstore, you can use any of the following methods to upload data to the Logstore.

**Note:**
We recommend that you use the SDK to upload logs.

- **Use the JS SDK**

  1. **Copy `loghub-tracking.js` to the web directory, and introduce the following script on the page:**

     Click to download.

     ```html
     <script type="text/javascript" src="loghub-tracking.js" async></script>
     ```

     **Note:**
     To keep page loading running, the script sends HTTP requests asynchronously. If data must be sent several times in the page loading process, the subsequent request overwrites the preceding HTTP request, and the browser shows the tracking request exits. Sending requests synchronously can help to avoid this problem. To send requests synchronously, replace the statement in the script.

     **Original script:**

     ```javascript
     this.httpRequest_.open("GET", url, true)
     ```

     **Replace the last parameter to send requests synchronously:**

     ```javascript
     this.httpRequest_.open("GET", url, false)
     ```

  2. **Create a Tracker object.**

     ```javascript
     var logger = new window.Tracker('${host}','${project}','${logstore }');
     logger.push('customer', 'zhangsan');
     logger.push('product', 'iphone 6s');
     ```
logger.push('price', 5500);
logger.logger();
logger.push('customer', 'lisi');
logger.push('product', 'ipod');
logger.push('price', 3000);
logger.logger();

The parameter meaning are as follows:

<table>
<thead>
<tr>
<th>Field</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>${host}</td>
<td>The domain name of the region where your Log Service is located.</td>
</tr>
<tr>
<td>${project}</td>
<td>The name of the project created in Log Service.</td>
</tr>
<tr>
<td>${logstore}</td>
<td>The name of the Logstore with the Web Tracking function enabled under ${project}.</td>
</tr>
</tbody>
</table>

After running the preceding commands, you can see the following two logs in Log Service:

customer:zhangsan
product:iphone 6s
price:5500

customer:lisi
product:ipod
price:3000

• Use HTTP GET request

curl --request GET 'http://${project}.${host}/logstores/${logstore}/track? APIVersion=0.6.0&key1=val1&key2=val2'

The parameter meanings are as follows.

<table>
<thead>
<tr>
<th>Field</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>${project}</td>
<td>The name of the project created in Log Service.</td>
</tr>
<tr>
<td>${host}</td>
<td>The domain name of the region where your Log Service is located.</td>
</tr>
<tr>
<td>${logstore}</td>
<td>The name of the Logstore with the Web Tracking function enabled under ${project}.</td>
</tr>
<tr>
<td>APIVersion=0.6.0</td>
<td>The reserved field, which is required.</td>
</tr>
<tr>
<td>Field</td>
<td>Definition</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>topic</strong>=yourtopic</td>
<td>Specify the log topic, reserved fields (optional).</td>
</tr>
<tr>
<td>key1=val1,key2=val2</td>
<td>The key-value pairs to be uploaded to Log Service. Multiple key-value pairs are supported, but you must make sure that the URL length is less than 16 KB.</td>
</tr>
</tbody>
</table>

- Use the HTML IMG tag

```html
<img src='http://${project}.${host}/logstores/${logstore}/track.gif?APIVersion=0.6.0&key1=val1&key2=val2'/>
<img src='http://${project}.${host}/logstores/${logstore}/track_ua.gif?APIVersion=0.6.0&key1=val1&key2=val2'/>
```

The parameter meanings are the same as those in Use HTTP GET request. In addition to uploading custom parameters, track_ua.gif transmits UserAgent and referer of in the HTTP header as log fields on the server.

Note:

To collect referer of the HTTPS page, the link of the preceding Web Tracking must be the HTTPS type.

After data is uploaded to Log Service, you can use LogSearch/Analytics of Log Service to search and analyze log data in real time, and display real-time analysis results with various visualization solutions. You can also consume data by using Consumer Library provided by Log Service.

### 5.2 Use DataWorks to export MaxCompute data to Log Service

**Scenario**

DataWorks is the data relay service of Alibaba Cloud. DataWorks can ship log files that Log Service collects to MaxCompute for storing and analyzing these log files. MaxCompute provides offline computing. If you require online analytical processing (OLAP), you can use DataWorks to export the log files that have been shipped to MaxCompute and the computing result to Log Service. Log Service then performs a real-time search and analysis of the exported data.
Implementation

LogHub Writer obtains the data that is generated by Reader from the DataWorks framework, and transforms the data types that are supported by DataWorks to the string type. When the data volume reaches the specified `batchSize`, LogHub Writer uses the Log Service Java SDK to transfer all the data to Log Service at a time. By default, LogHub Writer transfers 1,024 entries at a time. The value of `batchSize` is up to 4096.

Prerequisites

1. You have activated Log Service and created the project and Logstore.
2. You have activated MaxCompute and created tables.
3. You have activated DataWorks.

Procedure

1. Log on to the DataWorks console and create a LogHub data source.

   For more information about how to create a data source, see `#unique_221/unique_221_Connect_42_section_nkh_hnf_vdb`.
2. Create a synchronization task in script mode.
   a. Click Sync Tasks in the left-side navigation pane, and click Script Mode to configure the synchronization task.

Figure 5-1: Script Mode
b. Specify the parameters in the import template.

Figure 5-2: Import template

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source type</td>
<td>Select ODPS as the type of your data source.</td>
</tr>
<tr>
<td>data sources</td>
<td>The name of your data source. You can also click New Source to create a data source.</td>
</tr>
<tr>
<td>Type of objective</td>
<td>Select LogHub as the type of the shipping destination.</td>
</tr>
<tr>
<td>data sources</td>
<td>The name of the shipping destination. Select the LogHub data destination created in step 1, or click New Source to create a data destination.</td>
</tr>
</tbody>
</table>

Then, click confirmation to configure the synchronization task.

c. Enter your configuration.

The example is as follows:

```json
{
    "type": "job",
```


```
"version": "1.0",
"configuration": {
"setting": {
"errorLimit": {
"record": "0"
},
"speed": {
"mbps": "1",
"concurrent": 1,
"dmu": 1,
"throttle": false
}
},
"reader": {
"plugin": "odps",
"parameter": {
"accessKey":"*****",
"accessId":"*****",
"column": ["*"]
},
"writer": {
"plugin": "loghub",
"parameter": {
"endpoint": "",
"accessId": "",
"accessKey": "",
"project": "",
"logstore": "",
"batchSize": "1024",
"topic": "",
"time": "time_str",
"timeFormat": "%Y_%m_%d %H:%i:%S",
"column": [
"col0",
"col1",
"col2",
"col3",
"col4",
"col5"
],
"datasource": "sls"
}
}
```

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Required</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>endpoint</td>
<td>Yes</td>
<td>The endpoint of Log Service. For more information, see #unique_26.</td>
</tr>
<tr>
<td>Parameter</td>
<td>Required</td>
<td>Description</td>
</tr>
<tr>
<td>---------------</td>
<td>----------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>accessKeyId</td>
<td>Yes</td>
<td>The AccessKeyId of your Alibaba Cloud account or RAM user.</td>
</tr>
<tr>
<td>accessKeySecret</td>
<td>Yes</td>
<td>The AccessKeyId of your Alibaba Cloud account or RAM user.</td>
</tr>
<tr>
<td>project</td>
<td>Yes</td>
<td>The name of the destination project in Log Service.</td>
</tr>
<tr>
<td>logstore</td>
<td>Yes</td>
<td>The name of the destination Logstore in Log Service.</td>
</tr>
<tr>
<td>topic</td>
<td>No</td>
<td>The field in MaxCompute that you specify as the topic field in Log Service. It is an empty string by default.</td>
</tr>
<tr>
<td>batchSize</td>
<td>No</td>
<td>The number of entries that LogHub Writer transfers at a time. It is 1024 by default.</td>
</tr>
<tr>
<td>column</td>
<td>Yes</td>
<td>The column name in each entry.</td>
</tr>
<tr>
<td>time</td>
<td>No</td>
<td>The name of the time field.</td>
</tr>
</tbody>
</table>

**Note:**
- The columns that are not specified in the column parameter are dirty data.
- If the time field is not specified, the system time is used as the log time by default.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Required</th>
<th>Description</th>
</tr>
</thead>
</table>
| timeFormat  | If the time field is specified, timeFormat is required. | You can set timeFormat to the following format:  
  - bigint: unix timestamp.  
  - timestamp: time retrieved from the string, such as `%Y_%m_%d %H:%M:%S`.  
  
  If the time field is 1529382552 in the bigint type, the timeFormat field is bigint. If the time field is 2018_06_19 12:30:25 in the string type, the timeFormat field is `%Y_%m_%d %H:%M:%S`. |
| datasource  | Yes      | The data type that is defined in DataWorks. |
3. Save and run this task.

Click Save and specify the path to save this synchronization task. You can also run this task directly, or submit it to the scheduling system.

Figure 5-3: Run the synchronization task

- Run the task.
  
  Click Run to directly start synchronizing all the data.

- Schedule the task.
  
  Click Submit to submit the task to the scheduling system. Then, the scheduling system automatically runs this task according to your configuration.

Note:
We recommend that you set the scheduling cycle the same as the partition generation cycle. For example, if the partition is generated based on hourly collected data, the scheduling cycle is one hour.

For more information about scheduling the task, see *Ship data to MaxCompute via DataWorks*.

Data types

After you import MaxCompute data to Log Service using DataWorks, all data types are converted to the string type, as shown in the following table.
### MaxCompute data type

<table>
<thead>
<tr>
<th>Data type imported to LogHub</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long</td>
</tr>
<tr>
<td>Double</td>
</tr>
<tr>
<td>String</td>
</tr>
<tr>
<td>Data</td>
</tr>
<tr>
<td>Boolean</td>
</tr>
<tr>
<td>Bytes</td>
</tr>
</tbody>
</table>

### 5.3 Kafka protocol

In addition to the Logtail, SDK, and API, Log Service also allows you to write data into Log Service in compliance with the Kafka protocol. You can use the Kafka Producer SDK in various languages and collection agents that can export the collected data to Kafka.

#### Limits

- The supported Kafka protocol versions are from Kafka 0.8.0 to Kafka 2.1.1.
- You must use the SASL_SSL connection protocol for secure data transmission.
- If your Logstore contains multiple shards, you need to write data in load balancing mode.
- Currently, you can use only the producer or agent to write data into Log Service in compliance with the Kafka protocol.
### Configuration

If you use the Kafka protocol to collect data, you must set some parameters. The following table describes the required parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connection protocol</td>
<td>The connection protocol for secure data transmission. You must use SASL_SSL.</td>
<td>SASL_SSL</td>
</tr>
<tr>
<td>hosts</td>
<td>The cluster address for the initial connection. The port number for an intranet (either a classic network or VPC) address is 10011. The port number for an Internet address is 10012. You need to select the service endpoint where your target project is located. For more information, see #unique_223.</td>
<td>cn-hangzhou-intranet.log.aliyuncs.com:10011, cn-hangzhou.log.aliyuncs.com:10012</td>
</tr>
<tr>
<td>topic</td>
<td>The mapped Logstore name in Log Service. You must create a Logstore in advance.</td>
<td>test-logstore-1</td>
</tr>
<tr>
<td>username</td>
<td>The mapped project name in Log Service.</td>
<td>&lt;yourusername&gt;</td>
</tr>
<tr>
<td>Parameter</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>-------------</td>
<td>------------------------------------------------------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>password</td>
<td>The information about your AccessKey, which is in the format of ${access-key-id}#$${access-key-secret}. You need to replace ${access-key-id} with your AccessKey ID and ${access-key-secret} with your AccessKey Secret. We recommend that you use the AccessKey of a RAM user. For more information, see Grant a RAM user the permission to access Log Service.</td>
<td>&lt;yourpassword&gt;</td>
</tr>
<tr>
<td>Certificate</td>
<td>The directory of the certificate. Each domain name in Log Service has a CA certificate. You only need to use the default root certificate.</td>
<td>/etc/ssl/certs/ca-bundle.crt</td>
</tr>
</tbody>
</table>

Error codes

If you fail to collect log data in compliance with the Kafka protocol, the system returns a Kafka error code for the specific cause of failure. For more information about Kafka error codes, see the error list. The following table describes the specific error codes, description, and corresponding solutions.

<table>
<thead>
<tr>
<th>Error code</th>
<th>Description</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>NetworkException</td>
<td>The error message returned because a network error has occurred.</td>
<td>Wait for 1 second and try again.</td>
</tr>
<tr>
<td>TopicAuthorizationException</td>
<td>The error message returned because the authentication fails. Generally, your AccessKey</td>
<td>Enter a valid AccessKey and ensure that it has</td>
</tr>
<tr>
<td>Error code</td>
<td>Description</td>
<td>Solution</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>is invalid or has no</td>
<td>permission to write data into the corresponding project or Logstore.</td>
<td>the required write permission.</td>
</tr>
<tr>
<td>UnknownTopicOrPartitionException</td>
<td>The error message returned because either of the following errors has</td>
<td>1. Create a project and a Logstore in advance.</td>
</tr>
<tr>
<td></td>
<td>occurred:</td>
<td>2. Ensure that the region where the project is located is the same as</td>
</tr>
<tr>
<td></td>
<td>• The corresponding project or Logstore does not exist.</td>
<td>the region indicated by the endpoint that you entered.</td>
</tr>
<tr>
<td></td>
<td>• The region where the project is located is different from the region</td>
<td></td>
</tr>
<tr>
<td></td>
<td>indicated by the endpoint that you entered.</td>
<td></td>
</tr>
<tr>
<td>KafkaStorageException</td>
<td>The error message returned because a server error has occurred.</td>
<td>Wait for 1 second and try again.</td>
</tr>
</tbody>
</table>

Example

You want to write data into Log Service. The project in Log Service is named `test-project-1` and the Logstore is named `test-logstore-1`. The region where the project is located is `cn-hangzhou`. The AccessKey ID of the RAM user with the corresponding write permission is `<yourAccessKeyId>`, and the AccessKey Secret is `<yourAccessKeySecret>`. 
Example 1: Use Beats software to write data into Log Service

You can export the collected data to Kafka by using Beats software such as Metricbeat, Packetbeat, Winlogbeat, Auditbeat, Filebeat, and Heartbeat. For more information, see Configure the Kafka output. The sample code is as follows:

```yaml
output.kafka:
  # initial brokers for reading cluster metadata
  hosts: ["cn-hangzhou.log.aliyuncs.com:10012"]
  username: "<yourusername>"
  password: "<yourpassword>"
  ssl.certificateAuthorities:
  # message topic selection + partitioning
  topic: 'test-logstore-1'
  partition.round_robin:
    reachable_only: false

  required_aacks: 1
  compression: gzip
  max_message_bytes: 1000000
```

By default, the Beats software exports JSON-formatted logs to Kafka. You can also create a JSON type index for the content field. For more information, see #unique_224. The following figure shows a log sample.
Log Service

Data Collection / 5 Other collection methods

- Example 2: Use Collectd to write data into Log Service

Collectd is a daemon used to collect the performance metrics of a system or application on a regular basis. You can also use Collectd to export the collected data to Kafka. For more information, see the Write Kafka plug-in.

If you want to export the collected data from Collectd to Kafka, you need to install the Write Kafka plug-in and relevant dependencies. In the CentOS, you can directly run the `sudo yum install collectd-write_kafka` command to install the plug-in. For more information about the Red-Hat Package Manager (RPM) resources, see RPM resource collectd-write_kafka.

The sample code is as follows:

```xml
<Plugin write_kafka>
  Property "metadata.broker.list" "cn-hangzhou.log.aliyuncs.com:10012"
  Property "security.protocol" "sasl_ssl"
  Property "sasl.mechanism" "PLAIN"
  Property "sasl.username" "<yourusername>"
  Property "sasl.password" "<yourpassword>"
  Property "broker.address.family" "v4"
  <Topic "test-logstore-1">
    Format JSON
    Key "content"
  </Topic>
</Plugin>
```
In the preceding sample code, the format of the data exported to Kafka is set to JSON. In addition to JSON, Collectd also supports the Command and Graphite formats. For more information, see the Collectd configuration documentation.

If you use the JSON format, you can create a JSON type index for the content field. For more information, see **JSON type**. The following figure shows a log sample.

```
03-25 21:31:14 __source__ : rdkafka
    __tag__ : receive_time : 1553520674
    __topic__ : test
    ▼ content : {} ▼
       ▼ values : []
           0 : 25088000
       ▼ dstypes : []
           0 : "gauge"
       ▼ dsnames : []
           0 : "value"
           time : 1553520674.125
           interval : 10
           host : "[blank]"
           plugin : "memory"
           plugin_instance : ""
           type : "memory"
           type_instance : "slab_unrec"　
```

- **Example 3: Use Telegraf to write data into Log Service**

  **Telegraf** is a sub-project of InfluxData. It is the agent compiled in Go for collecting, processing, and aggregating metrics. It is designed to use less memory resources. Telegraf can be used to build services and collect the metrics of a third-party component through plug-ins. In addition, Telegraf has the integration feature. It can obtain metrics from the system where it runs, obtain metrics through a third-party API, and even monitor metrics through StatsD and Kafka consumer services.

  Telegraf can export data to Kafka. Therefore, you only need to modify the configuration file to use Telegraf to collect data and write data into Log Service. The sample code is as follows:

  ```
  [[outputs.kafka]]
  ## URLs of kafka brokers
  brokers = ["cn-hangzhou.log.aliyuncs.com:10012"]
  ## Kafka topic for producer messages
  ```
topic = "test-logstore-1"
routing_key = "content"
## CompressionCodec represents the various compression codecs recognized by
## Kafka in messages.
## 0 : No compression
## 1 : Gzip compression
## 2 : Snappy compression
## 3 : LZ4 compression
compression_codec = 1
## Optional TLS Config
tls_ca = "/etc/ssl/certs/ca-bundle.crt"
# tls_cert = "/etc/telegraf/cert.pem" # tls_key = "/etc/telegraf/
key.pem"
## Use TLS but skip chain & host verification
# insecure_skip_verify = false
## Optional SASL Config
sasl_username = "<yourusername>"
sasl_password = "<yourpassword>"
## Data format to output.
## https://github.com/influxdata/telegraf/blob/master/docs/
DATA_FORMATS_OUTPUT.md
data_format = "json"

Note:
You must set a valid `tls_ca` directory for Telegraf. You can use the default root certificate. The typical root certificate directory in a Linux environment is `/etc/ssl/certs/ca-bundle.crt`.

In the preceding sample code, the format of the data exported to Kafka is set to JSON. In addition to JSON, Telegraf also supports other formats such as Graphite and Carbon2. For more information, see Telegraf output data formats.

If you use the JSON format, you can create a JSON type index for the content field. For more information, see JSON type. The following figure shows a log sample.

- Example 4: Use Fluentd to write data into Log Service

   Fluentd is an open-source data collector that provides a unified logging layer. It allows you to collect data in a uniform manner so that you can easily use
and understand data. Fluentd is a member project of Cloud Native Computing Foundation (CNCF). It complies with the Apache 2 License protocol.

Fluentd provides many input, processing, and output plug-ins. Specifically, the *Kafka plug-in* can help Fluentd export data to Kafka. You only need to *install* and configure this plug-in.

The sample code is as follows:

```xml
<match **>
  @type kafka  
  # Brokers: You can choose either brokers or zookeeper.  
  brokers      cn-hangzhou.log.aliyuncs.com:10012  
  default_topic test-logstore-1  
  default_message_key content  
  output_data_type json  
  output_include_tag true  
  output_include_time true  
  sasl_over_ssl true  
  username <yourusername>  
  password <yourpassword>  
  ssl_ca_certs_from_system true  
  # ruby-kafka producer options  
  max_send_retries 10000  
  required_acks 1  
  compression_codec gzip
</match>
```

In the preceding sample code, the format of the data exported to Kafka is set to *JSON*. In addition to JSON, Fluentd also supports more than 10 formats. For more information, see *Fluentd Formatter*.

If you use the JSON format, you can create a JSON type index for the content field. For more information, see *JSON type*. The following figure shows a log sample.
Example 5: Use Logstash to write data into Log Service

Logstash is an open-source engine for collecting data in real time. Using Logstash, you can dynamically collect data from different sources, process the data (for example, filter or convert the data), and export the result to a target address. You can analyze the data further based on the output result.

Logstash provides a built-in Kafka output plug-in. It allows you to directly enable Logstash to write data into Log Service. However, you must configure the SSL certificate and the SASL jass file because Log Service uses the SASL_SSL connection protocol in compliance with the Kafka protocol.

1. Create a jaas file, and then save it to a target directory, such as /etc/kafka/kafka_client_jaas.conf.

```java
KafkaClient {
    org.apache.kafka.common.security.plain.PlainLoginModule required
    username="<yourusername>"
    password="<yourpassword>";
};
```

2. Set the SSL certificate, and then save it to a target directory, such as /etc/kafka/client-root.truststore.jks.

Each domain name in Log Service has a CA certificate. You only need to download the GlobalSign Root CA, and save the Base64-encoded root certificate to a target directory, such as /etc/kafka/ca-root. Then, run a keytool command to generate a JKS file. When you generate a JKS file for the first time, you need to set a password.

```
keytool -keystore client.truststore.jks -alias root -import -file /etc/kafka/ca-root
```

3. Configure the Logstash. The sample code is as follows:

```ruby
input { stdin { } }
output {
    stdout { codec => rubydebug }
    kafka {
        topic_id => "test-logstore-1"
        bootstrap_servers => "cn-hangzhou.log.aliyuncs.com:10012"
        security_protocol => "SASL_SSL"
        ssl_truststore_location => "/etc/client-root.truststore.jks"
        ssl_truststore_password => "123456"
        jaas_path => "/etc/kafka_client_jaas.conf"
        sasl_mechanism => "PLAIN"
        codec => "json"
        client_id => "kafka-logstash"
    }
}
```
In the preceding sample code, the format of the data exported to Kafka is set to JSON. In addition to JSON, Logstash also supports more than 10 formats. For more information, see Logstash Codec plug-ins.

If you use the JSON format, you can create a JSON type index for the content field. For more information, see JSON type. The following figure shows a log sample.

5.4 Collect data by using Syslog

Log Service allows you to collect data by using Syslog rather than using third-party agents to forward collected data.

Limits

- Syslog messages must conform to RFC 5424.
- The maximum size of each log entry is 64 KB.
- To ensure secure data transmission, TCP-based Transport Layer Security (TLS) 1.2 must be used for data transmission.
Parameters

When collecting data by using Syslog, you must associate Syslog with the endpoint of Log Service. The port number of Syslog is 10009. You must use the service endpoint of your project. For more information about service endpoints, see #unique_223. Example: cn-hangzhou-intranet.log.aliyuncs.com:10009. In addition, you must specify a project, Logstore, and AccessKey pair for Log Service in the STRUCTURED-DATA field of each Syslog message.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>STRUCTURED-DATA name</td>
<td>Each STRUCTURED-DATA field must have a name and the name must be Logservice.</td>
<td>Logservice</td>
</tr>
<tr>
<td>Project</td>
<td>A project in Log Service. Create a project in Log Service before you collect data.</td>
<td>test-project-1</td>
</tr>
<tr>
<td>Logstore</td>
<td>A Logstore in Log Service. You must create a Logstore before you collect data.</td>
<td>test-logstore-1</td>
</tr>
<tr>
<td>access-key-id</td>
<td>Your AccessKey ID. We recommend that you use the AccessKey ID of a RAM user account. For more information, see Authorize a RAM user to access Log Service.</td>
<td>&lt;yourAccessKeyId&gt;</td>
</tr>
<tr>
<td>access-key-secret</td>
<td>Your AccessKey Secret. We recommend that you use the AccessKey Secret of a RAM user account. For more information, see Grant a RAM user the permission to access Log Service.</td>
<td>&lt;yourAccessKeySecret&gt;</td>
</tr>
</tbody>
</table>

Sample log

Log Service parses received Syslog messages. To ensure the security of the AccessKey pair, the Logservice field is deleted by default. The following table describes fields that are received by Log Service.
<table>
<thead>
<tr>
<th>Field name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>source</strong></td>
<td>The HOSTNAME field defined in Syslog.</td>
</tr>
<tr>
<td><strong>topic</strong></td>
<td>The value of this field is syslog-forwarder.</td>
</tr>
<tr>
<td><strong>facility</strong></td>
<td>The facility (device/module) information defined in Syslog.</td>
</tr>
<tr>
<td><strong>program</strong></td>
<td>The process name.</td>
</tr>
<tr>
<td><strong>severity</strong></td>
<td>The severity of the log entry.</td>
</tr>
<tr>
<td><strong>priority</strong></td>
<td>The priority of the log entry.</td>
</tr>
<tr>
<td><strong>unixtimestamp</strong></td>
<td>The timestamp of the log entry. Unit: nanosecond.</td>
</tr>
<tr>
<td>content</td>
<td>The MSG field defined in Syslog.</td>
</tr>
</tbody>
</table>

For more information, see [RFC 5424](https://tools.ietf.org/html/rfc5424).

Sample configuration

You want to collect Syslog messages from your host and write them into Log Service. The project in Log Service is named `test-project-1` and the Logstore is named `test-logstore-1`. The project resides in the China (Hangzhou) region. The
AccessKey ID of the RAM user account with the write permission is `<yourAccessKeyId>`, and the AccessKey Secret is `<yourAccessKeySecret>`.

- Example 1: Use Rsyslog to forward logs to Log Service

Rsyslog is installed on Linux hosts by default to process system logs. You can configure Rsyslog to forward system logs to Log Service. The configuration files of Rsyslog differ between versions. You can run the `man rsyslogd` command to view the Rsyslog version.

### Note:
You must ensure that you have installed the `rsyslog-gnutls` package. You can run the `sudo apt-get install rsyslog-gnutls` or `sudo yum install rsyslog-gnutls` command to install the `rsyslog-gnutls` package.

1. Use parameters specific to your environment when referring to the sample code provided for the following two versions and add the code to the end of the Rsyslog configuration file. The address of the Rsyslog configuration file is `/etc/rsyslog.conf`.

- Rsyslog V8 and later

Set `$DefaultNetstreamDriverCAFile` to the location of the root certificate in the system.

```bash
# Set up disk assisted queues
$WorkDirectory /var/spool/rsyslog # where to place spool files
$ActionQueueFileName fwdRule1 # unique name prefix for spool files
$ActionQueueMaxDiskSpace 1g # 1gb space limit (use as much as possible)
$ActionQueueSaveOnShutdown on # save messages to disk on shutdown
$ActionQueueType LinkedList # run asynchronously
$ActionResumeRetryCount -1 # infinite retries if host is down
$ActionSendTCPRebindInterval 100 # close and re-open the connection to the remote host every 100 of messages sent.
# RsyslogGnuTLS set to default ca path
$DefaultNetstreamDriverCAFile /etc/ssl/certs/ca-bundle.crt
```

```bash
# Send messages to Loggly over TCP using the template
$template(name="LogServiceFormat" type="string" string="<%pri%>%timestamp:::date-rfc3339% %HOSTNAME% %app-name% %procid% %msgid% [logservice project="test-project-1" logstore="test-logstore-1"] access-key-id="<yourAccessKeyId>" access-key-secret="<yourAccessKeySecret>"] %msg%n")
```

# Issue: 20200323
Log Service

Data Collection / 5 Other collection methods

- Rsyslog V7 and earlier

Set $DefaultNetstreamDriverCAFile to the location of the root certificate in the system.

```bash
# Set up disk assisted queues
$WorkDirectory /var/spool/rsyslog # where to place spool files
$ActionQueueFileName fwdRule1 # unique name prefix for spool files
$ActionQueueMaxDiskSpace 1g # 1gb space limit (use as much as possible)
$ActionQueueSaveOnShutdown # save messages to disk on shutdown
$ActionQueueType LinkedList # run asynchronously
$ActionResumeRetryCount -1 # infinite retries if host is down
$ActionSendStreamDriver gtls
$ActionSendStreamDriverMode 1
$ActionSendStreamDriverAuthMode x509/name
$ActionSendStreamDriverPermittedPeers "*.log.aliyuncs.com"

# RsyslogGnuTLS set to default ca path
$DefaultNetstreamDriverCAFile /etc/ssl/certs/ca-bundle.crt
$ActionSendStreamDriver gtls
$ActionSendStreamDriverMode 1
$ActionSendStreamDriverAuthMode x509/name
$ActionSendStreamDriverPermittedPeer cn-hangzhou.log.aliyuncs.com

template(name="LogServiceFormat" type="string" string="<%pri%> 1 %timestamp:::date-rfc3339% %HOSTNAME% %app-name% %procid% %msgid% [logservice project="test-project-1" logstore="test-logstore-1" access-key-id="<yourAccessKeyId>" access-key-secret="<yourAccessKeySecret>"] %msg%"); template_escape(no);
```

2. Run the `sudo service rsyslog restart`, `sudo /etc/init.d/syslog-ng` restart, or `systemctl restart rsyslog` command to restart Rsyslog.

3. If no Syslog message is generated, you can run the `logger` command to generate several test logs, for example, `logger hello world!`.

- Example 2: Use Syslog-ng to forward logs to Log Service

Syslog-ng is an open-source log management daemon that enables you to collect data by using Syslog on Unix and Unix-like operating systems. You can run the `sudo yum install syslog-ng` or `sudo apt-get install syslog-ng` command to install Syslog-ng.

```bash
### Syslog-ng Logging Config for LogService ###

```bash
destination d_logservice{
tcp("cn-hangzhou.log.aliyuncs.com" port(10009))
```
Log Service

```plaintext
tls(peer-verify(required-untrusted))
template(LogServiceFormat));
};
log {
    source(s_src); # default use s_src
    destination(d_logservice);
};
### END Syslog-ng Logging Config for LogService ###

Note:
Rsyslog is installed on each ECS instance by default to take over Syslog. You must uninstall Rsyslog before you can use Syslog-ng because Rsyslog and Syslog-ng cannot work at the same time.

1. Use parameters specific to your environment when referring to the preceding sample code and add the code to the end of the Syslog-ng configuration file. The address of the Syslog-ng configuration file is /etc/syslog-ng/syslog-ng.conf.
2. Run the `sudo /etc/init.d/syslog-ng restart`, `sudo service syslog-ng restart`, or `sudo systemctl restart syslog-ng` command to restart Syslog-ng.
3. If no Syslog message is generated, you can run the `logger` command to generate several test logs, for example, `logger hello world!`.

FAQ

- How can I manually send a Syslog message to Log Service?

You can use the `ncat` command to send a Syslog message to Log Service. This helps check network connectivity and whether the AccessKey pair is authorized to send Syslog messages. If `ncat` is not installed on your host, you can run the `sudo yum install nmap-ncat` command to install `ncat`. For example, you can run the following command to send a Syslog message to Log Service. The project in Log Service is named `test-project-1` and the Logstore is named `test-logstore-1`. The project resides in the China (Hangzhou) region. The AccessKeyId of the RAM user account with the write permission is `<yourAccessKeyId>`, and the AccessKey Secret is `<yourAccessKeySecret>`.

Note:
The timestamp for sending Syslog messages uses UTC+0, such as 2019-03-28T03:00:15.003Z. However, in UTC+8, this time is 2019-03-28T11:00:15.003.

The ncat command cannot be used to determine whether the connection is interrupted. You must enter the information to be sent and press Enter to trigger message sending within 30 seconds after running the ncat command.

```
[root@iZbp145dd9fccuidd7g**** ~]# ncat --ssl cn-hangzhou.log.aliyuncs.com 10009
<34>1 2019-03-28T03:00:15.003Z mymachine.example.com su - ID47 [logservice project="test-project-1" logstore="test-logstore-1" access-key-id="<yourAccessKeyId>" access-key-secret="<yourAccessKeySecret>"] this is a test message
```

After you send the Syslog message, you can preview it in the Log Service console. For more information, see [Preview logs](#).

- What can I do if a manually triggered Syslog message transmission fails?

  Troubleshoot the failure according to error messages. For more information, see [Diagnose collection errors](#).

- How can I view Rsyslog error messages?

  Rsyslog error messages are stored in `/var/log/message`. You can run the `vim` command to view Rsyslog error messages.

  - Rsyslog error messages (Example 1):
    ```
dlopen: /usr/lib64/rsyslog/lmnsd_gtls.so: cannot open shared object file: No such file or directory
```
    This error message is returned when the `rsyslog-gnutls` package is not installed. You can run the `sudo apt-get install rsyslog-gnutls` or `sudo yum install rsyslog-gnutls` command to install Rsyslog. Then, you must restart Rsyslog.

  - Rsyslog error message (Example 2):
    ```
unexpected GnuTLS error -53 - this could be caused by a broken connection. GnuTLS reports:Error in the push function
```
    This error message is returned when the TCP connection is closed because it has been idle for a long period of time. You can ignore the error because Rsyslog automatically reconnects to the server.
How can I view Rsyslog-ng error messages?

Syslog-ng error messages are stored in journals by default. You can run
`systemctl status syslog-ng.service` and `journalctl -xe` commands to view
Rsyslog-ng error messages.

The error message returned when Syslog-ng failed to start:

```
Job for syslog-ng.service failed because the control process exited
with error code. See "systemctl status syslog-ng.service" and "
journalctl -xe" for details
```

If this error occurs, check whether the configuration file format is valid or
whether configuration conflicts exist. For example, you cannot configure
multiple `internal()` sources.

5.5 Logstash

5.5.1 Install Logstash

Log Service provides a Logstash plug-in that allows you to upload log data through
Logstash.

Context

Logstash is a popular open-source data collection program. You can install
the logstash-output-logservice plug-in to upload data to Log Service. For more
information, see Logstash plug-in on Github.

Procedure
1. Install the JDK.
   a. Download the JDK installer.
      Go to the Java official website, download the JDK installer as required, and then double-click the installer to install the JDK.
   b. Set the environment variables.
      Add or modify environment variables in advanced system settings.
      
      - **PATH**: `C:\Program Files\Java\jdk1.8.0_73\bin`
      - **CLASSPATH**: `C:\Program Files\Java\jdk1.8.0_73\lib;C:\Program Files\Java\jdk1.8.0_73\lib\tools.jar`
      - **JAVA_HOME**: `C:\Program Files\Java\jdk1.8.0_73`
   c. Verify that the JDK is installed.
      Run PowerShell or cmd.exe for verification.

      ```
      PS C:\Users\Administrator> java -version
      java version "1.8.0_73"
      Java(TM) SE Runtime Environment (build 1.8.0_73-b02)
      Java HotSpot(TM) 64-Bit Server VM (build 25.73-b02, mixed mode)
      PS C:\Users\Administrator> javac -version
      javac 1.8.0_73
      ```

2. Install Logstash.
   a. Download the installation package.
      Download the Logstash installation package.

      ```
      Note:

      - We recommend that you download Logstash 5.0 or later.
      - Logstash 6.4.3 can be installed and runs properly on the following operating systems: macOS 10.14.1, Windows 7, and CentOS 7.
      ```
   b. Install Logstash.
      Decompress the installation package to a specified directory.
3. Install the plug-in used by Logstash to write logs to Log Service.
   Install the plug-in online or offline based on the network environment where the server resides.

   • Online installation:
     The plug-in is hosted by RubyGems. For more information, see here.
     Run PowerShell or cmd.exe to go to the Logstash installation directory. Run the following command to install the logstash-output-logservice plug-in:
     
     ```
     PS C:\logstash-6.4.3> .\bin\logstash-plugin install logstash-output-logservice
     ```
   
   • Offline installation:
     Go to the logstash-output-logservice page, and click Download in the lower-right corner.
     If the server from which logs are collected cannot access the Internet, copy the downloaded gem package to a local directory. Run PowerShell or cmd.exe to go to the Logstash installation directory. Run the following command to install the logstash-output-logservice plug-in:
     
     ```
     PS C:\logstash-6.4.3> .\bin\logstash-plugin install C:\logstash-6.4.3\logstash-output-logservice-0.4.0.gem
     ```
   
   • Verification:
     
     ```
     PS C:\logstash-6.4.3> .\bin\logstash-plugin list
     ```
     Verify that logstash-output-logservice exists in the plug-in list of the server.

5.5.2 Create Logstash collection configurations

Context

Related plug-ins

• logstash-input-file

   This plug-in is used to collect log files in tail mode. For more information, see logstash-input-file.
path indicates the file path, which must use UNIX separators, for example, C:/test/multiline/*.log. Otherwise, fuzzy match is not supported.

- logstash-output-logservice

This plug-in is used to output the logs collected by the logstash-input-file plug-in to Log Service.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>endpoint</td>
<td>Log Service endpoint. Example: <a href="http://regionid.example.com">http://regionid.example.com</a>. For more information, see Log Service endpoint.</td>
</tr>
<tr>
<td>project</td>
<td>The project name of Log Service.</td>
</tr>
<tr>
<td>logstore</td>
<td>The Logstore name.</td>
</tr>
<tr>
<td>topic</td>
<td>The log topic name. The default value is null.</td>
</tr>
<tr>
<td>source</td>
<td>The log source. If this parameter is set to null, the IP address of the current machine is used as the log source. Otherwise, the log source is subject to the specified parameter value.</td>
</tr>
<tr>
<td>access_key_id</td>
<td>The AccessKey ID of the Alibaba Cloud account.</td>
</tr>
<tr>
<td>access_key_secret</td>
<td>The AccessKey Secret of the Alibaba Cloud account.</td>
</tr>
<tr>
<td>max_send_retry</td>
<td>The maximum number of retries performed when data packets cannot be sent to Log Service because of an exception. Data packets with retry failures are discarded. The retry interval is 200 ms.</td>
</tr>
</tbody>
</table>

Procedure

1. Create collection configurations

Create a configuration file in the C:\logstash-2.2.2-win\conf\ directory and then restart Logstash to apply the file.

You can create a configuration file for each log type. The file name format is *.conf. For easier management, we recommend that you create all the configuration files in the C:\logstash-2.2.2-win\conf\ directory.

Note:
The configuration file must be encoded in UTF-8 format without BOM. You can use Notepad++ to modify the file encoding format.

- **IIS logs**
  
  For more information, see *Use Logstash to collect IIS logs*.

- **CSV logs**
  
  Use the system time of log collection as the log uploaded time. For more information, see CSV log configuration.

- **Logs with built-in time**
  
  Take CSV log format as an example. Use the time in the log content as the log uploaded time. For more information, see *Use Logstash to collect CSV logs*.

- **General logs**
  
  By default, the system time of log collection is used as the log uploaded time. Log fields are not parsed. Single-line logs and multiline logs are supported.
  
  For more information, see *Use Logstash to collect other logs*.

2. Verify configuration syntax

   a. Run PowerShell or cmd.exe to go to the Logstash installation directory:

   ```
   PS C:\logstash-2.2.2-win\bin> .\logstash.bat agent --configtest --config C:\logstash-2.2.2-win\conf\iis_log.conf
   ```

   b. Modify the collection configuration file. Temporarily add a line of rubydebug configuration in the output phase to output the collection results to the console. Set the type field as per your needs.

   ```ruby
   output {
     If [type] = "***"{
       stdout { codec => rubydebug }
     logservice {
   }
   }
   ```

   c. Run PowerShell or cmd.exe to go to the Logstash installation directory and start the process:

   ```
   PS C:\logstash-2.2.2-win\bin> .\logstash.bat agent -f C:\logstash-2.2.2-win\conf
   ```

   After the verification, end the logstash.bat process and delete the temporary configuration item rubydebug.
What's next

When logstash.bat is started in PowerShell, the Logstash process is working in the frontend. Logstash is generally used for testing configurations and debugging collections. Therefore, we recommend that you set Logstash as a Windows service after the debugging is passed so as to enable Logstash to work in the backend and start automatically when power-on. For how to set Logstash as a Windows service, see Set Logstash as a Windows service.

5.5.3 Set Logstash as a Windows service

When logstash.bat is started in PowerShell, the Logstash process is working in the frontend. Logstash is generally used for testing configurations and debugging collections. Therefore, we recommend that you set Logstash as a Windows service after the debugging is passed so as to enable Logstash to work in the backend and start automatically when power-on.

Besides setting Logstash as a Windows service, you can also start, stop, modify, and delete the service by using command lines. For more information about how to use NSSM, see NSSM official document.

Add Logstash as a Windows service

This operation is generally performed when Logstash is deployed for the first time. If Logstash has been added, skip this step.

Run the following command to add Logstash as a Windows service.

- 32-bit system

```sql
C:\logstash-2.2.2-win\nssm-2.24\win32\nssm.exe install logstash "C:\logstash-2.2.2-win\bin\logstash.bat" "agent -f $C:\logstash-2.2.2-win\conf"
```

- 64-bit system

```sql
C:\logstash-2.2.2-win\nssm-2.24\win64\nssm.exe install logstash "C:\logstash-2.2.2-win\bin\logstash.bat" "agent -f $C:\logstash-2.2.2-win\conf"
```

Start the service

If the configuration file in the Logstash conf directory is updated, stop the Logstash service and then start it again.

Run the following command to start the service.
- **32-bit system**
  
  C:\logstash-2.2.2-win\nssm-2.24\win32\nssm.exe start logstash

- **64-bit system**
  
  C:\logstash-2.2.2-win\nssm-2.24\win64\nssm.exe start logstash

**Stop the service**

**Run the following command to stop the service.**

- **32-bit system**
  
  C:\logstash-2.2.2-win\nssm-2.24\win32\nssm.exe stop logstash

- **64-bit system**
  
  C:\logstash-2.2.2-win\nssm-2.24\win64\nssm.exe stop logstash

**Modify the service**

**Run the following command to modify the service.**

- **32-bit system**
  
  C:\logstash-2.2.2-win\nssm-2.24\win32\nssm.exe edit logstash

- **64-bit system**
  
  C:\logstash-2.2.2-win\nssm-2.24\win64\nssm.exe edit logstash

**Delete the service**

**Run the following command to delete the service.**

- **32-bit system**
  
  C:\logstash-2.2.2-win\nssm-2.24\win32\nssm.exe remove logstash

- **64-bit system**
  
  C:\logstash-2.2.2-win\nssm-2.24\win64\nssm.exe remove logstash

### 5.5.4 Advanced functions

Logstash provides multiple plug-ins to meet personalized requirements. For example:

- **grok**: Structurally parses logs into multiple fields by using regular expressions.
- **json_lines** and **json**: Structurally parses JSON logs.
- **date**: Parses and converts the date and time fields of logs.
5.5.5 Logstash error processing

If you encounter the following collection errors when using Logstash to collect logs, follow the corresponding suggestions and process the errors.

- Data with garbled characters in Log Service
  
  Logstash supports UTF-8 file encoding by default. Check whether input files are correctly encoded or not.

- Error message in the console
  
  The error `io/console not supported; tty will not be manipulated` is prompted in the console. However, the error does not affect the functions and can be ignored.

  If other errors occur, we recommend that you search Google or Logstash forums for help.

5.6 SDK collection

5.6.1 Producer Library

Aliyun LOG Java Producer is an easy-to-use and highly configurable Java library that helps you send data to Log Service. It is designed for Java applications that are running in big data and high concurrency scenarios.

For more information, see Aliyun LOG Java Producer on GitHub.

5.6.2 Log4j Appender

Log4j is an open source logging framework developed by Apache. You can use Log4j to print logs to a console, file, Graphical User Interface (GUI) component, socket server, NT kernel logger, or Unix syslog daemon. You can define the output format of each log entry. You can also define the level of each log entry to achieve fine-grained control over the generation of logs.

Log4j consists of three main components: loggers, appenders, and layouts.
Log Service

- Loggers define log entry levels to indicate the severity of log entries.

  Log entry levels are classified into ERROR, WARN, INFO, and DEBUG in descending order of severity.

- Appenders deliver log entries to different destinations.

  A destination can be a console or file.

- Layouts define the output format of logs.

  They specify how logs are displayed.

You can use Alibaba Cloud Log Log4j Appender to print logs to Log Service. For information about the download link and user guide, see Log4j Appender.

5.6.3 Logback Appender

Logback is an open source project developed by the founder of Log4j. You can use Logback to print logs to a console, file, Graphical User Interface (GUI) component, socket server, NT kernel logger, or Unix syslog daemon. You can define the output format of each log entry. You can also define the level of each log entry to achieve fine-grained control over the generation of logs.

You can use Aliyun Log Logback Appender to print logs to Log Service. The following shows the format of a log that is printed to Log Service.

```
level: ERROR
location: com.aliyun.openservices.log.logback.example.LogbackAppenderExample.main(LogbackAppenderExample.java:18)
message: error log
throwable: java.lang.RuntimeException: xxx
thread: main
time: 2018-01-02T03:15+0000
__Source___: xxx
__topic___: yyy
```

For information about the download link and user guide, see Logback Appender.

5.6.4 C Producer Library

Besides the Producer Library of Java version, LogHub also supports the Producer Library and Producer Lite Library of the C version, which provides you with a simple and high-performance one-stop log collection solution across platforms and with low consumption of resources.

For the GitHub project address, see:
Log Service

5.6.5 Go Producer Library

Aliyun LOG Go Producer Library is an easy-to-use and highly configurable Go library. It is tailored to Go applications running in big data scenarios with high concurrency. By using Aliyun LOG Go Producer Library, Go applications automatically resend failed logs and compress the data to be sent to improve data writing efficiency.

For more information, see Aliyun Log Go Producer on GitHub.

5.7 Common log formats

5.7.1 Log4j logs

Access Mode

Log Service supports collecting Log4j logs by using:

- LogHub Log4j Appender
- Logtail

Collect Log4j logs by using LogHub Log4j Appender

For more information, see Log4j Appender.

Collect Log4j logs by using Logtail

The log4j log consists of the first and second generations, and this document takes the default configuration of the first generation as an example, describes how to configure regular, if log4j is used 2. You need to modify the default configuration to print the date completely.

```
<Configuration status="WARN">
  <Appenders>
    <Console name="Console" target="SYSTEM_OUT">
      <PatternLayout pattern="%d{yyyy-MM-dd HH:mm:ssSSS zzz} [%t] %-5level %logger{36} - %msg%n"/>
    </Console>
  </Appenders>
  <Loggers>
    <Logger name="com.foo.Bar" level="trace">
      <AppenderRef ref="Console"/>
    </Logger>
    <Root level="error">
      <AppenderRef ref="Console"/>
    </Root>
  </Loggers>
</Configuration>
```

Issue: 20200323
For how to configure Logtail to collect Log4j logs, see **Python logs**. Select the corresponding configuration based on your network deployment and actual situation.

The automatically generated regular expression is only based on the log sample and does not cover all the situations of logs. Therefore, you must adjust the regular expression slightly after it is automatically generated.

**Log4j** log sample of Log4j default log format printed to a file is as follows:

```
2013-12-25 19:57:06,954 [10.207.37.161] WARN impl.PermanentTairDaoImpl - Fail to Read Permanent Tair,key:e:470217319319741_1,result:com.example.tair.Result@172e3ebc[rc=code=-1, msg=connection error or timeout,value=,flag=0]
```

Matching of the beginning of a line in multiline logs (use IP to indicate the beginning of a line):

```
\d+-\d+-\d+ \[(\[^\]\]*)\] \S+\S+\S-\s.
```

The regular expression used to extract log information:

```
(\d+-\d+-\d+\s\d+:\d+:\d+\s\[(\[^\]\]*)\]\S+\S+\S-\s\.)
```

Time conversion format:

```
%Y-%m-%d %H:%M:%S
```

**Extraction results of the log sample:**

<table>
<thead>
<tr>
<th>Key</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>time</td>
<td>2013-12-25 19:57:06,954</td>
</tr>
<tr>
<td>ip</td>
<td>10.207.37.161</td>
</tr>
<tr>
<td>level</td>
<td>WARN</td>
</tr>
<tr>
<td>class</td>
<td>impl.PermanentTairDaoImpl</td>
</tr>
<tr>
<td>message</td>
<td>Fail to Read Permanent Tair,key:e:470217319319741_1,result:com.example.tair.Result@172e3ebc[rc=code=-1, msg=connection error or timeout,value=,flag=0]</td>
</tr>
</tbody>
</table>
5.7.2 Python logs

The logging module of Python provides a general logging system, which can be used by third-party modules or applications. The logging module provides different log levels and logging methods such as files, HTTP GET/POST, SMTP, and Socket. You can customize a logging method as needed. The logging module is the same as Log4j except that they have different implementation details. The logging module provides the logger, handler, filter, and formatter features.

To collect Python logs, we recommend you to use logging handler directly:

- Automatically upload Python logs by using log handler
- Log handler automatically parses logs in K-V format
- Log handler automatically parses logs in JSON format

Python log format

The log format specifies the output format of log records in formatter. The construction method of formatter needs two parameters: message format string and message date string. Both of the parameters are optional.

Python log format:

```python
import logging
import logging.handlers
LOG_FILE = 'tst.log'
handler = logging.handlers.RotatingFileHandler(LOG_FILE, maxBytes = 1024*1024, backupCount = 5) # Instantiate the handler
fmt = '%(asctime)s - %(filename)s:%(lineno)s - %(name)s - %(message)s'
formatter = logging.Formatter(fmt)   # Instantiate the formatter
handler.setFormatter(formatter)      # Add the formatter to the handler
logger = logging.getLogger('tst')    # Obtain the logger named tst
logger.addHandler(handler)           # Add the handler to the logger
logger.setLevel(logging.DEBUG)
logger.info('first info message')
logger.debug('first debug message')
```

Field description

The formatter is configured in the %(key)s format, that is, replacing the dictionary keywords. The following keywords are provided:

<table>
<thead>
<tr>
<th>Format</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>%(name)s</td>
<td>The logger name of the generated log.</td>
</tr>
<tr>
<td>Format</td>
<td>Meaning</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>%(levelname)s</td>
<td>The log level in text format, including DEBUG, INFO, WARNING, ERROR, and CRITICAL.</td>
</tr>
<tr>
<td>%(pathname)s</td>
<td>The full path of the source file where the statement that outputs the log resides (if available).</td>
</tr>
<tr>
<td>%(filename)s</td>
<td>The file name.</td>
</tr>
<tr>
<td>%(module)s</td>
<td>The name of the module where the statement that outputs the log resides.</td>
</tr>
<tr>
<td>%(funcName)s</td>
<td>The name of the function that calls the log output.</td>
</tr>
<tr>
<td>%(lineno)d</td>
<td>The code line where the function statement that calls the log output resides (if available).</td>
</tr>
<tr>
<td>%(created)f</td>
<td>The time (in the UNIX standard time format) when the log is created, which indicates the number of seconds since 1970-1-1 00:00:00 UTC.</td>
</tr>
<tr>
<td>%(relativeCreated)d</td>
<td>The interval (in milliseconds) between the log created time and the time that the logging module is loaded.</td>
</tr>
<tr>
<td>%(asctime)s</td>
<td>The log creation time, which is in the format of “2003-07-08 16:49:45,896” by default (the number after the comma (,) is the number of milliseconds).</td>
</tr>
<tr>
<td>%(msecs)d</td>
<td>The log creation time in the milliseconds.</td>
</tr>
<tr>
<td>%(thread)d</td>
<td>The thread ID (if available).</td>
</tr>
<tr>
<td>%(threadName)s</td>
<td>The thread name (if available).</td>
</tr>
<tr>
<td>%(process)d</td>
<td>The process ID (if available).</td>
</tr>
<tr>
<td>%(message)s</td>
<td>The log message.</td>
</tr>
</tbody>
</table>
Log sample

Log sample


Common Python logs and the corresponding regular expressions:

- Log format

```
2016-02-19 11:03:13.410 - test.py:19 - tst - first debug message
```

Regular expression:
```
(\d+-\d+-\d+\s\S+)\s+-\s+(\[^:]+):(\d+)\s+-\s+(\w+)\s+-\s+(. *)
```

- Log format

```
%(asctime)s - %(filename)s:%(lineno)s - %(levelname)s %(pathname)s %(module)s %(funcName)s %(created)f %(thread)d %(threadName)s %(process)d %(name)s - %(message)s
```

Log sample

```
2016-02-19 11:06:52.514 - test.py:19 - 10 DEBUG test.py test <module > 1455851212.514271 139865996687072 MainThread 20193 tst - first debug message
```

Regular expression:
```
(\d+-\d+-\d+\s\S+)\s+-\s+(\[^:]+):(\d+)\s+-\s+(\d+)\s+(\w+)\s+(\S+)\s+(\w+)\s+(\S+)\s+(\S+)\s+(\d+)\s+(\w+)\s+-\s+(. *)
```

Configure Logtail to collect Python logs

For the detailed procedures of collecting Python logs by using Logtail, see #unique_242. Select the corresponding configuration based on your network deployment and actual situation.

1. Create a project and a Logstore. For detailed procedures, see #unique_111.
2. On the Logstores page, click the Data Import Wizard icon.
3. Select a data source.

   Select the Text File.
4. Configure the data source.
   
   a. Enter the Configuration Name and Log Path, and then select the Full Regex Mode from the mode drop-down list.
   
   b. Turn on the Singleline switch.
   
   c. Enter Log Sample.

   ![Configuration Interface]

   d. Turn on the Extract Field switch.
   
   e. Configure Regular Expression.

   A. Generates a regular expression by selecting strings of the log sample.

   If the automatically generated regular expression does not match your log sample, you can generate a regular expression by selecting strings of the log sample. Log Service supports selecting strings to automatically parse the log sample, that is, to automatically generates a regular expression for each selected field. In Log Sample, select log fields and click Generate RegEx. A regular expression of each selected field is displayed in the Regular
Expression column. You can generate a full regular expression for the log sample through multiple selections.

B. Modify the regular expression.

Considering the format of the actual log data may have minor changes, click Manually Input to adjust the automatically generated regular expression.
according to the actual situations to conform to all log formats that may occur in the collection process.

C. Validate the regular expression.

Click Validate after modifying the regular expression. If the regular expression is correct, extracted results are displayed. Modify the regular expression if any errors exist.

f. Confirm Extraction Results.

View the parsing results of the log fields and enter corresponding keys for the log extraction results.

Assign a descriptive field name for each log field extraction result. For example, assign time for the time field. If you do not use the system time, you must specify a field where value is time, and name its key as time.

g. Turn on the System Time switch.

If you use the system time, the time of each log is the time when the Logtail client parses the log.

h. (Optional) Configure Advanced options.
i. Click Next.

After completing Logtail configuration, apply the configuration to the machine group to collect Python logs.
5.7.3 Node.js logs

By default, Node.js logs are printed to the console, which makes the data collection and troubleshooting inconvenient. By using Log4js, logs can be printed to files and log format can be customized, which is convenient for data collection and coordination.

```javascript
var log4js = require('log4js');
log4js.configure({
    appenders: [
        {
            type: 'file', // file output
            filename: 'logs/access.log',
            maxLogSize: 1024,
            backups: 3,
            category: 'normal'
        }
    ]
});
var logger = log4js.getLogger('normal');
logger.setLevel('INFO');
logger.info("this is a info msg");
logger.error("this is a err msg");
```

Log format

After the log data is stored in the text file format by using Log4js, the log is displayed in the following format in the file:

```
[2016-02-24 17:42:38.946] [INFO] normal - this is a info msg
[2016-02-24 17:42:38.951] [ERROR] normal - this is a err msg
```

Log4js has six output levels, including trace, debug, info, warn, error, and fatal in ascending order.

Collect Node.js logs by using Logtail

For how to configure Logtail to collect Log4j logs, see Python logs. Select the corresponding configuration based on your network deployment and actual situation.

The automatically generated regular expression is only based on the log sample and does not cover all the situations of logs. Therefore, you must adjust the regular expression slightly after it is automatically generated. Therefore, you must adjust the regular expression slightly after it is automatically generated. See the following Node.js log samples for reference and write a correct and comprehensive regular expression for your log.

See the following common Node.js logs and the corresponding regular expressions:
• Log sample 1:
  - Log sample:
    ```
    [2016-02-24 17:42:38.946] [INFO] normal - this is a info msg
    ```
  - Regular expression type
    ```
    \[[([\^])\]+]\s\[[([\^])\]+]\s(\w+)\s-(. *)
    ```
  - Extracted fields:
    time, level, loggerName and message.

• Log sample 2:
  - Log sample:
    ```
    [2016-01-31 12:02:25.844] [INFO] access - 42.120.73.203 - - "GET /user/projects/ali_sls_log? ignoreError=true HTTP/1.1" 304 - "http://aliyun.com/" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_3) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/48.0.2564.97 Safari/537.36"
    ```
  - Regular expression type
    ```
    \[[([\^])\]+]\s\[[([\w])\]+]\s(\w+)\s-\s(\w+)\s-\s-\s(\S+)\s-\s(\S+)\s-\s(\[^\"]\)+\s(\[^\"]\)+. *
    ```
  - Extracted fields:
    time, level, loggerName, ip, request, status, referer and user_agent.

5.7.4 WordPress logs

Default WordPress log format

Raw sample log:

```
172.64.0.2 - - [07/Jan/2016:21:06:39 +0800] "GET /wp-admin/js/password-strength-meter.min.js? ver=4.4 HTTP/1.0" 200 776 "http://wordpress.c4ala0eecd194316955231dccc4adfb7.cn-hangzhou.alicontainer.com/wp-admin/install.php" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_5)
```
AppleWebKit/537.36 (KHTML, like Gecko) Chrome/47.0.2526.106 Safari/537.36

tching of the beginning of a line in multiline logs (use IP to indicate the beginning of a line):
\d+\./\d+\./\d+\./\d+\s-\s.*

The regular expression used to extract log information:
(\S+) - - \[(\[^\]\]*)\] "(\S+)\s"\[(^\+)+\]\s"(\^\+)+\"

Time conversion format:
%d/%b/%Y:%H:%M:%S

Extraction results of the log sample:

<table>
<thead>
<tr>
<th>Key</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ip</td>
<td>10.10.10.1</td>
</tr>
<tr>
<td>time</td>
<td>07/Jan/2016:21:06:39 +0800</td>
</tr>
<tr>
<td>method</td>
<td>GET</td>
</tr>
<tr>
<td>url</td>
<td>/wp-admin/js/password-strength-meter.min.js?ver=4.4 HTTP/1.0</td>
</tr>
<tr>
<td>status</td>
<td>200</td>
</tr>
<tr>
<td>length</td>
<td>776</td>
</tr>
<tr>
<td>ref</td>
<td><a href="http://wordpress.c4a1a0aecedb1943169555231dcc4adf7.cn-hangzhou.aliyuncs.com/wp-admin/install.php">http://wordpress.c4a1a0aecedb1943169555231dcc4adf7.cn-hangzhou.aliyuncs.com/wp-admin/install.php</a></td>
</tr>
<tr>
<td>user-agent</td>
<td>Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_5) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/47.0.2526.106 Safari/537.36</td>
</tr>
</tbody>
</table>

5.7.5 ThinkPHP logs

ThinkPHP is a Web application development framework based on the PHP language.

Log format

Logs are printed in the following format in ThinkPHP:

```php
<?php
Think\Log::record('D method instantiation does not find the model class');
```
Log example

```
[ 2016-05-11T21:03:05+08:00 ] 10.10.10.1 /index.php
INFO: [ app_init ] --START--
INFO: Run Behavior\BuildLiteBehavior [ RunTime:0.000014s ]
INFO: [ app_init ] --END-- [ RunTime:0.000091s ]
INFO: [app_begin] -- start --
INFO: Run Behavior\ReadHtmlCacheBehavior [ RunTime:0.000038s ]
INFO: [ app_begin ] --END-- [ RunTime:0.000076s ]
INFO: [ view_parse ] --START--
INFO: Run Behavior\ParseTemplateBehavior [ RunTime:0.000068s ]
INFO: [ view_parse ] --END-- [ RunTime:0.000104s ]
INFO: [ view_filter ] --START--
INFO: Run Behavior\WriteHtmlCacheBehavior [ RunTime:0.000032s ]
INFO: [ view_filter ] --END-- [ RunTime:0.000062s ]
INFO: [ app_end ] --START--
INFO: Run Behavior\ShowPageTraceBehavior [ RunTime:0.000032s ]
INFO: [ app_end ] --END-- [ RunTime:0.000070s ]
ERR: D method instantiation does not find the model class
```

Configure Logtail to collect ThinkPHP logs

For the complete process of collecting ThinkPHP logs by using Logtail, see *Python logs*. Select the corresponding configuration based on your network deployment and actual situation.

The automatically generated regular expression is only based on the log sample and does not cover all the situations of logs. Therefore, you must adjust the regular expression slightly after it is automatically generated.

ThinkPHP logs are multiline logs whose mode is not fixed. The following fields can be extracted from the ThinkPHP logs: time, access IP, accessed URL, and printed message. The message field contains multiple lines of information and can only be packaged to one field because the mode is not fixed.

Logtail collects configuration parameters of ThinkPHP logs
Regular expression at the beginning of the line:
```
\[s\d+-\d+-\w+:\d+:\d+:+\d+:\d+s.
```

Regular expression:
```
\[s\(\d+-\d+-\w+:\d+:+\d+s\)]\s+(\S+)\s+(\S+)\s+(.*
```

Time expression:
```
%Y-%m-%dT%H:%M:%S
```

5.7.6 Use Logstash to collect IIS logs

You need to modify the configuration file to parse the IIS log fields before you use logstash to capture the IIS log.

Log sample

View IIS log configurations, select the W3C format (default field setting), and save the format to put it into effect.

```
2016-02-25 01:27:04 112.74.74.124 GET /goods/list/0/1.html - 80 - 66.249.65.102 Mozilla/5.0+(compatible;+Googlebot/2.1;++http://www.google.com/bot.html) 404 0 2 703
```

Collection configuration

```
input {
  file {
    type => "iis_log_1"
    path => "[C:/inetpub/logs/LogFiles/W3SVC1/*.log]"
    start_position => "beginning"
  }
  filter {
    if [type] == "iis_log_1" {
      # ignore log comments
      if [message] =~ "^#" {
        drop {}
      }
    }
    grok {
      # check that fields match your IIS log settings
    }
    date {
      match => [ "log_timestamp", "YYYY-MM-dd HH:mm:ss" ]
      timezone => "Etc/UTC"
    }
    useragent {
      source=> "useragent"
      prefix=> "browser"
    }
```
Log Service

Data Collection / 5 Other collection methods

```ruby
mutate {
  remove_field => [ "log_timestamp"]
}
}

output {
  if [type] == "iis_log_1" {
    logservice {
      codec => "json"
      endpoint => "***"
      project => "***"
      logstore => "***"
      topic => ""
      source => ""
      access_key_id => "***"
      access_key_secret => "***"
      max_send_retry => 10
    }
  }
}
```

**Note:**

- The configuration file must be encoded in UTF-8 format without BOM. You can use Notepad++ to modify the file encoding format.
- *path* indicates the file path, which must use delimiters in the UNIX format, for example, *C:/test/multiline/*.log*. Otherwise, fuzzy match is not supported.
- The *type* field must be modified unified and kept consistent in the file. If a machine has multiple Logstash configuration files, the type field in each configuration file must be unique. Otherwise, data cannot be processed properly.

Related plug-ins: *file* and *grok*.

**Restart Logstash to apply configurations**

Create a configuration file in the *conf* directory and restart Logstash to apply the file. See *Set Logstash as a Windows service* for more information.

### 5.7.7 Use Logstash to collect CSV logs

You need to modify the configuration file to parse the CSV log fields before you use logsturg to capture the CSV log. The acquisition of the CSV log can use the system time of the acquisition log as the upload log time, you can also use the time in the contents of the log as the upload log time. For different definitions of log time, there are two ways to configure logstroudsburg to collect CSV logs.
Use the system time as the uploaded log time

- **Log sample**


- **Collection configuration**

  ```ruby
  input {
    file {
      type => "csv_log_1"
      path => ["C:/test/csv/*.log"]
      start_position => "beginning"
    }
  }
  filter {
    if [type] == "csv_log_1" {
      csv {
        separator => ","
        columns => ["ip", "a", "date", "time", "b", "latency", "status", "size", "method", "url", "file"]
      }
    }
  }
  output {
    if [type] == "csv_log_1" {
      logservice {
        codec => "json"
        endpoint => "***"
        project => "***"
        logstore => "***"
        topic => ""
        source => ""
        access_key_id => "***"
        access_key_secret => "***"
        max_send_retry => 10
      }
    }
  }
  ```

**Note:**

- The configuration file must be encoded in UTF-8 format without BOM. You can use Notepad++ to modify the file encoding format.
- *path* indicates the file path, which must use delimiters in the UNIX format, for example, `C:/test/multiline/*.log`. Otherwise, fuzzy match is not supported.
- *type* field must be modified unitedly and kept consistent in the file. If a machine has multiple Logstash configuration files, *type* field in each
configuration file must be unique. Otherwise, data cannot be processed properly.

Related plug-ins: *file* and *csv*.

- Restart Logstash to apply configurations

Create a configuration file in the *conf* directory and restart Logstash to apply the file. For more information, see Set *Set Logstash as a Windows service* as a Windows service.

Upload the log field content as the log time

- Log sample

| 10.116.14.201,--,Feb 25 2016 14:03:44,WSVC7,1332,200,0,GET,project/shenzhen-test/logstore/logstash/detail,C:\test\csv\test_csv_withtime\log |

- Collection configuration

```ruby
input {
  file {
    type => "csv_log_2"
    path => ["C:/test/csv_withtime/*.log"]
    start_position => "beginning"
  }
}
filter {
  if [type] == "csv_log_2" {
    csv {
      separator => "",
      columns => ["ip", "a", "datetime", "b", "latency", "status", "size", "method", "url", "file"]
    }
    date {
      match => [ "datetime", "MMM dd YYYY HH:mm:ss" ]
    }
  }
}
output {
  if [type] == "csv_log_2" {
    logservice {
      codec => "json"
      endpoint => "***"
      project => "***"
      logstore => "***"
      topic => ""
      source => ""
      access_key_id => "***"
      access_key_secret => "***"
      max_send_retry => 10
    }
  }
```
5.7.8 Use Logstash to collect other logs

You can modify the configuration file to parse log fields before you use logstash to capture logs.

Upload using system time as log time

- **Log sample**

  ```
  2016-02-25 15:37:01 [main] INFO com.aliyun.sls.test_log4j - single line log
    java.lang.ArithmeticException: / by zero
    at com.aliyun.sls.test_log4j.divide(test_log4j.java:23) ~[bin/?:?]
    at com.aliyun.sls.test_log4j.main(test_log4j.java:13) [bin/?:?]
  ```

- **Collection configuration**

  ```
  input {
    file {
      type => "common_log_1"
      path => ["C:/test/multiline/*.log"]
      start_position => "beginning"
    }
  }
  ```

### Note:

- The configuration file must be encoded in UTF-8 format without BOM. You can use Notepad++ to modify the file encoding format.
- *path* indicates the file path, which must use delimiters in the UNIX format, for example, *C:/test/multiline/*.log*. Otherwise, fuzzy match is not supported.
- *type* field must be modified unitedly and kept consistent in the file. If a machine has multiple Logstash configuration files, *type* field in each configuration file must be unique. Otherwise, data cannot be processed properly.

**Related plug-ins:** *file* and *csv*.

- Restart Logstash to apply configurations

  Create a configuration file in the *conf* directory and restart Logstash to apply the file. For more information, see *Set Logstash as a Windows service* as a Windows service.
Log Service

Data Collection / 5 Other collection methods

codec => multiline {
    pattern => "^\d{4}-\d{2}-\d{2} \d{2}:\d{2}:\d{2}\$"
    negate => true
    auto_flush_interval => 3
    what => previous
}
}
}
output {
    if [type] == "common_log_1" {
        logservice {
            codec => "json"
            endpoint => "***"
            project => "***"
            logstore => "***"
            topic => ""
            source => ""
            access_key_id => "***"
            access_key_secret => "***"
            max_send_retry => 10
        }
    }
}

Note:
- The configuration file must be encoded in UTF-8 format without BOM. You can use Notepad++ to modify the file encoding format.
- path indicates the file path, which must use delimiters in the UNIX format, for example, C:/test/multiline/*.log. Otherwise, fuzzy match is not supported.
- type field must be modified unitedly and kept consistent in the file. If a machine has multiple Logstash configuration files, the type field in each configuration file must be unique. Otherwise, data cannot be processed properly.

Related plug-ins: file and multiline(for a single-line log file, remove the codec => multiline configuration).

• Restart Logstash to apply configurations

Create a configuration file in the conf directory and restart Logstash to apply the file. For more information, see Set Logstash as a Windows service.

5.7.9 Unity3D logs

Context

Issue: 20200323
Unity3D is an integrated game development tool compatible with multiple platforms. Developed by Unity Technologies, this tool allows a player to easily create various interactive contents such as 3D video game, architectural visualization, and real-time 3D animation. Unity3D is a fully integrated and professional game engine.

You can use the Web Tracking function of Log Service to collect Unity3D logs conveniently. This document introduces how to use the Web Tracking function to collect Unity logs to Log Service by collecting the `Unity Debug.Log`.

**Procedure**

1. Activate the Web Tracking function
   - For more information, see [Web Tracking](#).

2. Register Unity3D LogHandler
   - Create a C# file `LogOutputHandler.cs` in the Unity editor. Enter the following codes and modify three member variables in the codes, which are:
     - `project`, indicating the name of the log project.
     - `logstore`, indicating the name of the Logstore.
     - `serviceAddr`, indicating the address of the log project.
   - For more information, see [C# source code](#).

```csharp
using UnityEngine;
using System.Collections;
public class LogOutputHandler : MonoBehaviour
{
    //Register the HandleLog function on scene start to fire on debug.log events
    public void OnEnable()
    {
        Application.logMessageReceived += HandleLog;
    }
    //Remove callback when object goes out of scope
    public void OnDisable()
    {
        Application.logMessageReceived -= HandleLog;
    }
    string project = "your project name";
    string logstore = "your logstore name";
    string serviceAddr = "http address of your log service project";
    //Capture debug.log output, send logs to Loggly
    public void HandleLog(string logString, string stackTrace,
                          LogType type)
    {
        string parameters = "";
        parameters += "Level=" + WWW.EscapeURL(type.ToString());
        parameters += "&";
        parameters += "Message=" + WWW.EscapeURL(logString);
```
parameters += "&";
parameters += "Stack_Trace=" + WWW.EscapeURL(stackTrace);
parameters += "&";
// Add any User, Game, or Device MetaData that would be
useful to finding issues later
parameters += "Device_Model=" + WWW.EscapeURL(SystemInfo.
deviceModel);
string url = "http://" + project + "." + serviceAddr + "/
logstores/" + logstore + "/track? APIVersion=0.6.0&" + parameters;
StartCoroutine(SendData(url));

The preceding codes can asynchronously send logs to Alibaba Cloud Log Service
. You can add more fields that you want to collect in the example.

3. Generate Unity logs

In the project, create the LogglyTest.cs file and add the following codes:

using UnityEngine;
using System.Collections;
public class LogglyTest : MonoBehaviour {
    void Start () {
        Debug.Log ("Hello world");
    }
}

4. Preview the log in the console.

After completing the preceding steps, run the Unity program. Then, you can
preview your sent logs in the Log Service console.

The preceding example provides the methods for collecting logs such as Debug
.Log, Debug.LogError, and Debug.LogException. The component object model
of Unity, its program crash API, and other types of Log APIs can be used to
conveniently collect the device information on the client.
6 Best practices

6.1 Collect IoT or embedded development logs

Internet of Things (IoT) is enjoying high growth. More and more IoT devices are being applied to our daily life, such as smart routers, various TV dongles, Tmall Genie, and robot vacuum cleaners, bringing us the convenience of intelligence. According to Gartner's prediction, 20 billion smart devices will be put into service by the end of 2020, which gives a glimpse of the huge market in this field. The embedded development model in the traditional software field is faced with great challenges in the IoT field. In addition to the large number and wide distribution, IoT devices are difficult to debug and restricted in hardware. As a result, traditional device log solutions cannot meet the demands perfectly.

Based on years of Logtail development experience and the characteristics of IoT devices, the Log Service team has customized a log data collection solution for IoT devices: C Producer Library.

Embedded development requirements

The IoT or embedded development engineers must have profound knowledge and experience in development and the capability to manage, monitor, and diagnose
a large number of black boxes. Embedded development mainly has the following requirements:

- **Data collection**: How can the engineers collect data from millions or even tens of millions of devices distributed all around the world in real time?
- **Debugging**: How can the engineers use one solution to meet the requirements of both online data collection and real-time debugging in development?
- **Online diagnosis**: When an error occurs on an online device, how can the engineers locate the device quickly and check the error context?
- **Monitoring**: How many devices are currently online? How is the working status distribution? How is the geographic distribution? How does a device send alerts in real time when an error occurs?
- **Real-time data analysis**: How is the data generated by devices integrated with real-time computing and big data warehouses to build user profiles?

**Problem investigation costs most of the time:**

1. Black box environment
2. Distributed and disperse

**Time consumed in Ops scenario**

**Major challenges in the IoT field**

When thinking about the solutions to the preceding problems, we find that the approaches in the traditional software field are ineffective in the IoT field. The main challenges arise from the following characteristics of IoT devices:

- **Large numbers of devices**: In the traditional O&M field, a company managing 10,000 servers is qualified for a large one. However, managing 100,000 online devices is only a small threshold in the IoT field.
• Wide distribution: The deployed hardware devices are usually distributed around the country or even the world.

• Black box: IoT devices are mostly in unknown states. It is difficult to log on to and debug these devices.

• Restrictions in resources: IoT devices are relatively restricted in hardware to reduce costs, for example, the total memory size may only be 32 MB. As a result, traditional log collection approaches for PCs do not work in the IoT field.

C Producer Library: a log data collection solution customized by Log Service

Logtail, the client of Log Service, is deployed on millions of x86 servers. For more information, see Logtail technology sharing: and . In addition, Log Service provides a variety of collection solutions:

• Mobile SDK: It collects data from Android or iOS platforms with tens of millions of daily active user (DAUs).

• Web Tracking (JS): Similar to Baidu Tongji and Google Analytics, it uses a lightweight collection mode without signature.

In the IoT field, based on years of Logtail development experience and the characteristics of IoT devices in terms of CPU, memory, disk, network, and application mode, we have developed a log data collection solution for IoT devices: C Producer Library.

**Features of C Producer Library**

Like a lightweight Logtail, C Producer Library offers high stability, high performance, and low resource consumption. Although it does not have the feature
of real-time configuration management in Logtail, C-Producer Library has 70% of the features of Logtail, including:

- Multiple tenants: C Producer Library can process various types of logs (such as Metric, DebugLog, and ErrorLog) according to their priorities. Multiple clients can be configured and each client can be separately configured with the collection priority and target project or Logstore.
- Context query: Logs generated by the same client are in the same context, and the relevant logs before and after a log can be viewed.
- Concurrent sending and resumable upload: The upper limit of cache can be set. Logs fail to be written when the upper limit is exceeded.

In addition, C Producer Library provides the following features specific to IoT devices, including:

- Local debugging: Logs can be exported to local devices. You can set the rotation, log quantity, and rotation size.
- Fine-grained resource control: Different cache upper limits and aggregation modes can be set for different types of data or logs.
- Log cache compression: The cache of the data failed to be sent can be compressed to reduce the memory usage of devices.

Advantages of C Producer Library

As a custom solution for IoT devices, C Producer Library has obvious advantages in the following aspects:
• Highly concurrent write on the client: Hundreds of thousands of logs can be written every second with a configurable sending thread pool. For more information, see "Performance test" in this topic.

• Low resource consumption: Only 70% of the CPU is occupied when 200,000 log entries are written every second. The resources are not affected when up to 100 log entries are generated on low-performance hardware (for example, Raspberry Pi) every second.

• No data copies on disks of client logs: Data is directly sent to the server through the network after being generated.

• Client computing separated from I/O logic: Logs are generated asynchronously, without blocking the working thread.

• Multiple priorities: Different clients can be configured with different priorities to make sure that logs with higher priorities are sent first.

• Local debugging: Local debugging can be configured to facilitate you to test the applications locally when the network is unavailable.

In the preceding scenarios, C Producer Library simplifies application development. You do not have to consider log collection details or worry about the impact of log collection on your business operations. This makes data collection significantly easier.

To make it distinct, we made a comparison between C Producer Library and other embedded collection solutions. The following table lists the comparison results.
<table>
<thead>
<tr>
<th>Type</th>
<th>Platform</th>
<th>C Producer Library</th>
<th>Other solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Programming</td>
<td>Mobile + Embedded</td>
<td>Mobile-based</td>
<td></td>
</tr>
<tr>
<td>Context</td>
<td>Supported</td>
<td>Not supported</td>
<td></td>
</tr>
<tr>
<td>Multiple logs</td>
<td>Supported</td>
<td>Not supported (one type of logs)</td>
<td></td>
</tr>
<tr>
<td>Custom format</td>
<td>Supported</td>
<td>Not supported (several limited fields are provided)</td>
<td></td>
</tr>
<tr>
<td>Priority</td>
<td>Supported</td>
<td>Not supported</td>
<td></td>
</tr>
<tr>
<td>Environment parameter</td>
<td>Configurable</td>
<td>Configurable</td>
<td></td>
</tr>
<tr>
<td>Stability</td>
<td>Concurrency</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Concurrency</td>
<td>LZ4 (balance between efficiency and performance) + Gzip</td>
<td>Optimized</td>
<td></td>
</tr>
<tr>
<td>Compression algorithm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low resource consumption</td>
<td>Optimized</td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td>Transmission</td>
<td>Resumable upload</td>
<td>Supported</td>
<td>By default, resumable upload is not supported. Secondary development is required.</td>
</tr>
<tr>
<td>Access point</td>
<td>8 (in China) + 8 (outside China)</td>
<td>Hangzhou</td>
<td></td>
</tr>
<tr>
<td>Debugging</td>
<td>Local log</td>
<td>Supported</td>
<td>Supported in manual mode</td>
</tr>
<tr>
<td>Parameter configuration</td>
<td>Supported</td>
<td>Not supported</td>
<td></td>
</tr>
<tr>
<td>Real-time performance</td>
<td>Visible on the server side</td>
<td>1 second (99.9%) to 3 seconds (maximum)</td>
<td>1 to 2 hours</td>
</tr>
<tr>
<td>Custom processing</td>
<td>More than 15 interconnection modes</td>
<td>Customized real-time and offline solution</td>
<td></td>
</tr>
</tbody>
</table>
C Producer Library + Log Service solution

C Producer Library integrates with Alibaba Cloud Log Service to form a complete set of log solutions for IoT devices.

- Large scale
  - Supports writing hundreds of millions of log entries on the client in real time.
  - Supports writing petabytes of data every day.

- High speed
  - Fast data collection: Data can be consumed after being written without any latency.
  - Quick query: Billions of data records can be processed and queried within 1 second by using a complex query statement (with five conditions).
  - Rapid analysis: Hundreds of millions of data records can be aggregated and analyzed within 1 second by using a complex analysis statement (aggregated with five dimensions and the GroupBy statement).

- Wide interconnection
  - Seamlessly integrated with various Alibaba Cloud products.
  - Compatible with various open-source storage, computing, and visual systems.

Download and use

Download URL: GitHub

One application can create multiple producers, and each producer can include multiple clients. Each client can be independently configured with the target address, log level, local debugging, cache size, custom identifier, and topic.
For more information about installation methods and operation steps, see `README`.

Performance test

Environment configuration

- High-performance scenario: traditional x86 servers.
- Low-performance scenario: Raspberry Pi (low power consumption environment).

The following figure shows the configurations.

<table>
<thead>
<tr>
<th>High-performance scenario</th>
<th>Low-performance scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>• CPU: Intel(R) Xeon(R) CPU E5-2682 v4 @ 2.50 GHz</td>
<td></td>
</tr>
<tr>
<td>• Memory: 64 GB</td>
<td></td>
</tr>
<tr>
<td>• Operating system: Linux version 2.6.32-220.23.2.el5.x86_64</td>
<td></td>
</tr>
<tr>
<td>• GCC: 4.1.2</td>
<td></td>
</tr>
<tr>
<td>• C-Producer: Dynamic library is 162 KB. Static library is 140 KB. (Use the static library for test. The binary after compilation is 157 KB. All are stripped.)</td>
<td></td>
</tr>
<tr>
<td>Type: Raspberry Pi 3B</td>
<td></td>
</tr>
<tr>
<td>CPU: Broadcom BCM2837 1.2 GHz A53 64 bit. (Use the host USB for power supply. The frequency is lowered to 600 MHz.)</td>
<td></td>
</tr>
<tr>
<td>Memory: 1 GB DDR2</td>
<td></td>
</tr>
<tr>
<td>Operating system: Linux 4.9.41-v7+ #1023 SMP armv71 GNU/Linux</td>
<td></td>
</tr>
<tr>
<td>GCC: 6.3.0 (Raspbian 6.3.0-18+rp1)</td>
<td></td>
</tr>
<tr>
<td>C-Producer: Dynamic library is 179 KB. Static library is 162 KB. (Use the static library for test. The binary after compilation is 287 KB. All are stripped.)</td>
<td></td>
</tr>
</tbody>
</table>

C Producer Library configuration
- ARM (Raspberry Pi)
  - Cache: 10 MB
  - Aggregation time: 3 seconds (If any of the conditions is met, namely, aggregation time, aggregation data package size, and aggregation log quantity, the data is packaged and sent.)
  - Aggregation data package size: 1 MB
  - Aggregation log quantity: 1,000
  - Sending thread: 1
  - Custom tag: 5

- x86
  - Cache: 10 MB
  - Aggregation time: 3 seconds (If any of the conditions is met, namely, aggregation time, aggregation data package size, and aggregation log quantity, the data is packaged and sent.)
  - Aggregation data package size: 3 MB
  - Aggregation log quantity: 4,096
  - Sending thread: 4
  - Custom tag: 5

Sample log

1. The total data volume is approximately 600 Bytes for 10 key-value pairs.
2. The total data volume is approximately 350 Bytes for 9 key-value pairs.

```
__source__: 11.164.233.187
__tag__:i: 2
__tag__:i5: 6
__tag__:a: b
__tag__:c: d
__tag__:tag_key: tag_value
__topic__: topic_test
_file_: /disk1/workspace/tools/aliyun-log-c-sdk/sample/log_producer_sample.c
_function_: log_producer_post_logs
_level_: LOG_PRODUCER_LEVEL_WARN
_line_: 248
_thread_: 40978304
LogHub: Real-time log collection and consumption
Search/Analytics: Query and real-time analysis
Interconnection: Grafana and JDBC/SQL92
Visualized: dashboard and report functions
```
Test results on the x86 platform

- C Producer Library sends up to 90 MB data every second. It consumes only 70% of CPU and 140 MB of memory to upload 200,000 log entries every second.
- When the sending speed of the server is 200 entries every second, data sending basically does not have any impact on CPU (reduced to less than 0.01%).
- The average time it takes the client thread to send an entry of data (or generate a log entry) is 1.2 μs.

Test results on the Raspberry Pi platform

- In the Raspberry Pi test, the frequency of CPU is only 600 MHz. Therefore, the performance is approximately 10% of that of the server. The highest sending speed is 20,000 log entries every second.
- When the sending speed of Raspberry Pi is 20 entries every second, data sending basically does not have any impact on CPU (reduced to less than 0.01%).
- Raspberry Pi uses a USB to connect to a PC shared network. The average time it takes the client thread to send an entry of data (or generate a log entry) is about 12 μs.
6.2 Collect logs through Web Tracking

When sending an important email, you will set the read receipt tag in the email to make sure that the recipient has read the email. You will receive a reminder when the recipient reads the email.

The read receipt mode is widely used in many scenarios, such as:

- Checking whether the recipient has read the sent leaflet.
- Checking how many users have clicked a promoted webpage.
- Analyzing user access conditions on a mobile application marketing activity page.

Traditional website- and webmaster-based solutions cannot be used for custom collection and statistics, as they face the following difficulties:

- Hard to meet the individual needs: User behavior data is not generated at the mobile client. The user behavior data includes some parameters out of operations-based custom needs, such as the source, channel, environment, and behavior parameters.
- High development difficulty and cost: To meet the need of data collection and analysis, you must first purchase the cloud host, public network IP address, development data receiving server, and message-oriented middleware. In addition, you must adopt mutual backup to guarantee the high availability of the services. Then, develop the server and perform tests.
- Hard to use: After data is transmitted to the server, engineers must cleanse the results and import them to the database to generate data for operations.
- Inelastic: The usage of users cannot be predicated. Therefore, a large resource pool must be reserved.

From the preceding aspects, when a content-oriented operations need is raised, a big challenge is how to quickly meet the collection and analysis needs of such user behavior data.

*Log Service* provides Web Tracking, JS, or Tracking Pixel SDK for the preceding lightweight tracking collection scenarios. With this feature, tracking and data reporting can be completed within 1 minute. In addition, Log Service provides 500 MB *FreeTier quota* per month for each account, allowing you to process business without any cost.
Features

The collection and analysis solution is based on Alibaba Cloud Log Service, a one-stop service for log data. Log Service allows you to quickly complete the collection, consumption, shipping, query, and analysis of large amounts of log data without the need for development. This improves both the O&M efficiency and the operational efficiency. Log Service has the following features:

- **LogHub** for real-time collection and consumption. It is interconnected with Blink, Flink, Spark Streaming, Storm, and Kepler.
- **LogShipper** for data shipping. It is interconnected with MaxCompute, EMR, Object Storage Service (OSS), and Function Compute.
- **LogSearch and Analytics** for query and real-time analysis. It is interconnected with DataV, Grafana, Zipkin, and Tableau.

Advantages on the collection side

Log Service provides more than 30 data collection approaches and complete solutions for servers, mobile clients, embedded devices, and various development languages. For example:

- **Logtail**: the log collection agent for x86 servers.
- **Android or iOS SDK**: oriented to mobile clients.
- **C Producer Library**: oriented to devices with limited CPU or memory, and smart devices.
The lightweight collection solution Web Tracking in this topic only uses an HTTP GET request to transmit data to a Log Service Logstore. It is applicable to scenarios where no verification is required, such as static websites, advertising, document promotion, and mobile data collection. The following figure shows the characteristics of Web Tracking, when compared with other log collection solutions.

Web Tracking access process

The term Web Tracking (also named Tracking Pixel) is from the image tag in the HTML syntax. A 0-pixel image can be embedded on a page, which is invisible to users by default. When you access the page and the image is loaded, a GET request is initiated, transmitting parameters to the server.

To use Web Tracking, follow these steps:

1. Enable the Web Tracking feature for the Logstore.

   By default, the Logstore does not allow anonymous writing. Before using the Logstore, you must enable the Web Tracking feature for the Logstore.
2. Write data to the Logstore by tracking.

You can write data in any of the following ways:

- Use an HTTP Get request to write data.

```
curl --request GET 'http://${project}.${sls-host}/logstores/${logstore}/track? APIVersion=0.6.0&key1=val1&key2=val2'
```

- Embed an HTML image tag. Data is automatically written to the Logstore when a user visits the page.

```
<img src='http://${project}.${sls-host}/logstores/${logstore}/track.gif? APIVersion=0.6.0&key1=val1&key2=val2'/>
or
<img src='http://${project}.${sls-host}/logstores/${logstore}/track_ua.gif? APIVersion=0.6.0&key1=val1&key2=val2'/>
```

In addition to uploading custom parameters, the server also uses UserAgent and referer in the HTTP header as fields in logs.

- Use the SDK for JavaScript to write data.

```
<script type="text/javascript" src="loghub-tracking.js" async></script>

var logger = new window.Tracker('${sls-host}','${project}','${logstore}');
logger.push('customer', 'zhangsan');
logger.push('product', 'iphone 6s');
logger.push('price', 5500);
logger.logger();
```

For more information, see Web Tracking.

Case about multi-channel content promotion

Scenarios

Operational staff cannot wait to send new contents (such as new functions, activities, games, and articles) to users because this is the first and most important step to gain users.

Taking game release as an example. A great expense is put in game promotion, for example, 10,000 advertisements are invested. About 2,000 people load the advertisements, which accounts for 20% of the total number of advertisements. About 800 people click the advertisement, and fewer people finally download the game, register accounts, and try the game.
Therefore, obtaining the content promotion effectiveness accurately and in real time is important for the business. To reach the overall promotion goal, operational staff usually use various channels for promotion, for example:

- In-site messages, official website blogs, and homepage banners.
- SMS messages, emails, and leaflets.
- New media, such as Sina Weibo, DingTalk group, WeChat public account, Zhihu forum, and TouTiao.

Procedure

Step 1 Enable the Web Tracking feature

Create a Logstore (for example, myclick) in Log Service and enable the Web Tracking feature.
Step 2 Generate a Web Tracking tag

1. Add an identity for the article (article=1001) to be promoted in each promotion channel, and generate a Web Tracking tag as follows (taking the img tag as an example):

   • In-site message channel (mailDec)

   ```
   <img src="http://example.cn-hangzhou.log.aliyuncs.com/logstores/myclick/track_ua.gif?APIVersion=0.6.0&from=mailDec&article=1001" alt="" title="">
   ```

   • Official website channel (aliyunDoc)

   ```
   <img src="http://example.cn-hangzhou.log.aliyuncs.com/logstores/myclick/track_ua.gif?APIVersion=0.6.0&from=aliyundoc&article=1001" alt="" title="">
   ```

   • Email channel (email)

   ```
   <img src="http://example.cn-hangzhou.log.aliyuncs.com/logstores/myclick/track_ua.gif?APIVersion=0.6.0&from=email&article=1001" alt="" title="">
   ```

   You can add other channels after the from parameter or add more parameters to be collected in the URL.

2. Put the img tag in the promotion content and release it.

Step 3 Analyze logs

After tracking collection, you can use the LogSearch and Analytics features of Log Service to query and analyze large amounts of log data in real time. In addition to the built-in dashboard, Log Service supports the interconnection methods such as DataV, Grafana, and Tableau to achieve result analysis visualization.

The following figure displays the collected log data. You can enter a keyword in the search box to query logs.
You can also enter an SQL statement after the query to perform real-time analysis and visualization in seconds.
1. Design the query statements.

The real-time analysis statements for user clicking or reading logs are as follows. For more information about the fields and analysis scenarios, see Analysis syntax.

- Current total traffic and page views
  
  ```
  * | select count(1) as c
  ```

- Curve of the page views per hour
  
  ```
  * | select count(1) as c, date_trunc('hour',from_unixtime(__time__)) as time group by time order by time desc limit 100000
  ```

- Ratios of page views in each channel
  
  ```
  * | select count(1) as c, f group by f desc
  ```

- Devices where the page views are from
  
  ```
  * | select count_if(ua like '%Mac%') as mac, count_if(ua like '%Windows%') as win, count_if(ua like '%iPhone%') as ios, count_if(ua like '%Android%') as android
  ```

- Regions where the page views are from
  
  ```
  * | select ip_to_province(__source__) as province, count(1) as c group by province order by c desc limit 100
  ```

2. Configure the real-time data to a dashboard that is refreshed in real time.

Note:

An invisible img tag records your access to this topic. You can try to find it in the source code of the page.
6.3 Collect Container Service logs

Log Service collects Kubernetes cluster logs by using Logtail. After logs are collected, you can preview, query, and analyze logs in real time. You can also publish the query and analysis results on a dashboard through DataV.

Scenario

You need to collect logs from a Kubernetes cluster in Hangzhou by using Log Service, query and analyze the collected access logs, and then display the query and analysis results in charts through DataV.

Prerequisites

1. Your AccessKey is obtained.
2. Log Service is activated. A project and a Logstore are created.
3. DataV is activated.

Configuration process

1. **Deploy a DaemonSet for Logtail and configure a Logtail machine group.**

   Create a machine group with a custom ID in the Log Service console. Then, you do not need to worry about log collection if the Kubernetes cluster is scaled up or down in the future.

2. **Create a collection configuration for the server.**

   Create a collection configuration in the Log Service console. All the collection configurations are created for the Log Service server. No local configuration is needed.

3. **Import data to DataV.**

Step 1: Deploy a DaemonSet for Logtail

1. Connect to your Kubernetes cluster.
2. Set relevant parameters.

   a. Download the Log Service YAML file template and open it in the vi editor.

   b. Replace all `${your_xxxx}` variables in the env settings of the template with actual values.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>${your_region_name}</code></td>
<td>The region where the Log Service project you created resides. For more information about available regions, see <em>Set startup parameters</em>.</td>
</tr>
<tr>
<td></td>
<td><strong>Note:</strong> Do not confuse the hyphen (−) with the underscore ( _) in the region name.</td>
</tr>
<tr>
<td><code>${your_aliyun_user_id}</code></td>
<td>The user ID of your Alibaba Cloud account, which is of the string type. For more information about how to view the ID, see <em>Configure AliUids for ECS servers under other Alibaba Cloud accounts or on-premises IDCs</em>.</td>
</tr>
<tr>
<td><code>${your_machine_group_name}</code></td>
<td>The custom ID of the machine group in your cluster. For more information about how to set the custom ID, see <em>Create a server group with a custom identifier</em>.</td>
</tr>
</tbody>
</table>

**Note:**

A. You need to obtain an AccessKey of your Alibaba Cloud account. For more information, see *Create an AccessKey* in Quick Start.

B. **Do not modify the** `volumeMounts` **and** `volumes` **settings in the template**. Otherwise, Logtail may not work properly.
C. You can customize the startup parameter settings of the Logtail container if the following conditions are met:

A. The following three environment variables are configured before you start the Logtail container: ALIYUN_LOGTAIL_USER_DEFINED_ID, ALIYUN_LOGTAIL_USER_ID, and ALIYUN_LOGTAIL_CONFIG.

B. The domain socket of Docker is mounted to /var/run/docker.sock.

C. The root directory is mounted to the /logtail_host directory of the Logtail container if you need to collect the files of other containers or hosts.

c. Deploy a DaemonSet for Logtail.

Use the following sample code:

[root@iZu kubernetes]# curl http://logtail-release.oss-cn-hangzhou.aliyuncs.com/docker/k8s/logtail-daemonset.yaml > logtail-daemonset.yaml
[root@iZu kubernetes]# vi logtail-daemonset.yaml
...
   env:
    - name: "ALIYUN_LOGTAIL_CONFIG"
      value: "/etc/ilogtail/conf/cn_hangzhou/ilogtail_config.json"
    - name: "ALIYUN_LOGTAIL_USER_ID"
      value: "16542189653****"
    - name: "ALIYUN_LOGTAIL_USER_DEFINED_ID"
      value: "k8s-logtail"
...
[root@iZu kubernetes]# kubectl apply -f logtail-daemonset.yaml

Step 2: Configure a Logtail machine group

1. Log on to the Log Service console and click the target project.
2. On the left-side navigation submenu, click the Server Groups icon. The machine group list appears in the left-side navigation pane.
3. Click the icon next to Server Groups and select Create Machine Group.
4. In the Create Machine Group dialog box, select Custom ID from the Identifier drop-down list. Enter the value of ALIYUN_LOGTAIL_USER_DEFINED_ID set in the previous step in the Custom Identifier field.

After configuring the machine group, wait for 1 minute, and then click the machine group name to view the heartbeat status of nodes where a DaemonSet for Logtail is deployed. For more information, see View status in Manage server groups.

Step 3: Create a collection configuration for the server

1. On the Overview page of the target project, click Import Data to start the configuration process.
2. Select a data type.
   
   In the Import Data dialog box, select Docker Standard Output - Container.
3. Select a Logstore.
   
   If you start the collection configuration process by clicking the plus sign (+) next to Data Import under a Logstore, the system skips this step.
4. Create a machine group.

Before creating a machine group, ensure that Logtail is installed.

- Servers of Alibaba Group: By default, Logtail is installed on these servers. If Logtail is not installed on a server, contact Alibaba Cloud as prompted.
- Elastic Compute Service (ECS) instances: Select ECS instances and click Install. ECS instances that are running in Windows do not support one-click installation of Logtail. In this case, you need to manually install Logtail. For more information, see Install Logtail in Windows.
- On-premises servers: Install Logtail as prompted. For more information about how to install Logtail, see Install Logtail in Linux or Install Logtail in Windows based on your operating system.

After installing Logtail, click Complete Installation to create a machine group. If you have created a machine group, click Use Existing Server Groups.

5. Configure the machine group.

Select a machine group and move it from Source Server Groups to Applied Server Groups.

6. Configure the data source.

In the Specify Data Source step, enter your collection configuration. The following configuration is an example. For more information about configuration items, see Container standard output.

```json
{
  "inputs": [
    {
      "type": "service_docker_stdout",
      "detail": {
        "Stdout": true,
        "Stderr": true,
        "IncludeLabel": {
          "app": "monitor"
        },
        "ExcludeLabel": {
          "type": "pre"
        }
      }
    }
  ]
}
```

7. Configure query and analysis.

An index is created by default. You can modify it as needed.
Step 4: Import data to DataV

1. Log on to the DataV console. On the Data Sources tab, click Add Source. In the Add Data Source dialog box, enter required information such as the data type and your AccessKey, and then click OK.

2. Create an area chart and set Area Chart parameters.

   Enter the following code in the Query field:

   ```
   {
     "projectName": "sls-zc-test-hz-pub",
     "logStoreName": "nginx_access",
     "topic": "",
     "from": 1518332000,
     "to": 1518352301,
     "query": "* | select count(1) as pv, date_format(from_unixtime(__time__ - __time__%3600), '%Y/%m/%d %H:%i:%s') as time group by time order by time limit 1000",
     "line": 100,
     "offset": 0
   }
   ```

Important notes

During the test, you can enter the UNIX time for the from and to fields in the sample code, for example, 1509897600. After you publish the chart, you can set the from and to parameters in the URL of the chart to specify the start time and end time of data to be queried. For example, when you preview a chart, the URL is http://datav.aliyun.com/screen/86312. If you open the URL http://datav.aliyun.com/screen/86312?from=1510796077&to=1510798877, the chart can query and display data within the specified time range.

After configuring the chart, click the Preview and Publish icons. A dashboard is created.

6.4 Build a service to read logs from mobile apps directly

In the era of mobile Internet, it is increasingly common to upload data through mobile apps. We expect that logs in mobile apps can be directly uploaded to Log Service, instead of being transferred by an app server, so that users can focus more on their business logic development.

In normal mode, the AccessKey of the Alibaba Cloud account is required when logs are written to Log Service for authentication and anti-tampering. If a mobile app accesses Log Service in this mode, you need to save your AccessKey information.
on the mobile client, which poses a data security risk of AccessKey disclosure. Once the AccessKey is disclosed, you must upgrade the mobile app and replace the AccessKey, which is too costly. Another approach for uploading logs from mobile clients to Log Service is to use users' app servers. However, in this mode, if the number of mobile apps is large, the app servers must carry all the data on mobile clients. This mode has a high requirement on the server specification.

To avoid the preceding problems, Log Service provides a more secure and convenient scheme to collect mobile app logs. It uses RAM to build a direct data transfer service for mobile apps based on mobile services. In contrast to the scheme of directly using the AccessKey to access Log Service, you do not need to store the AccessKey on the app side in this scheme. This eliminates the risk of AccessKey disclosure. A temporary token with a lifecycle gives more safety. You can also configure more complex access control policies for the token, for example, limiting the access permission of the IP address segment. The cost of this scheme is low. You do not need many app servers because the mobile apps are directly connected to the cloud platform and only the control flow is sent to the app server.

You can create a RAM role for operating Log Service and configure an app as a RAM user to assume this role, so that you can build a Log Service-based direct log transfer service for mobile apps within 30 minutes. Direct data transfer is a service that allows mobile apps to directly access Log Service, with only the control flow sent to the app server.

Advantages

Using RAM to build a Log Service-based direct data transfer service for mobile apps has the following advantages:

- This access mode is more secure, providing flexible and temporary permission assignment and authentication.
- Fewer app servers are required, reducing the cost. The mobile apps are directly connected to the cloud platform and only the control flow is sent to the app server.
- Log Service supports high concurrency. A large number of users can use the service at the same time. Large upload and download bandwidths are provided.
- Log Service supports auto scaling, providing unlimited storage space.

The following figure shows the architecture.
### Description

<table>
<thead>
<tr>
<th>Module</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Android or iOS mobile app</td>
<td>The app on the mobile phone of the end user and the log source.</td>
</tr>
<tr>
<td>LOG</td>
<td>Alibaba Cloud Log Service, responsible for storing log data uploaded by the app.</td>
</tr>
<tr>
<td>RAM or STS</td>
<td>Alibaba Cloud RAM, responsible for providing the user identity management and resource access control services, and generating temporary access credentials.</td>
</tr>
<tr>
<td>App server</td>
<td>The app background service developed for the Android or iOS apps. It manages tokens used by the app for data upload and download, and metadata information of data uploaded by users in the app.</td>
</tr>
</tbody>
</table>

### Configuration process

1. The app applies for a temporary access credential from the app server.

   To avoid the risk of information leakage, the Android or iOS app does not store the AccessKey ID and AccessKey secret. Therefore, the app must request a temporary upload credential (a token) from the app server. The token is only
valid for a certain period. For example, if a token is set to be valid for 30 minutes (which can be specified by the app server), the Android or iOS app can use this token to access Log Service within the next 30 minutes. However, 30 minutes later, the app must request a new token.

2. The app server checks the validity of the request and then returns a token to the app.

3. After obtaining the token, the mobile app can access Log Service.

This topic describes how the app server applies for the token from the RAM service and how the Android or iOS app obtains the token.

Procedure

1. Authorize a RAM role to operate Log Service.

   Create a RAM role for operating Log Service and configure an app as a RAM user to assume this role. For more information, see #unique_257.

   After the configuration is complete, you can obtain the following information:
   • AccessKey ID and AccessKey secret of the RAM user
   • Resource path of the role

2. Build an app server.

   This topic provides sample programs in multiple languages. The download URLs are listed at the end of this topic.

   Each downloaded language package contains the configuration file config.json, which contains the following information:

   ```json
   {
     "AccessKeyId" : "",
     "AccessKeySecret" : "",
     "RoleArn" : "",
     "TokenExpireTime" : "900",
     "PolicyFile": "policy/write_policy.txt"
   }
   ```

   Note:

   1. AccessKeyId: the ID of your AccessKey.
   3. RoleArn: the resource path of the role.
4. TokenExpireTime: the expiration time of the token obtained by the Android or iOS app. The minimum value is 900, in seconds. The default value can be left unchanged.

5. PolicyFile: The file that lists the permissions that the token can grant. The default value does not need to be changed.

This topic describes two token files that define the most common permissions in the policy directory.

- write_policy.txt: specifies a token that grants the write permission for the project to this user.
- readonly_policy.txt: specifies a token that grants the read permission for the project to this user.

You can also design your policy file as required.

Response format

// Correct response
{
  "StatusCode":200,
  "AccessKeyId":"STS.3p***dgagdasdg",
  "AccessKeySecret":"rpnwO9***tGdrddgsR2YrTtI",
  "SecurityToken":"CAES+wMIARKAAZhjH0EUOihJMQBmjRywXq7MQ/cjLYg80Ah1ek0Jm6XMhr9oc5s`ä`3qaPer8p1YaXINTDiCFZWFkv1Hf1pQhuxfKBc+mRR9KAhHuEfqH+rdjZqjTF7p2m1wXJP86Gqk+G2MpHrUe6TYBkJ43GhhTVFMuM3BZaj3VjZW0X8I0QRI1FKZjIiEjMzE0MjY0NzgM5MTE4NjkxMSoLY2xpZGSSDgSDGAGESGTETqOi6c2RrLWRlbW8vKgoUYWNzOm9zczqOqo6c2RrLWRlbW9KEDExNDg5M2AzMDcyNDY4MTShSBT120DQywag9bc3N1bWVUUm9sZVVzZXJgAGoSMzMTQyNjQ3MzxxMTg2OTEcglzzZGstZGVtbzI=",
  "Expiration":"2017-11-12T07:49:09Z",
}

// Error response
{
  "StatusCode":500,
  "ErrorCode":"InvalidAccessKeyId.NotFound",
  "ErrorMessage":"Specified access key is not found."
}

Correct response description: The following table describes the five variables that constitute a token.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>StatusCode</td>
<td>The result returned when the app retrieves the token. The app returns 200 if the token is successfully retrieved.</td>
</tr>
<tr>
<td>Variable</td>
<td>Description</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>AccessKeyId</td>
<td>The AccessKey ID that the Android or iOS app obtains when initializing LogClient.</td>
</tr>
<tr>
<td>AccessKeySecret</td>
<td>The AccessKey secret that the Android or iOS app obtains when initializing LogClient.</td>
</tr>
<tr>
<td>SecurityToken</td>
<td>The token that the Android or iOS app uses to access Log Service.</td>
</tr>
<tr>
<td>Expiration</td>
<td>The time period before the token expires. The Android SDK automatically determines the validity of the token and then retrieves a new one as needed.</td>
</tr>
</tbody>
</table>

Error response description:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>StatusCode</td>
<td>The result returned when the app retrieves the token. The app returns 500 if the token fails to be retrieved.</td>
</tr>
<tr>
<td>ErrorCode</td>
<td>The error cause.</td>
</tr>
<tr>
<td>ErrorMessage</td>
<td>The error description.</td>
</tr>
</tbody>
</table>

Running method of the sample code:

For Java 1.7 or later, after downloading and decompressing the package, create a Java project, copy the dependency, code, and configuration to the project, and run the main function. By default, the program listens on port 7080 and waits for the HTTP request. Perform the operations in other languages in a similar way.

3. The mobile client constructs an HTTP request to obtain a token from the app server.

The formats of HTTP request and response are as follows:

Request URL: GET https://localhost:7080/

Response:

```json
{
  "StatusCode":"200",
  "AccessKeyId":"STS.XXXXXXXXXXXXXXXXX",
  "AccessKeySecret":"",
  "SecurityToken":"",
  "Expiration":"2017-11-20T08:23:15Z"
}
```
Note:
All examples in this topic are used to demonstrate how to set up a server. You can implement custom development based on these examples.

Source code download

Sample code of the app server: PHP, Java, Ruby, Node.js.

6.5 Collect data over the public network

In some scenarios, you may need to collect data from clients, such as mobile clients, HTML webpages, PCs, servers, hardware devices, and cameras, over the public network for real-time processing.

In a traditional architecture, the preceding feature can be achieved by integrating front-end servers with Kafka. Now, you can use Log Service LogHub to replace such architecture with a solution that is more reliable, cost-effective, elastic, and secure.

Scenarios

Data can be collected from clients, such as mobile clients, external servers, and webpages, over the public network. After data is collected, data applications, such as real-time computing and data warehouse, are required to store or consume data.
Solution 1: Integrate front-end servers with Kafka

Kafka does not support the RESTful protocol and is more commonly used in clusters. Therefore, you need to set up an NGINX server as a public network proxy, and then use LogStash or API to write data to Kafka through NGINX.

The following table lists the required devices.

<table>
<thead>
<tr>
<th>Device</th>
<th>Quantity</th>
<th>Configuration</th>
<th>Purpose</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECS instance</td>
<td>2</td>
<td>Single-core CPU, 2 GB memory</td>
<td>Front-end server, load balancing, and mutual backup</td>
<td>RMB 108 per unit per month</td>
</tr>
<tr>
<td>Server Load Balancer</td>
<td>1</td>
<td>Standard</td>
<td>Pay-as-you-go instance</td>
<td>RMB 14.4 per month (lease) + RMB 0.8 per GB (data traffic)</td>
</tr>
<tr>
<td>Kafka or ZooKeeper</td>
<td>3</td>
<td>Single-core CPU, 2 GB memory</td>
<td>Data writing and processing</td>
<td>RMB 108 per unit per month</td>
</tr>
</tbody>
</table>

Solution 2: Use LogHub

You can use mobile SDKs, Logtail, or Web Tracking JS to write data to the LogHub endpoint.

The following table lists the required devices.

<table>
<thead>
<tr>
<th>Device</th>
<th>Purpose</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>LogHub</td>
<td>Real-time data collection</td>
<td>Less than RMB 0.2 per GB. Click here to check the billing method.</td>
</tr>
</tbody>
</table>

Scenario comparison

Scenario 1: Up to 10 GB of data is collected each day, generating about 1 million write requests. The 10 GB in this example is the compressed size. The actual data volume ranges from 50 GB to 100 GB.

Solution 1:

---
Server Load Balancer (lease): 0.02 × 24 × 30 = RMB 14.4
Server Load Balancer (data traffic): 10 × 0.8 × 30 = RMB 240
ECS cost: 108 × 2 = RMB 216
Kafka ECS: free if shared with other services  
Total: RMB 470.4 per month

Solution 2:
--------------
LogHub traffic: $10 \times 0.2 \times 30 = \text{RMB 60}$  
Number of LogHub requests: $0.12$ (assuming 1 million requests per day)  
$\times 30 = \text{RMB 3.6}$  
Total: RMB 63.6 per month

Scenario 2: Up to 1 TB of data is collected each day, generating about 100 million write requests.

Solution 1:
--------------
Server Load Balancer (lease): $0.02 \times 24 \times 30 = \text{RMB 14.4}$  
Server Load Balancer (data traffic): $1,000 \times 0.8 \times 30 = \text{RMB 24,000}$  
ECS cost: $108 \times 2 = \text{RMB 216}$  
Kafka ECS: free if shared with other services  
Total: RMB 24,230.4 per month

Solution 2:
--------------
LogHub traffic: $1,000 \times 0.15 \times 30 = \text{RMB 4,500}$ (tiered pricing)  
Number of LogHub requests: $0.12 \times 100$ (assuming 100 million requests per day)  
$\times 30 = \text{RMB 360}$  
Total: RMB 4,860 per month

Solution comparison

The comparison between the preceding scenarios shows that you can use LogHub to collect data from the public network at a competitive cost. Furthermore, Solution 2 outperforms Solution 1 in the following aspects:

- Auto scaling: LogHub supports collecting MBs or even PBs of data each day.
- Abundant permission control options: You can use Access Control List (ACL) to control the read and write permissions.
- HTTPS compatibility: Data is encrypted before transmission.
- Log shipping at no cost: No additional development is required for shipping logs to data warehouses.
- Detailed data monitoring: You can gain more information about your business.
- Abundant SDKs for connecting to upstream and downstream systems: Like Kafka, LogHub provides abundant SDKs to connect to downstream systems. It can be deeply integrated with Alibaba Cloud and open-source products.

For information, see the product page of Log Service.
6.6 Collect data from multiple channels

Log Service LogHub supports real-time data collection and consumption. It supports more than 30 real-time collection approaches.

Data is usually collected in two different modes as described in the following table. This topic describes how to use the streaming import feature of LogHub to collect data in real time.

<table>
<thead>
<tr>
<th>Mode</th>
<th>Advantage</th>
<th>Disadvantage</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Batch import</td>
<td>High throughput, focusing on historical data</td>
<td>Poor real-time performance</td>
<td>FTP-based import, uploading from OSS, mailing hard drives, and SQL-based data export to LogHub</td>
</tr>
<tr>
<td>Streaming import</td>
<td>Real-time, what you see is what you get (WYSIWYG), focusing on real-time data</td>
<td>High requirements on the collection terminals</td>
<td>LogHub-based data collection, HTTP-based uploading, Internet of Things (IoT) data collection, and Queue</td>
</tr>
</tbody>
</table>

Background

"I Want Take-away" is an e-commerce website with a platform involving users, restaurants, and couriers. Users can place their take-away orders on the website, app, WeChat, or Alipay. Once receiving an order from a user, a merchant starts preparing the ordered dishes. At the same time, the system automatically notifies the nearest couriers. Then, one of the couriers accepts the order and delivers the dishes to the user.
Operation requirements

During operations, the following issues are identified:

- It is difficult to attract new users. In some other cases, a hefty advertising investment in various marketing channels, such as webpage ads and WeChat push messages, attracts some new users. However, it is difficult to evaluate the effectiveness of each channel.
- Users often complain about the slow delivery. However, it is difficult to locate the delayed phase, for example, order taking, delivery, or preparing the order. In addition, "I Want Take-away" has no clue how to improve the situation.
- User operations teams often organize promotions, such as giving away coupons, but cannot get expected results.
- In terms of scheduling, "I Want Take-away" wants to know how to facilitate merchants to stock up food for peak hours and dispatch more couriers to a specific area.
- As for the customer service, "I Want Take-away" wants to analyze the reason why some users fail to place orders, for example, whether any inappropriate actions were performed by the users before they report the failure or whether any system errors occur.
Data collection challenges

In digital operations, the first step is to figure out how to centrally collect the distributed log data. The following challenges are encountered during the process:

- Multiple channels: advertisers and street promotions (leaflets)
- Multiple terminals: webpages, official social media accounts, mobile phones, desktop browsers, and mobile browsers
- Heterogeneous networks: Virtual Private Cloud (VPC), user-created Internet data center (IDC), and Alibaba Cloud Elastic Compute Service (ECS)
- Various development languages: Java for the core system, NGINX for front-end servers, and C++ for the back-end payment system
- Devices: merchants' devices adopting different architectures, such as x86 and ARM

In the past, you need to complete a large amount of diversified work to collect the logs distributed externally and internally for unified management. But now, you can use the collection feature of LogHub for unified access.

Unified log management and configuration

1. Create a log management project, for example, myorder.
2. Create Logstores for logs generated from different data sources, for example:

- wechat-server (for storing WeChat server access logs)
- wechat-app (for storing WeChat server application logs)
- wechat-error (for storing WeChat error logs)
- alipay-server
- alipay-app
- deliver-app (for storing logs related to the courier app status)
- deliver-error (for storing delivery error logs)
- web-click (for storing HTML5 webpage click logs)
- server-access (for storing server-side access logs)
- server-app (for storing application logs)
- coupon (for storing application coupon logs)
- pay (for storing payment logs)
- order (for storing order logs)

3. To cleanse raw data and run Extract-Transform-Load (ETL) jobs on the raw data, create intermediate Logstores.

User promotion log collection

The following two approaches are commonly used to attract new users:

- Offer coupons upon website registration.
- Offer coupons for scanning QR codes on other channels, such as:
  - QR codes on leaflets
  - QR codes on webpages

Implementation

Define the following registration server link and generate QR codes for leaflets and webpages. When a user scans the QR code on a webpage for registration, you can identify the user as being from a specific source, and create logs accordingly.

```
http://examplewebsite/login?source=10012&ref=kd4b
```

When receiving a request, the server generates the following logs:

```
2016-06-20 19:00:00 e41234ab342ef034,102345,5k4d,467890
```
The logs contain the following parameters:

- **time**: the time of registration.
- **session**: the current session of the browser. It is used for behavior tracking.
- **source**: the source channels. For example, Campaign A is labeled as 10001, leaflets 10002, and elevator advertisements 10003.
- **ref**: the account of someone who recommends the user to sign up. If no one recommends the user to sign up, leave this parameter blank.
- **params**: other parameters.

**Collection methods:**

- The application generates logs to hard disks by using *Logtail*.
- The application writes data to Log Service by using the *SDK*.

**Server-side data collection**

Alipay or WeChat official account programming is a typical web-side model that generally utilizes the following four types of logs:

- **NGINX or Apache access logs**: used for monitoring and real-time statistics.

```plaintext
10.1.168.193 -- [01/Mar/2012:16:12:07 +0800] "GET /Send?AccessKeyId=8225105404 HTTP/1.1" 200 5 "-" "Mozilla/5.0 (X11; Linux i686 on x86_64; rv:10.0.2) Gecko/20100101 Firefox/10.0.2"
```

- **NGINX or Apache error logs**

```plaintext
2016/04/18 18:59:01 [error] 26671#0: *20949999 connect() to unix:/tmp/fastcgi.socket failed (111: Connection refused) while connecting to upstream, client: 10.101.1.1, server: , request: "POST /logstores/test_log HTTP/1.1", upstream: "fastcgi://unix:/tmp/fastcgi.socket:", host: "ali-tianchi-log.cn-hangzhou-devcommon-intranet.sls.aliyuncs.com"
```

- **Application-layer logs**: capture the event details, such as the time, location, result, delay, method, and parameters. This type of log usually ends with extended parameters.

```json
{
    "time": "2016-08-31 14:00:04",
    "localAddress": "10.178.93.88:0",
    "methodName": "load",
    "param": ["31851502"],
    "result": "...
    "serviceName": "com.example",
    "startTime": 1472623203994,
}
```
Application-layer error logs: record the error details, such as the time, code line, error code, and reason.


Implementation

- To write logs to a local file, specify regular expressions in the Logtail Config to write the logs to the specified Logstore.
- To collect logs generated in a Docker container, integrate Container Service with Log Service.
- To collect logs for Java programs, use Log4J Appender (without saving logs to hard disks) or LogHub Producer Library (for high-concurrency client-side write).
- To collect logs for C#, Python, Java, PHP, and C programs, use the corresponding SDKs.

Access to end user logs

- Mobile clients: Use the SDK for iOS or Android, or Mobile Analytics (MAN) for log access.
- ARM devices: Use the native C code for cross-compiling.
- Merchant platform devices: Use the corresponding SDK on x86 servers for log access. Use the native C code on ARM devices for cross-compiling.

Collection of user behavior data for desktop or mobile websites

The user behavior can be divided into the following two types:

- User behavior that has interaction with the back-end server, such as placing an order, logging on, and logging out.
- User behavior that has no interaction with the back-end server, including sending requests that are processed directly at the front end, such as scrolling or closing a page.
Implementation

- To collect data of user behavior that has interaction with the back-end server, refer to the implementation approaches for server-side data collection.
- To collect data of user behavior that has no interaction with the back-end server, use Tracking Pixel or JS Library.

Server log O&M

Example:

- **System logs**
  
  Aug 31 11:07:24 zhouqi-mac WeChat[9676]: setupHotkeyListenning event
  NSEvent: type=KeyDown loc=(0,703) time=115959.8 flags=0 win=0x0
  winNum=7041 ctxt=0x0 chars="u" unmodchars="u" repeat=0 keyCode=32

- **Application debug logs**
  
  __FILE__:build/release64/sls/shennong_worker/ShardDataIndexManager.cpp
  __LEVEL__:WARNING
  __LINE__:238
  __THREAD__:31502
  offset:816103453552
  saved_cursor:146978055385742676
  seek count:62900
  seek data redo
  log:pangu://localcluster/redo_data/41/example/2016_08_30/250_1472555483
  user_cursor:146978055385885689973

- **Trace logs**
  
  [2013-07-13 10:28:12.772518] [DEBUG] [26064] __TRACE_ID__:661353951201 __item__:Class:Function]_end__ request_id:1734117
  user_id:124 context:.....

Implementation

For more information, see the implementation approaches for server-side data collection.

Data collection in different network environments

LogHub provides access points in each region with the following three access approaches:
6.7 Manage logs

The LogHub feature of Log Service allows you to use Logtail to collect logs.

Log management scenario

A server or container stores a large amount of log data generated by applications in different directories.

- Developers publish or unpublish applications.
- The server can scale out during peak hours and scale in during off-peak hours.
- The log data is queried, monitored, and warehoused based on changing requirements.
Challenges

- Fast application publishing and increasing log types

Each application generates access, operations, logic, and error logs. When new applications are associated with each other or with existing applications, the volume of logs explodes.

The following table lists the different types of logs collected for a takeout website.

<table>
<thead>
<tr>
<th>Type</th>
<th>Application</th>
<th>Log name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web</td>
<td>NGINX</td>
<td>wechat-nginx, which stores WeChat server NGINX logs</td>
</tr>
<tr>
<td>Web error</td>
<td>NGINX</td>
<td>alipay-nginx, which stores Alipay server NGINX logs</td>
</tr>
<tr>
<td></td>
<td>NGINX</td>
<td>server-access, which stores server-side access logs</td>
</tr>
<tr>
<td></td>
<td>NGINX error</td>
<td>alipay-nginx, which stores NGINX error logs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>...</td>
</tr>
<tr>
<td>Web app</td>
<td>Tomcat</td>
<td>alipay-app, which stores Alipay server application logic</td>
</tr>
<tr>
<td></td>
<td></td>
<td>...</td>
</tr>
<tr>
<td>App</td>
<td>Mobile app</td>
<td>deliver-app, which stores status logs of the courier app</td>
</tr>
<tr>
<td>App error</td>
<td>Mobile app</td>
<td>deliver-error, which stores delivery error logs</td>
</tr>
<tr>
<td>Web</td>
<td>HTML5</td>
<td>web-click, which stores HTML5 page view (PV) logs</td>
</tr>
<tr>
<td>Server</td>
<td>Server</td>
<td>Internal logic logs on the server side</td>
</tr>
<tr>
<td>Syslog</td>
<td>Server</td>
<td>Server system logs</td>
</tr>
</tbody>
</table>
• Log consumption for different purposes

For example, access logs can be downloaded for metering and billing. Operations logs can be queried by a database administrator (DBA). These logs require business intelligence (BI) analysis and end-to-end monitoring.

• Business and environment changes

With the rapid development of the Internet, you need to adapt to continuous business and environment changes:
- Application server scale-out
- Servers as machines
- New application deployment
- New log consumers

Ideal management architecture

An ideal management architecture requires:
- A well-defined, low-cost framework
- A stable, reliable, and unattended mechanism, for example, a mechanism that allows you to scale in or out servers as needed
- Standardized application deployment without complicated configurations
- High performance that meets log processing requirements

Log Service solution

Log Service LogHub uses Logtail to collect logs. This process involves the following components:
- Project: a management container.
- Logstore: the source of a type of logs.
- Machine group: the directory and format of logs.
- Configuration: the path to the log source.

The relationships between these components are as follows:
- A project includes multiple Logstores, machine groups, and configurations. Different projects can be used to meet different business requirements.
- Each application can have multiple types of logs. Each type of log has a Logstore and a fixed directory (with the same configuration).

app --> logstore1, logstore2, logstore3
app --&gt; config1, config2, config3

- A single application can be deployed on multiple machine groups. Multiple applications can be deployed on a single machine group.

app --&gt; machineGroup1, machineGroup2
machineGroup1 --&gt; app1, app2, app3

- The collection directories defined in the Logtail configurations are applied to the specified machine groups, and logs are collected into any Logstore.

config1 * machineGroup1 --&gt; Logstore1
config1 * machineGroup2 --&gt; Logstore1
config2 * machineGroup1 --&gt; Logstore2

Benefits

- **High efficiency:** The Web Console or SDK is provided for batch management.
- **Large scale:** Millions of machines and applications can be managed.
- **Real-time response:** Collection configurations take effect within minutes.
- **High elasticity**
  - The machine identification feature supports auto scaling of servers.
  - LogHub supports auto scaling.
- **High stability and reliability:** No human intervention is required.
- **Abundant query capabilities in log processing, such as real-time computing, offline analysis, and indexing:**
  - LogHub: real-time collection and consumption. LogHub uses more than 30 methods to collect massive data for real-time downstream consumption.
  - LogShipper: stable and reliable log shipping. LogShipper ships log data from LogHub to Object Storage Service (OSS), MaxCompute, or Table Store for storage and big data analysis.
  - LogSearch: real-time data indexing and querying. LogSearch allows you to query logs in a centralized manner no matter where active server logs are located.
6.8 Use logging handlers to automatically upload Python logs

This topic describes how to use logging handlers to automatically upload Python logs.

Benefits

Logging handlers provided by Python SDK can automatically upload logs of Python applications to Log Service without storing the logs on a local disk. Compared with other tools that store logs in files before uploading them, logging handlers have the following benefits:

- Logs are uploaded in real time without being locally stored.
- Massive volumes of logs can be uploaded in an asynchronous manner.
- Logging handlers are easy to configure, requiring no modification of code. The configurations immediately take effect after you modify the configuration file. The physical location of the host is not required for uploading logs.
- JSON formatted log information and key-value formatted log information are automatically parsed.

Configurations

Logging handlers are compatible with the Python logging module. For more information about the Python logging module, see *Python Logging*.

The Python logging module allows you to configure logging by using a script or using a configuration file. The following example describes how to configure logging by using the *logging.conf* file.

```python
[loggers]
keys=root,sls

[handlers]
keys=consoleHandler, slsHandler

[formatters]
keys=simpleFormatter, rawFormatter

[logger_root]
level=DEBUG
handlers=consoleHandler

[logger_sls]
level=INFO
handlers=consoleHandler, slsHandler
qualname=sls
propagate=0

[handler_consoleHandler]
```
A root logging handler and a sls logging handler are configured. The sls logging handler is an object that is created from the `aliyun.log.QueuedLogHandler` class.

The following shows the parameters that are passed for the configurations of the sls logging handler. For more information about the parameters, see the parameter list.

```python
args=(os.environ.get('ALIYUN_LOG_SAMPLE_ENDPOINT', ''), os.environ.get('ALIYUN_LOG_SAMPLE_ACCESSID', ''), os.environ.get('ALIYUN_LOG_SAMPLE_ACCESSKEY', ''), os.environ.get('ALIYUN_LOG_SAMPLE_TMP_PROJECT', ''), "logstore")
```

Note:

`os.environ` is used to obtain the values of environment variables. You can also enter actual parameter values.

Upload logs

You can use the logging configuration file to print logs to Log Service.

```python
import logging
import logging.config

# Configurations
logging.config.fileConfig('logging.conf')
logger = logging.getLogger('sls')

# Use logger
logger.info("test1")

try:
    1/0
except ZeroDivisionError as ex:
Then, logs are automatically uploaded to Log Service. To use the search and analytics feature, you must enable the index of the corresponding Logstore.

Configure the index for a Logstore

Enable the index of the Logstore that receives logs and configure the index for specific fields. We recommend that you use `#unique_263` to configure the index as follows:

```bash
aliyunlog log_update_index --project_name="project1" --logstore_name="logstore1" --index_detail="file:///Users/user1/loghandler_index.json"
```

For more information about the configuration file, see `python_logging_handler_index.json`.

Change the log fields to be collected

The following table shows the log fields that Log Service support. By default, all of these fields are collected.

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>message</td>
<td>The content of the log.</td>
</tr>
<tr>
<td>record_name</td>
<td>The name of the logging handler. In the preceding example, the record_name field is <code>sls</code>.</td>
</tr>
<tr>
<td>level</td>
<td>The level of the log, such as INFO and ERROR.</td>
</tr>
<tr>
<td>file_path</td>
<td>The full path of the configuration file.</td>
</tr>
<tr>
<td>func_name</td>
<td>The name of the function that is used to print the log.</td>
</tr>
<tr>
<td>line_no</td>
<td>The number of the code line where the function is called.</td>
</tr>
<tr>
<td>module</td>
<td>The module where the function resides.</td>
</tr>
<tr>
<td>thread_id</td>
<td>The ID of the current thread.</td>
</tr>
<tr>
<td>thread_name</td>
<td>The name of the current thread.</td>
</tr>
<tr>
<td>process_id</td>
<td>The ID of the current process.</td>
</tr>
<tr>
<td>process_name</td>
<td>The name of the current process.</td>
</tr>
</tbody>
</table>

You can choose the fields to be collected based on the `fields` parameter of `AliyunLogHandler`. For more information, see `aliyun.log.LogFields`.
In the following example, the preceding configuration file is modified to only collect the level, func_name, module, and line_no fields.

```
[handler_slsHandler]
class=aliyun.log.QueuedLogHandler
level=INFO
formatter=rawFormatter
args=('cn-beijing.log.aliyuncs.com', 'ak_id', 'ak_key', 'project1', "logstore1", 'mytopic', ['level', 'func_name', 'module', 'line_no'] )
```

**Note:**

- The message field is collected regardless of your configurations.
- You can also add a prefix and suffix to the names of these fields by using the `buildin_fields_prefix` and `buildin_fields_suffix` parameters. An example of a prefix or suffix is `__level__`.

Configure logging by using a JSON script:

```python
#encoding: utf8
import logging, logging.config, os

# Configurations
conf = {'version': 1,
    'formatters': {'rawformatter': {'class': 'logging.Formatter',
                                    'format': '%(message)s'}},
    'handlers': {'sls_handler': {'()':
                                'aliyun.log.QueuedLogHandler',
                                'level': 'INFO',
                                'formatter': 'rawformatter',
                                # custom args:
                                'end_point': os.environ.get('ALIYUN_LOG_SAMPLE_ENDPOINT', ''),
                                'access_key_id': os.environ.get('ALIYUN_LOG_SAMPLE_ACCESSID', ''),
                                'access_key': os.environ.get('ALIYUN_LOG_SAMPLE_ACCESSKEY', ''),
                                'project': 'project1',
                                'log_store': "logstore1"
                               }
               },
    'loggers': {'sls': {'handlers': ['sls_handler', ],
                        'level': 'INFO',
                        'propagate': False}}
}
logging.config.dictConfig(conf)

# Use the logger
logger = logging.getLogger('sls')
```
logger.info("Hello world")

Note:
QueuedLogHandler is initialized by passing named parameters. For more information about parameters, see the parameters.

For more information about how to use dicConfig in Python, see dictConfig.

Use Python logging handlers for uWSGI logs

QueuedLogHandler cannot perform as expected due to the scheduling of uWSGI processes. The following two logging handlers are provided as alternatives to upload uWSGI logs.

- UwsgiQueuedLogHandler. It has the same functions and configurations as the QueuedLogHandler class. To use this class, you must install the third-party library uwsgidecorators.

- SimpleLogHandler. It can upload logs in real time and has the same configurations as the QueuedLogHandler class. We recommend that you only use this class for testing purposes.
7 FAQ

7.1 Troubleshoot collection errors

If the log collection fails or the collection status is abnormal when you use Logtail, follow these steps to troubleshoot the errors.

Procedure

1. Check whether the Logtail heartbeat in the machine group is normal

Log on to the Log Service console and click Machine Status to view the status of the machine group. For more information, see Manage a machine group. If the heartbeat status is normal, move to the next step.

   If the heartbeat status is fail, see Logtail heartbeat error for troubleshooting.

2. Check whether the collection configuration is created and applied to the machine group

After you confirm that the Logtail client status is normal, check the following configurations.

   a) Check whether Logtail configuration is created

      For more information, see Logtail configuration. Make sure that the log monitoring directory and the log file name match with the files on the machine. The directory does not support fuzzy match and must be set to an absolute path, while the log file name supports fuzzy match.

   b) Check whether Logtail configuration is applied to the machine group

      See Manage configurations in Manage a machine group. Check if the target configuration is applied to the machine group.

3. Check for collection errors

   If Logtail is properly configured, check whether new data is generated in real time in the log file. Logtail collects incremental data only, it does not read inventory files if the files are not updated. If the log file is updated but the
updates cannot be queried in Log Service, diagnose the problem in the following ways:

- Diagnose collection errors
  
  See *Diagnose collection errors* to handle the errors according to the error type reported by Logtail.

- View Logtail logs
  
  Client logs include key INFO logs and all the WARNING and ERROR logs. To see complete and real-time errors, view the client logs in the following paths:
  
  - **Linux**: `/usr/local/ilogtail/ilogtail.LOG`
  - **Linux**: `/usr/local/ilogtail/logtail_plugin.LOG` (logs of input sources such as HTTP, MySQL binlog, and MySQL query results)
  - **Windows x64**: `C:\Program Files (x86)\Alibaba\Logtail\logtail_*.log`
  - **Windows x32**: `C:\Program Files\Alibaba\Logtail\logtail_*.log`

- Usage exceeds the limit
  
  - To collect large volumes of logs, files, or data, you can modify the Logtail startup parameters for higher log collection throughput. For more information, see *Set startup parameters*.

  If the problem persists, open a ticket to contact Log Service engineers and attach the key information collected during troubleshooting to the ticket.

### 7.2 What can I do if the Logtail client has no heartbeat?

If the heartbeat status of the Logtail machine group is abnormal when you use Logtail to collect logs, you can troubleshoot the error manually or by using the Logtail automatic diagnostic tool.

After you install Logtail on your server to collect logs, the Logtail client regularly sends heartbeat packets to the server. If the machine group status shows that the Logtail client has no heartbeat, the Logtail client is disconnected from the server. In this case, you can troubleshoot the error manually or by using the *Logtail automatic diagnostic tool* as required.
- Automatic diagnosis: Log Service provides the Logtail automatic diagnostic tool for Linux. For more information, see How do I use the Logtail automatic diagnostic tool?

- Manual diagnosis: If the Logtail diagnostic tool fails to troubleshoot the error or your server is running in Windows, troubleshoot the error by performing the following steps.

1. Check whether Logtail is installed

Run the following command to check whether Logtail is installed. If you do not install Logtail, see Install Logtail in Linux or Install Logtail in Windows to install it. You must install Logtail based on the region of your Log Service project and the network type.

Check the Logtail installation status:
If `ilogtaild` is running is returned, Logtail is installed, as shown in the following example:

```
[root@localhost ~]# sudo /etc/init.d/ilogtaild status
ilogtail is running
```

**Windows:**

1. On Control Panel, choose System and Security > Administrative Tools, and then double-click Services.
2. Check the running status of LogtailDaemon and LogtailWorker. If they are running normally, Logtail is installed.

If Logtail is running, go to the next step.

2. Check whether the Logtail installation parameter is correct

Before installing Logtail, you must to specify the correct network endpoint. That is, you must select a correct Logtail installation parameter from `Table 3-3: Region names for Logtail installation` based on the region of the Log Service project and decide how to install Logtail based on the network type. If the installation script or parameter is incorrect, the Logtail client may have no heartbeat.

The Logtail configuration file `ilogtail_config.json` records the Logtail installation parameter and the installation method that you used. This file is stored in:

- **Linux:** `/usr/local/ilogtail/ilogtail_config.json`
- **Windows x64:** `C:\Program Files (x86)\Alibaba\Logtail\ilogtail_config.json`
- **Windows x32:** `C:\Program Files\Alibaba\Logtail\ilogtail_config.json`
1. Check the installation parameter.

Check whether the region of the network endpoint recorded in the `ilogtail_config.json` file is the same as that of your Log Service project.

For example, the returned information in the following figure indicates that Logtail is installed on an ECS instance in China (Hangzhou).

![Figure 7-1: Check the installation parameter](image)

2. Check the installation method.

Telnet the domain name configured in the `ilogtail_config.json` file to check whether Logtail is properly installed based on the network type of the server.

For example, the domain name recorded in the `ilogtail_config.json` file is `cn-hangzhou-intranet`. You can run the `telnet logtail.cn-hangzhou-intranet.log.aliyuncs.com 80` command to check the network connectivity. If Logtail is connected to the server, Logtail is properly installed.

For example, run the following command to check whether Logtail is connected to an ECS instance that is running in Linux:

```
[root@*********** ~]# telnet logtail.cn-hangzhou-intranet.log.aliyuncs.com 80
Trying 100*0*7*5...
Connected to logtail.cn-hangzhou-intranet.log.aliyuncs.com.
```
If the telnet command fails, the Logtail installation parameter is incorrect, so that the installation command is incorrect. For more information about how to select a correct installation parameter, see Install Logtail in Linux or Install Logtail in Windows.

If Logtail is properly installed, go to the next step.

3. Check whether the IP address configuration of the machine group is correct.

The server IP address obtained by the Logtail client must be configured in the machine group. Otherwise, the Logtail client has no heartbeat in the machine group or Logtail cannot collect logs.

The Logtail client obtains the server IP address as follows:

- If no hostname is bound, the Logtail client obtains the IP address of the first NIC of the server.
- If a hostname is bound in the /etc/hosts file, the Logtail client obtains the bound IP address. You can run the hostname command to view the hostname.

Troubleshooting procedure

1. Check the server IP address obtained by the Logtail client.

The ip field in the app_info.json file indicates the server IP address obtained by the Logtail client. This file is stored in:

- Linux: /usr/local/ilogtail/app_info.json
- Windows x64: C:\Program Files (x86)\Alibaba\Logtail\app_info.json
- Windows x32: C:\Program Files\Alibaba\Logtail\app_info.json

Note:

- Logtail cannot work if the ip field in the app_info.json file is empty. In this case, you must to configure an IP address for the server and restart Logtail.
The `app_info.json` file only records information. Any modification to this file does not change the server IP address obtained by the Logtail client.

Figure 7-2: Check the server IP address obtained by the Logtail client

```
[root@iZbp1f3ce8nd9q2l7bd42 ~]# cat /usr/local/logtail/app_info.json
{
    "UUID": "D75A4533-44B6-46C8-B071-614BC7A19685",
    "hostname": "iZbp1f3ce8nd9q2l7bd42",
    "instance_id": "AE9FDA16-B279-11E8-A011-80163E0E5573_192.108.35.4_1536309632",
    "ip": "192.168.35.4",
    "logtail_version": "6.16.19",
    "os": "linux; 3.10.0-693.2.2.el7.x86_64; #1 SMP Tue Sep 12 22:26:13 UTC 2017; x86_64",
    "update_time": "2018-09-07 16:40:32"
}
```
2. Check the IP address configuration of the machine group.

Log on to the Log Service console, click the target project, and then click Logtail Machine Group in the left-side navigation pane. On the Machine Groups page, click Status in the Actions column of the target machine group.

Figure 7-3: Check the machine group status

<table>
<thead>
<tr>
<th>No.</th>
<th>ip</th>
<th>Heartbeat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>192.1[red]0.10.15</td>
<td>FAIL</td>
</tr>
<tr>
<td>2</td>
<td>192.1[red]0.10.16</td>
<td>FAIL</td>
</tr>
</tbody>
</table>

If the server IP addresses configured in the machine group do not include the server IP address obtained by the Logtail client, you must modify the IP address configuration.

- If an incorrect server IP address is configured in the machine group, modify the IP address and save it. Then, check the heartbeat status again in 1 minute.
- If you have modified the network configuration of the server where Logtail is installed, for example, /etc/hosts, you must restart Logtail to obtain the new server IP address. In addition, you must modify the server IP address configured in the machine group based on the `ip` field in the `app_info.json` file.

Logtail restart methods

- Linux:

  ```bash
  sudo /etc/init.d/ilogtaild stop
  ```
sudo /etc/init.d/ilogtaild start

- Windows: On Control Panel, choose System and Security > Administrative Tools, and then double-click Services. Find LogtailWorker and restart it.

If the server IP address obtained by the Logtail client is configured in the machine group, go to the next step.

4. Check whether an AliUid is configured for the ECS instance under another Alibaba Cloud account

If your ECS instance and Log Service project belong to different Alibaba Cloud accounts, or you use a server deployed in an on-premises IDC or provided by another cloud product vendor, you must configure an AliUid to authorize the server where Logtail is installed.

Check whether a file named after your AliUid exists in the /etc/ilogtail/users directory.

If such a file does not exist, configure an AliUid. For more information, see Configure AliUids for ECS servers under other Alibaba Cloud accounts or on-premises IDCs.

**Note:**

- The AliUid must be an Alibaba Cloud account ID.
- You can view your Alibaba Cloud account ID in the Alibaba Cloud console as follows: Move the point over your avatar and choose User Info > Security Settings.

Figure 7-4: View your Alibaba Cloud account ID
If the problem persists, submit a ticket to Log Service engineers. Along with the ticket, provide the information about your project, Logstore, and machine group, the app_info.json and ilogtail_config.json files, and the output of the Logtail automatic diagnostic tool.

7.3 Do I need to update Logtail settings after the network type is changed?

After the network type is changed from classic network to VPC, you need to restart Logtail and update the settings of the Logtail machine group.

After Logtail is installed, if your ECS network type is changed from classic network to VPC, you need to update Logtail settings by performing the following steps:

1. Restart Logtail as the admin user.

   • In Linux:
     
     ```bash
     sudo /etc/init.d/ilogtaild stop
     sudo /etc/init.d/ilogtaild start
     ```

   • In Windows:
     
     In Control Panel, choose System and Security > Administrative Tools. Open the Services program, locate the LogtailWorker file, and then right-click the file and click Restart in the shortcut menu.
2. Update the machine group settings.

- Custom identity
  
  If a custom identity is configured for the machine group, a VPC is accessible without the need to manually update the machine group settings.

- IP address
  
  If the IP address of the ECS server is configured for the machine group, you need to replace the machine group IP address to the one that is obtained after the Logtail restart, namely, the ip field in the app_info.json file.

  The app_info.json file is stored in:

  - /usr/local/ilogtail/app_info.json in Linux
  - C:\Program Files (x86)\Alibaba\Logtail\app_info.json in Windows x64
  - C:\Program Files\Alibaba\Logtail\app_info.json in Windows x32

7.4 Diagnose collection errors

Errors may occur during log collection by Logtail, such as regular expression parsing failures, incorrect file paths, and traffic exceeding the shard service capability. Currently, the diagnosis function is provided in the Log Service console for diagnosing log collection errors.

Procedure

1. Log on to the Log Service console, and then click the target project name.
2. On the Logstores page, click Diagnose in the Log Collection Mode column.

Figure 7-5: Diagnosis
3. Check log collection errors.

In the displayed dialog box, view the list of log collection errors. To view error details, move your cursor to the Error Type column.

For more information, see Log collection error types.

Figure 7-6: View collection errors

4. Query log collection errors of a specified machine

To query all log collection errors occurred to a specific machine, enter the IP address of the machine in the search box on the query page. Logtail reports errors every 5 minutes.

After fixing these errors and resuming business, check if the errors persist based on the timeframe. Historical error reports are still displayed before they expire. Ignore the historical error reports and only check whether new errors occurred after these historical errors are fixed.

Note:

To view all the complete log lines that are discarded because of parsing failure, you can log on to the machine to view the /usr/local/ilogtail/ilogtail.LOG file.
7.5 Log collection error types

On the Logstores page, you can click Diagnose of a Logstore to view all log collection errors about it. This topic describes the specific error types and handling methods.

If you encounter an error not mentioned in this topic, you can open a ticket and submit error details.

<table>
<thead>
<tr>
<th>Error type</th>
<th>Description</th>
<th>Handling method</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOGFILE_PERMINSSION_ALARM</td>
<td>Logtail has no permission to read the specified file.</td>
<td>Check the Logtail startup account on the server. We recommend that you start Logtail as the root user.</td>
</tr>
<tr>
<td>SPLIT_LOG_FAIL_ALARM</td>
<td>The line start regular expression cannot match line starts of the log, and the log cannot be split into lines.</td>
<td>Check the correctness of the line start regular expression. If the log contains only one line, you can set the line start regular expression to . * .</td>
</tr>
<tr>
<td>MULTI_CONFIG_MATCH_ALARM</td>
<td>Each file can only be collected by one Logtail Config.</td>
<td>Check whether a file is collected in multipleConfigs. If yes, delete unnecessaryConfigs.</td>
</tr>
<tr>
<td>REGEX_MATCH_ALARM</td>
<td>The log content does not match the regular expression upon regular expression parsing.</td>
<td>Copy a part from the unmatched content as a sample for a rematch, and then generate a new regular expression.</td>
</tr>
<tr>
<td>PARSE_LOG_FAIL_ALARM</td>
<td>Log parsing fails because the formats of JSON and delimiter logs do not conform to format definitions.</td>
<td>Click the error message to view details.</td>
</tr>
<tr>
<td>Error type</td>
<td>Description</td>
<td>Handling method</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>CATEGORY_CONFIG_ALARM</td>
<td>The Logtail Config is invalid.</td>
<td>We recommend that you modify the regular expression because generally the process of extracting a file path as the Topic through the regular expression fails. If the error is due to another reason, open a ticket and submit error details.</td>
</tr>
<tr>
<td>LOGTAIL_CRASH_ALARM</td>
<td>Logtail cannot respond because its server resource usage has exceeded the upper limit.</td>
<td>Modify the upper limits of the CPU usage and memory usage by following the instructions provided in Set startup parameters. You can also open a ticket for additional support.</td>
</tr>
<tr>
<td>REGISTER_INOTIFY_FAIL_ALARM</td>
<td>Registering log monitoring has failed on Linux. A possible cause is that Logtail does not have the permission to access the folder or the folder has been deleted.</td>
<td>Check whether Logtail has the permission to access the folder and whether the folder has been deleted.</td>
</tr>
<tr>
<td>DISCARD_DATA_ALARM</td>
<td>The CPU resources for configuring Logtail are insufficient or the incoming traffic to Log Service is restricted.</td>
<td>Modify the upper limit of the CPU usage or limits on concurrent incoming traffic to Log Service by following the instructions provided in Set startup parameters. You can also open a ticket for additional support.</td>
</tr>
<tr>
<td>Error type</td>
<td>Description</td>
<td>Handling method</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| SEND_DATA_FAIL_ALARM          | • The Alibaba Cloud account have not created any AccessKey (AK).  
• The Logtail client cannot connect to the Log Service server, or the connection quality is poor.  
• The writing quota on the server is insufficient.                                                                                                 | • Use the Alibaba Cloud account to create an AK.  
• Check the local Config file `/usr/local/ilogtail/ilogtail_config.json` and run `curl<endpoint>` to check whether any result is returned.  
• Increase the number of Shards for the Logstore so that more data can be written to the Logstore.                                                                                           |
<p>| REGISTER_INOTIFY_FAIL_ALARM   | Logtail fails to register inotify watcher for the log directory.                                                                                                                                               | Check whether the directory exists. If yes, check the directory permissions.                                                                                                                                 |
| SEND_QUOTA_EXCEED_ALARM       | The log writing amount exceeds the limit.                                                                                                                                                                     | Expand the Shard capacity in the console.                                                                                                                                                                    |
| READ_LOG_DELAY_ALARM          | Log collection lags behind log generation. In normal cases, this is because the CPU resources for configuring Logtail are insufficient or the incoming traffic to Log Service is restricted.                              | Modify the upper limit of the CPU usage or limits on concurrent incoming traffic to Log Service by following the instructions provided in <a href="https://cloud.enviro.com/">Set startup parameters</a>. You can also open a ticket for additional support. |
| DROP_LOG_ALARM                | Log collection lags behind log generation, and unprocessed log rotations outnumber 20. In normal cases, this is because the CPU resources for configuring Logtail are insufficient or the incoming traffic to Log Service is restricted.                 | Modify the upper limit of the CPU usage or limits on concurrent incoming traffic to Log Service by following the instructions provided in <a href="https://cloud.enviro.com/">Set startup parameters</a>. You can also open a ticket for additional support. |</p>
<table>
<thead>
<tr>
<th>Error type</th>
<th>Description</th>
<th>Handling method</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOGDIR_PERMSSION_ALARM</td>
<td>Logtail has no permission to read the log monitoring directory.</td>
<td>Check whether the log monitoring directory exists. If yes, check the directory permissions.</td>
</tr>
<tr>
<td>ENCODING_CONVERT_ALARM</td>
<td>Code conversion fails.</td>
<td>Check whether the log encoding format conforms to the specified format.</td>
</tr>
<tr>
<td>OUTDATED_LOG_ALARM</td>
<td>The log is outdated because the time when Logtail received the log has exceeded more than 12 hour and the log is expired. Possible causes are as follows:  · Log parsing is more than 12 hours behind schedule.  · Custom time fields are incorrect.  · The time of the log recording program is incorrect.</td>
<td>· Check whether READ_LOG_DELAY_ALARM exists. If yes, use the method of handling READ_LOG_DELAY_ALARM to handle this error. If no, check the time filed settings.  · Check the time filed settings. If the time field settings are correct, check whether the time of the log recording program is correct.  You can also open a ticket for additional support.</td>
</tr>
<tr>
<td>STAT_LIMIT_ALARM</td>
<td>The number of files in the Logtail Config directory exceeds the upper limit.</td>
<td>Check whether the Logtail Config directory contains an excessive number of files and subdirectories. If yes, configure the root monitoring directory and the maximum directory monitoring depth as needed.</td>
</tr>
<tr>
<td>Error type</td>
<td>Description</td>
<td>Handling method</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>DROP_DATA_ALARM</td>
<td>When the log collection process exits, writing logs to the local disk expires. In this case, the logs that have not been written to the local disk will be discarded.</td>
<td>Modify the upper limit of the CPU usage or limits on concurrent incoming traffic to Log Service by following the instructions provided in Set startup parameters. Generally, the error is caused by severe collection blocks. You can also open a ticket for additional support.</td>
</tr>
<tr>
<td>INPUT_COLLECT_ALARM</td>
<td>An error occurs during input source collection.</td>
<td>Handle the error according to the error message.</td>
</tr>
<tr>
<td>HTTP_LOAD_ADDRESS_ALARM</td>
<td>The address in the HTTP input source is invalid.</td>
<td>Check the validity of the address.</td>
</tr>
<tr>
<td>HTTP_COLLECT_ALARM</td>
<td>An error occurs during HTTP input source collection.</td>
<td>Handle the error according to the error message. In normal cases, the error is caused by expiration.</td>
</tr>
<tr>
<td>FILTER_INIT_ALARM</td>
<td>An error occurs during filer initialization.</td>
<td>Handle the error according to the error message. In normal cases, the error is caused by invalid filter regular expressions.</td>
</tr>
<tr>
<td>INPUT_CANAL_ALARM</td>
<td>An error occurs when MySQL binlogs run.</td>
<td>Handle the error according to the error message. The canal service may restart when Logtail Config is updated. Errors caused by service restart can be ignored.</td>
</tr>
<tr>
<td>Error type</td>
<td>Description</td>
<td>Handling method</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>--------------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>CANAL_INVALID_ALARM</td>
<td>The internal state of MySQL binlogs is abnormal.</td>
<td>Check whether the table schema is being modified when the error occurs. This error generally occurs when meta data changes are caused by table schema modifications during binlog running. Open a tick if the cause is another one.</td>
</tr>
<tr>
<td>MYSQL_INIT_ALARM</td>
<td>An error occurs during MySQL initialization.</td>
<td>Handle the error according to the error message.</td>
</tr>
<tr>
<td>MYSQL_CHECKPOING_ALARM</td>
<td>The MySQL checkpoint format is incorrect.</td>
<td>Check whether to modify the checkpoint settings in the current Logtail Config. If the error persists after you modify the checkpoint settings, open a ticket and submit error details.</td>
</tr>
<tr>
<td>MYSQL_TIMEOUT_ALARM</td>
<td>The MySQL query expires.</td>
<td>Check whether the error is caused by MySQL server faults or abnormal network status.</td>
</tr>
<tr>
<td>MYSQLPARSE_ALARM</td>
<td>Parsing MySQL query results fails.</td>
<td>Check whether the checkpoint format configured in MySQL matches the format of the corresponding field.</td>
</tr>
<tr>
<td>AGGREGATOR_ADD_ALARM</td>
<td>Logtail fails to add data to the queue.</td>
<td>Ignore the error if the actual data amount is large because the error is caused by excessive fast data sending.</td>
</tr>
<tr>
<td>Error type</td>
<td>Description</td>
<td>Handling method</td>
</tr>
<tr>
<td>-------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| ANCHOR_FIND_ALARM       | Possible error causes are anchor plug-in faults, Config faults, or mismatch between the Config and log. | Click the error message to view details, which may contain the following error types. Check whether the corresponding Config encounters faults accordingly.  
· anchor cannot find key: The SourceKey is specified in the Config but its corresponding field cannot be found in the log.  
· anchor no start: The keywords specified by Start cannot be found in the value of SourceKey.  
· anchor no stop: The keywords specified by Stop cannot be found in the value of SourceKey. |
<p>| ANCHOR_JSON_ALARM       | An error occurs when the anchor plug-in performs JSON expansion on the keywords specified by Start and Stop. | Click the error message to view details. Check the keywords and the related Config. Check whether there is any Config fault or invalid log. |
| CANAL_RUNTIME_ALARM     | An error occurs when the binlog plug-in runs.                                | Click the error message to view details, and then handle the error accordingly. The error is related to the connected MySQL master database. |</p>
<table>
<thead>
<tr>
<th>Error type</th>
<th>Description</th>
<th>Handling method</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHECKPOINT_INVALID_ALARM</td>
<td>The plug-in fails to parse the checkpoint.</td>
<td>Click the error message to view details, and then handle the error according to the checkpoint key, checkpoint content (the first 1,024 bytes), and other information.</td>
</tr>
<tr>
<td>DIR_EXCEED_LIMIT_ALARM</td>
<td>The number of directories for simultaneous monitoring exceeds the upper limit.</td>
<td>Check whether the Config of the current Logstore and other Configs applied on Logtail contain excessive directories. If yes, configure the root monitoring directory and the maximum directory monitoring depth as needed.</td>
</tr>
<tr>
<td>DOCKER_FILE_MAPPING_ALARM</td>
<td>Logtail fails to add Docker file mapping by executing commands.</td>
<td>Click the error message to view details, and then handle the error accordingly.</td>
</tr>
<tr>
<td>DOCKER_FILE_MATCH_ALARM</td>
<td>The specified file cannot be found in Docker.</td>
<td>Click the error message to view details, and then handle the error according to the container information and file path.</td>
</tr>
<tr>
<td>DOCKER_REGEX_COMPILE_ALARM</td>
<td>The docker stdout plug-in fails to construct a regular expression based on BeginLineRegex in the Config.</td>
<td>Click the error message to view details, and then check whether the regular expression is correct.</td>
</tr>
<tr>
<td>Error type</td>
<td>Description</td>
<td>Handling method</td>
</tr>
<tr>
<td>-------------------------</td>
<td>--------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>DOCKER_STDOUT_INIT_ALARM</td>
<td>The docker stdout collection initialization fails.</td>
<td>Click the error message to view details, which may contain the following error types:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>· host... version... error: Check whether the Docker engine specified in the Config is accessible.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>· load checkpoint error: Ignore the error if there is no impact because the error is caused by checkpoint loading failure.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>· container...: Set either stdout or stderr as a label. The error is caused because the specified container has an invalid label value. Handle the error according to the error details.</td>
</tr>
<tr>
<td>DOCKER_STDOUT_START_ALARM</td>
<td>The stdout file size exceeds the upper limit during docker stdout collection initialization.</td>
<td>Ignore the error because, in normal cases, the stdout file already exists at the first collection.</td>
</tr>
<tr>
<td>DOCKER_STDOUT_STAT_ALARM</td>
<td>The docker stdout plug-in cannot check the stdout file.</td>
<td>Ignore the error because the container cannot access the stdout file after the container exits.</td>
</tr>
<tr>
<td>FILE_READER_EXCEED_ALARM</td>
<td>The number of objects opened by Logtail exceeds the upper limit.</td>
<td>Check whether the Config settings are appropriate because the error is caused by excessive files being collected.</td>
</tr>
<tr>
<td>Error type</td>
<td>Description</td>
<td>Handling method</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>GEOIP_ALARM</td>
<td>The geoip plug-in is faulty.</td>
<td>Click the error message to view details, which may contain the following error types:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• invalid ip...: The plug-in fails to obtain the IP address. Check whether SourceKey in the Config is correct or whether an invalid log exists.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• parse ip...: The plug-in fails to parse the city information based on the obtained IP address. Handle the error according to the error details.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• cannot find key...: The plug-in cannot find the specified SourceKey from the log. Check whether the Config is faulty or whether an invalid log exists.</td>
</tr>
<tr>
<td>HTTP_INIT_ALARM</td>
<td>The http plug-in incorrectly compiles the ResponseStringMatch regular expression specified in the Config.</td>
<td>Click the error message to view details, and then check whether the regular expression is correct.</td>
</tr>
<tr>
<td>HTTP_PARSE_ALARM</td>
<td>The http fails to receive HTTP responses.</td>
<td>Click the error message to view details, and then check the Config or the requested HTTP server.</td>
</tr>
<tr>
<td>INIT_CHECKPOINT_ALARM</td>
<td>The binlog plug-in fails to load the checkpoint. In this case, the plug-in will ignore the checkpoint and recollect the log.</td>
<td>Click the error message to view details, and then determine whether the error can be ignored.</td>
</tr>
<tr>
<td>Error type</td>
<td>Description</td>
<td>Handling method</td>
</tr>
<tr>
<td>----------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>LOAD_LOCAL_EVENT_ALARM</td>
<td>Logtail handles a local event.</td>
<td>Ignore the error if it is caused by manual operations. For other cases, open a ticket and submit error details. Click the error message to view details, and then handle the error according to the file name, Config name, project, Logstore, and other information.</td>
</tr>
<tr>
<td>LOG_REGEX_FIND_ALARM</td>
<td>The processor_split_log_regex and processor_split_log_string plug-ins cannot obtain the SplitKey specified by the Config from the log.</td>
<td>Click the error message to view details, and then check whether the Config is faulty.</td>
</tr>
<tr>
<td>LUMBER_CONNECTION_ALARM</td>
<td>The server cannot be powered off when the service_lumberjack plugin is stopped.</td>
<td>Click the error message to view details, and then handle the error accordingly. In normal cases, this error can be ignored.</td>
</tr>
</tbody>
</table>
| LUMBER_LISTEN_ALARM        | An error occurs when the service_lumberjack plugin is being initiated for log monitoring. | Click the error message to view details, which may contain the following error types:    
  - init tls error...: Check whether the TLS configurations are correct.  
  - listen init error...: Check whether the address-related settings are correct. |
<table>
<thead>
<tr>
<th>Error type</th>
<th>Description</th>
<th>Handling method</th>
</tr>
</thead>
<tbody>
<tr>
<td>LZ4_COMPRESS_FAIL_ALARM</td>
<td>An error occurs when Logtail executes LZ4 compression.</td>
<td>Click the error message to view details, and then handle the error according to the values of log lines, project, category, and region.</td>
</tr>
<tr>
<td>MYSQL_CHECKPOINT_ALARM</td>
<td>The MySQL plug encounters a checkpoint error.</td>
<td>Click the error message to view details, which may contain the following error types:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• init checkpoint error ...: Initializing the checkpoint fails. In this case, check whether the checkpoint column specified by the Config and the corresponding values are correct.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• not matched checkpoint...: The checkpoint information does not match. In this case, check whether the mismatch is caused by manual operations, for example, Config updates. If yes, ignore the error.</td>
</tr>
<tr>
<td>NGINX_STATUS_COLLECT_ALARM</td>
<td>An error occurs when the nginx_status plug-in obtains the server status.</td>
<td>Click the error message to view details, and then handle the error according to the URL and other information.</td>
</tr>
<tr>
<td>NGINX_STATUS_INIT_ALARM</td>
<td>The nginx_status plug-in fails to initiate and parse the URL specified by the Config.</td>
<td>Click the error message to view details, and then check whether the address is correct according to the URL.</td>
</tr>
<tr>
<td>Error type</td>
<td>Description</td>
<td>Handling method</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>OPEN_FILE_LIMIT_ALARM</td>
<td>Logtail cannot open new files because the number of opened files has exceeded the upper limit.</td>
<td>Click the error message to view details, and then handle the error according to the log file path, project, Logstore, and other information.</td>
</tr>
<tr>
<td>OPEN_LOGFILE_FAIL_ALARM</td>
<td>An error occurs when Logtail opens a file.</td>
<td>Click the error message to view details, and then handle the error according to the log file path, project, Logstore, and other information.</td>
</tr>
<tr>
<td>PARSE_DOCKER_LINE_ALARM</td>
<td>The service_docker_stdout plug-in fails to parse the log.</td>
<td>Click the error message to view details, which may contain the following error types:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• parse docker line error: empty line: The log is empty.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• parse json docker line error...: The plug-in fails to parse the log in JSON format. In this case, handle the error according to the error message and the first 512 bytes of the log.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• parse cri docker line error...: The plug-in fails to parse the log in CRI format. In this case, handle the error according to the error message and the first 512 bytes of the log.</td>
</tr>
<tr>
<td>Error type</td>
<td>Description</td>
<td>Handling method</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| PLUGIN_ALARM       | An error occurs when the plug-in is initialized or called.                  | Click the error message to view details, which may contain the following error types. Handle the error accordingly.  
  - init plugin error...: Initiating the plug-in fails.  
  - hold on error...: Stopping the plug-in fails.  
  - resume error...: Recovering the plug-in fails.  
  - start service error...: Starting service input-type plug-ins fails.  
  - stop service error...: Stopping service input-type plug-ins fails. |
<p>| PROCESSOR_INIT_ALARM | The regex plug-in fails to compile the Regex regular expression specified by the Config. | Click the error message to view details, and then check whether the regular expression is correct. |</p>
<table>
<thead>
<tr>
<th>Error type</th>
<th>Description</th>
<th>Handling method</th>
</tr>
</thead>
</table>
| PROCESS_TOO_SLOW_ALARM     | Logtail parses logs too slowly.          | 1. Click the error message to view details, and then determine whether the slow parsing is normal according to the number of logs, buffer size, and parsing time.  
<pre><code>                          |                                                                                          | 2. If the slow parsing is abnormal, check whether inappropriate parsing configurations exist. For example, the processes on the node where Logtail resides occupy excessive CPU resources, or an inefficient regular expression exists. |
</code></pre>
<p>| REDISPARSE_ADDRESS_ALARM   | The redis plug-in fails to parse the ServerUrls specified by the Config.                  | Click the error message to view details, and then check the URL.                                                                                |
| REGEX_FIND_ALARM           | The regex plug-in cannot find the fields specified by SourceKey in the Config from the log. | Click the error message to view details, and then check whether the SourceKey is incorrect or an invalid log exists.                       |</p>
<table>
<thead>
<tr>
<th>Error type</th>
<th>Description</th>
<th>Handling method</th>
</tr>
</thead>
<tbody>
<tr>
<td>REGEX_UNMATCHED_ALARM</td>
<td>The regex plug-in fails to match the log.</td>
<td>Click the error message to view details, which may contain the following error types. Handle the error accordingly, for example, determine whether the Config is correct.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- unmatch this log content...: The log cannot match the regular expression in the Config.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- match result count less...: The number of matched logs is less than that of Keys specified in the Config.</td>
</tr>
<tr>
<td>SAME_CONFIG_ALARM</td>
<td>There are Configs with the same name in a Logstore. In this case, Logtail chooses one to collect the log, and the others will be discarded.</td>
<td>Click the error message to view details, and then handle the error according to the Config path and other information.</td>
</tr>
<tr>
<td>SPLIT_FIND_ALARM</td>
<td>The split_char and split_string plug-ins cannot find the fields specified by SourceKey in the Config from the log.</td>
<td>Click the error message to view details, and then check whether SourceKey settings are incorrect or an invalid log exists.</td>
</tr>
<tr>
<td>SPLIT_LOG_ALARM</td>
<td>The number of parsed fields parsed by the processor_split_char and processor_split_string plug-ins does not match that of fields specified by SplitKeys.</td>
<td>Click the error message to view details, and then check whether SourceKey settings are incorrect or an invalid log exists.</td>
</tr>
<tr>
<td>Error type</td>
<td>Description</td>
<td>Handling method</td>
</tr>
<tr>
<td>----------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>STAT_FILE_ALARM</td>
<td>An error occurs when the plug-in collects files through theLogFileReader object.</td>
<td>Click the error message to view details, and handle the error according to the file path and other information.</td>
</tr>
<tr>
<td>SERVICE_SYSLOG_INIT_ALARM</td>
<td>The service_syslog plug-in initialization fails.</td>
<td>Click the error message to view details, and check whether Address in the Config is correct.</td>
</tr>
<tr>
<td>SERVICE_SYSLOG_STREAM_ALARM</td>
<td>An error occurs when the service_syslog plug-in collects data through TCP.</td>
<td>Click the error message to view details, which may contain the following error types. Handle the error accordingly.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• accept error...: An error occurs during Accept execution. In this case, the plug-in waits for a while and restarts.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• setKeepAlive error...: Setting Keep Alive fails. In this case, the plug-in ignores the error and runs properly.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• connection i/o timeout ...: Reading data through TCP expires. In this case, the plug-in resets the expiration duration and reads data properly.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• scan error...: An error occurs when the plug-in reads data through TCP. In this case, the plug-in waits for a while and tries again.</td>
</tr>
<tr>
<td>Error type</td>
<td>Description</td>
<td>Handling method</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>SERVICE_SYSLOG_PACKET_ALARM</td>
<td>An error occurs when the service_syslog plug-in collects data through UDP.</td>
<td>Click the error message to view details, which may contain the following error types. Handle the error accordingly.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• connection i/o timeout ...: Reading data through UDP expires. In this case, the plug-in resets the expiration duration and reads data properly.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• read from error...: An error occurs when the plug-in reads data through UDP. In this case, the plug-in waits for a while and tries again.</td>
</tr>
</tbody>
</table>

7.6 Query local collection status

Logtail is used to query its own health status and log collection progress, helping you troubleshoot log collection issues and customize status monitoring for log collection.

1. User guide
   a. all command
   b. active command
   c. logstore command
   d. logfile command
   e. history command

2. Return values
3. Use cases

a. Monitor the running status of Logtail
b. Monitor log collection progress
c. Determine whether or not Logtail has finished collecting log files
d. Troubleshoot log collection issues

User guide

If a Logtail client supporting status query function is installed, you can query local log collection status by entering commands on the client. To install Logtail, see Install Logtail in Linux.

Enter the `/etc/init.d/logtaild -h` command on the client to check if the client supports querying local log collection status. If the `logtail insight, version` keyword is returned, it indicates that this function is supported on the Logtail client.

```
/etc/init.d/logtaild
Usage: ./logtaild { start | stop (graceful, flush data and save checkpoints) | force-stop | status | -h for help}$
logtail insight, version : 0.1.0
command list :
   status all [index]
      get logtail running status
   status active [--logstore | --logfile] index [project] [logstore]
      list all active logstore | logfile. if use --logfile, please add project and logstore. default --logstore
   status logstore [--format=line | json] index project logstore
      get logstore status with line or json style. default --format=line
   status logfile [--format=line | json] index project logstore fileFullPath
      get log file status with line or json style. default --format=line
   status history beginIndex endIndex project logstore [fileFullPath]
      query logstore | logfile history status.
   index : from 1 to 60. in all, it means last $(index) minutes; in active/logstore/logfile/history, it means last $(index)*10 minutes
```

Currently, Logtail supports the following query commands, command functions, time intervals to query and time windows for result statistics:

<table>
<thead>
<tr>
<th>Command</th>
<th>Functions</th>
<th>Time interval to query</th>
<th>Time window for statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>all</td>
<td>Query the running status of Logtail.</td>
<td>Last 60 min</td>
<td>1 min</td>
</tr>
<tr>
<td>Command</td>
<td>Functions</td>
<td>Time interval to query</td>
<td>Time window for statistics</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
<td>------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>active</td>
<td>Query Logstores or log files that are currently active (that is, with data collected).</td>
<td>Last 600 min</td>
<td>10 minutes.</td>
</tr>
<tr>
<td>logstore</td>
<td>Query the collection status of a Logstore.</td>
<td>Last 600 min</td>
<td>10 minutes.</td>
</tr>
<tr>
<td>logfile</td>
<td>Query the collection status of a log file.</td>
<td>Last 600 min</td>
<td>10 minutes.</td>
</tr>
<tr>
<td>history</td>
<td>Query the collection status of a Logstore or log file over a period of time.</td>
<td>Last 600 min</td>
<td>10 minutes.</td>
</tr>
</tbody>
</table>

**Note:**

- The `index` parameter in the command represents the index value of the time window, which is counted from the current time. Its valid range is 1–60. If the time window for statistics is one minute, windows in the last \((index, index-1]\) minutes are queried. If the time window for statistics is 10 minutes, windows in the last \((10*index, 10*(index-1)]\) minutes are queried.

- All query commands belong to status subcommands, so the main command is `status`.

**Command format**

`/etc/init.d/ilogtaild status all [ index ]`

**Note:**

The `all` command is used to view the running status of Logtail. The `index` parameter is optional. If left blank, 1 is taken by default.
### Example

```
/etc/init.d/ilogtaild status all 1
  ok
/etc/init.d/ilogtaild status all 10
  busy
```

### Output description

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
<th>Priority</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>ok</td>
<td>The current status is normal.</td>
<td>None.</td>
<td>No action is needed.</td>
</tr>
<tr>
<td>busy</td>
<td>The current collection speed is high and the Logtail status is normal.</td>
<td>None.</td>
<td>No action is needed.</td>
</tr>
<tr>
<td>many_log_files</td>
<td>The number of logs being collected is large.</td>
<td>Low</td>
<td>Check if the configuration contains files that do not need to be collected.</td>
</tr>
<tr>
<td>process_block</td>
<td>Current log parsing is blocked.</td>
<td>Low</td>
<td>Check if logs are generated too quickly. If you still get this output,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Check your needs to modify the upper limit of CPU usage or the limit on concurrent sending by using network.</td>
</tr>
</tbody>
</table>

*Set startup parameters*
<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
<th>Priority</th>
<th>Resolution:</th>
</tr>
</thead>
<tbody>
<tr>
<td>send_block</td>
<td>Current sending is blocked.</td>
<td>Relatively high</td>
<td>blocked. Check if logs are generated too quickly and if the network status is normal. If you still get this output, <em>Set startup parameters</em> as per your needs to modify the upper limit of CPU usage or the limit on concurrent sending by using network.</td>
</tr>
<tr>
<td>send_error</td>
<td>Failed to upload log data.</td>
<td>High</td>
<td>To troubleshoot the issue, see <em>Diagnose collection errors</em>.</td>
</tr>
</tbody>
</table>

**active command**

**Command format**

```
/etc/init.d/ilogtaild status active [--logstore] index
/etc/init.d/ilogtaild status active --logfile index project-name logstore-name
```

**Note:**

- The `active [--logstore] index` command is used to query Logstores that are currently active. The `--logstore` parameter can be omitted without changing the meaning of the command.
- The `active --logfile index project-name logstore-name` command is used to query all active log files in a Logstore for a project.
- The active command is used to query active log files level by level. We recommend that you first locate the currently active Logstore and then query active log files in this Logstore.

**Example**

```
/etc/init.d/ilogtaild status active 1
sls-zc-test : release-test
sls-zc-test : release-test-ant-rpc-3
```
sls-zc-test : release-test-same-regex-3
/etc/init.d/ilogtaild status active --logfile 1 sls-zc-test release-test
/disk2/test/normal/access.log

Output description

- To run the `active --logstore index` command, all currently active Logstores are output in the format of `project-name : logstore-name`. To run the `active --logfile index project-name logstore-name` command, the complete paths of active log files are output.
- A Logstore or log file with no log collection activity in the current query window does not appear in the output.

logstore command

Command format

```
/etc/init.d/ilogtaild status logstore [--format={line|json}] index
project-name logstore-name
```

Note:

- The logstore command is used to output the collection statuses of the specified project and Logstore in LINE or JSON format.
- If the `--format=` parameter is not configured, `--format=line` is selected by default. The echo information is output in LINE format. Note that `--format` parameter must be placed behind `logstore`.
- If this Logstore does not exist or has no log collection activity in the current query window, you get an empty output in LINE format or a null value in JSON format.

Example

```
/etc/init.d/ilogtaild status logstore 1 sls-zc-test release-test-same
time_begin_readable : 17-08-29 10:56:11
time_end_readable : 17-08-29 11:06:11
time_begin : 1503975371
time_end : 1503975971
project : sls-zc-test
logstore : release-test-same
status : ok
config : ##1.0##sls-zc-test$same
read_bytes : 65033430
parse_success_lines : 230615
parse_fail_lines : 0
```
last_read_time : 1503975970
read_count : 687
avg_delay_bytes : 0
max_unsend_time : 0
min_unsend_time : 0
max_send_success_time : 1503975968
send_queue_size : 0
send_network_error_count : 0
send_network_quota_count : 0
send_network_discard_count : 0
send_success_count : 302
send_block_flag : false
sender_valid_flag : true
/etc/init.d/logtaild status logstore --format=json 1 sls-zc-test
release-test-same
{
    "avg_delay_bytes" : 0,
    "config" : "##1.0##sls-zc-test$same",
    "last_read_time" : 1503975970,
    "logstore" : "release-test-same",
    "max_send_success_time" : 1503975968,
    "max_unsend_time" : 0,
    "min_unsend_time" : 0,
    "parse_fail_lines" : 0,
    "parse_success_lines" : 230615,
    "project" : "sls-zc-test",
    "read_bytes" : 65033430,
    "read_count" : 687,
    "send_block_flag" : false,
    "send_network_discard_count" : 0,
    "send_network_error_count" : 0,
    "send_network_quota_count" : 0,
    "send_queue_size" : 0,
    "send_success_count" : 302,
    "sender_valid_flag" : true,
    "status" : "ok",
    "time_begin" : 1503975371,
    "time_begin_readable" : "17-08-29 10:56:11",
    "time_end" : 1503975971,
    "Maid" : "17-08-29 11:06:11"
}

### Output description

<table>
<thead>
<tr>
<th>Reserved Word</th>
<th>Meaning</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Status</strong></td>
<td>The overall status of this Logstore. For specific statuses, descriptions, and change methods, see the following table.</td>
<td>None.</td>
</tr>
<tr>
<td>time_begin_readable</td>
<td>The start time that can be read.</td>
<td>None.</td>
</tr>
<tr>
<td>time_end_readable</td>
<td>The end time that can be read.</td>
<td>None.</td>
</tr>
<tr>
<td>Reserved Word</td>
<td>Meaning</td>
<td>Unit</td>
</tr>
<tr>
<td>---------------</td>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>time_begin</td>
<td>The start time of statistics.</td>
<td>UNIX timestamp, measured in seconds.</td>
</tr>
<tr>
<td>time_end</td>
<td>The end time of statistics.</td>
<td>UNIX timestamp, measured in seconds.</td>
</tr>
<tr>
<td>project</td>
<td>The project name.</td>
<td>None.</td>
</tr>
<tr>
<td>logstore</td>
<td>The Logstore name.</td>
<td>None.</td>
</tr>
<tr>
<td>config</td>
<td>The collection configuration name, which is globally unique and consisted of #1. @#, project, $, and config.</td>
<td>None.</td>
</tr>
<tr>
<td>read_bytes</td>
<td>The number of logs read in the window.</td>
<td>Byte</td>
</tr>
<tr>
<td>parse_success_lines</td>
<td>The number of successfully parsed log lines in the window.</td>
<td>Line</td>
</tr>
<tr>
<td>parse_fail_lines</td>
<td>The number of log lines that failed to be parsed in the window.</td>
<td>Line</td>
</tr>
<tr>
<td>last_read_time</td>
<td>The last read time in the window.</td>
<td>UNIX timestamp, measured in seconds.</td>
</tr>
<tr>
<td>Read_count</td>
<td>The number of times that logs are read in the window.</td>
<td>Number</td>
</tr>
<tr>
<td>avg_delay_bytes</td>
<td>The average of the differences between the current offset and the file size each time logs are read in the window.</td>
<td>Byte</td>
</tr>
<tr>
<td>max_unsend_time</td>
<td>The maximum time that unsent data packets are in the send queue when the window ends. The value is 0 when the queue is empty.</td>
<td>UNIX timestamp, measured in seconds.</td>
</tr>
<tr>
<td>Reserved Word</td>
<td>Meaning</td>
<td>Unit</td>
</tr>
<tr>
<td>----------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>min_unsend_time</td>
<td>The minimum time that unsent data packets are in the send queue when the window ends. The value is 0 when the queue is empty.</td>
<td>UNIX timestamp, measured in seconds.</td>
</tr>
<tr>
<td>max_send_success_time</td>
<td>The maximum time that data is successfully sent in the window.</td>
<td>UNIX timestamp, measured in seconds.</td>
</tr>
<tr>
<td>send_queue_size</td>
<td>The number of unsent data packets in the current send queue when the window ends.</td>
<td>Packet</td>
</tr>
<tr>
<td>send_network_error_count</td>
<td>The number of data packets that failed to be sent in the window because of network errors.</td>
<td>Packet</td>
</tr>
<tr>
<td>send_network_quota_count</td>
<td>The number of data packets that failed to be sent in the window because the quota is exceeded.</td>
<td>Packet</td>
</tr>
<tr>
<td>send_network_discard_count</td>
<td>The number of discarded data packets in the window because of data exceptions or insufficient permissions.</td>
<td>Packet</td>
</tr>
<tr>
<td>send_success_count</td>
<td>The number of successfully sent data packets in the window.</td>
<td>Packet</td>
</tr>
<tr>
<td>send_block_flag</td>
<td>Whether or not the send queue is blocked when the window ends.</td>
<td>None.</td>
</tr>
<tr>
<td>Reserved Word</td>
<td>Meaning</td>
<td>Unit</td>
</tr>
<tr>
<td>------------------</td>
<td>---------------------------------------------------------------------------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>sender_valid_flag</td>
<td>Whether or not the send flag of this Logstore is valid when the window ends. True means the flag is valid, and false means the flag is disabled because of network errors or quota errors.</td>
<td>None.</td>
</tr>
</tbody>
</table>

### Logstore status

<table>
<thead>
<tr>
<th>Status</th>
<th>Meaning</th>
<th>Handling method</th>
</tr>
</thead>
<tbody>
<tr>
<td>ok</td>
<td>The status is normal.</td>
<td>No action is needed.</td>
</tr>
<tr>
<td>process_block</td>
<td>Log parsing is blocked.</td>
<td>Check if logs are generated too quickly. If you still get this output, Configure <a href="#">set startup parameters</a> as per your needs to modify the upper limit of CPU usage or the limit on concurrent sending by using network.</td>
</tr>
<tr>
<td>parse_fail</td>
<td>Log parsing failed.</td>
<td>Check whether or not the log format is consistent with the log collection configuration.</td>
</tr>
<tr>
<td>send_block</td>
<td>Current sending is blocked.</td>
<td>blocked. Check if logs are generated too quickly and if the network status is normal. If you still get this output, <a href="#">set startup parameters</a> as per your needs to modify the upper limit of CPU usage or the limit on concurrent sending by using network.</td>
</tr>
<tr>
<td>Status</td>
<td>Meaning</td>
<td>Handling method</td>
</tr>
<tr>
<td>-------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>sender_invalid</td>
<td>An exception occurred when sending log data.</td>
<td>Check the network status. If the network is normal, see <em>Diagnose collection errors</em> in Query diagnosis errors to troubleshoot the issue.</td>
</tr>
</tbody>
</table>

**logfile command**

**Command format**

```
/etc/init.d/ilogtaild status logfile [--format={line|json}] index
project-name logstore-name fileFullPath
```

**Note:**

- The logfile command is used to output the collection status of a specific log file in LINE or JSON format.
- If the `--format=` parameter is not configured, `--format=line` is selected by default. The echo information is output in LINE format.
- If this log file does not exist or has no log collection activity in the current query window, you get an empty output in LINE format or a null value in JSON format.
- The `--format` parameter must be placed behind `logfile`.
- The `fileFullpath` must be a full path name.

**Example**

```
/etc/init.d/ilogtaild status logfile 1 sls-zc-test release-test-same /disk2/test/normal/access.log
time_begin_readable : 17-08-29 11:16:11
time_end_readable : 17-08-29 11:26:11
time_begin : 1503976571
time_end : 1503977171
project : sls-zc-test
logstore : release-test-same
status : ok
config : ##1.0##sls-zc-test$same
file_path : /disk2/test/normal/access.log
file_dev : 64800
file_inode : 22544456
file_size_bytes : 17154060
file_offset_bytes : 17154060
read_bytes : 65033430
parse_success_lines : 230615
parse_fail_lines : 0
last_read_time : 1503977170
read_count : 667
avg_delay_bytes : 0
```
/etc/init.d/ilogtaild status logfile --format=json 1 sls-zc-test
release-test-same /disk2/test/normal/access.log
{
    "avg_delay_bytes" : 0,
    "config" : "##1.0##sls-zc-test$same",
    "file_dev" : 64800,
    "file_inode" : 22544456,
    "file_path" : "/disk2/test/normal/access.log",
    "file_size_bytes" : 17154060,
    "last_read_time" : 1503977170,
    "logstore" : "release-test-same",
    "parse_fail_lines" : 0,
    "parse_success_lines" : 230615,
    "project" : "sls-zc-test",
    "read_bytes" : 65033430,
    "read_count" : 667,
    "read_offset_bytes" : 17154060,
    "status" : "ok",
    "time_begin" : 1503976571,
    "time_begin_readable" : "17-08-29 11:16:11",
    "time_end" : 1503977171,
    "time_end_readable" : "17-08-29 11:26:11"
}

Output description

<table>
<thead>
<tr>
<th>Reserved Word</th>
<th>Meaning</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Status</td>
<td>The collection status of this log file in the current query window. See the status of logstore command.</td>
<td>None.</td>
</tr>
<tr>
<td>time_begin_readable</td>
<td>The start time that can be read.</td>
<td>None.</td>
</tr>
<tr>
<td>time_end_readable</td>
<td>The end time that can be read.</td>
<td>None.</td>
</tr>
<tr>
<td>time_begin</td>
<td>The start time of statistics.</td>
<td>UNIX timestamp, measured in seconds.</td>
</tr>
<tr>
<td>time_end</td>
<td>The end time of statistics.</td>
<td>UNIX timestamp, measured in seconds.</td>
</tr>
<tr>
<td>project</td>
<td>The project name.</td>
<td>None.</td>
</tr>
<tr>
<td>logstore</td>
<td>The Logstore name.</td>
<td>None.</td>
</tr>
<tr>
<td>file_path</td>
<td>The path of the log file.</td>
<td>None.</td>
</tr>
<tr>
<td>file_dev</td>
<td>The device ID of the log file.</td>
<td>None.</td>
</tr>
<tr>
<td>file_inode</td>
<td>The inode of the log file.</td>
<td>None.</td>
</tr>
<tr>
<td>Reserved Word</td>
<td>Meaning</td>
<td>Unit</td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>file_size_bytes</td>
<td>The size of the last scanned file in the window.</td>
<td>Byte</td>
</tr>
<tr>
<td>read_offset_bytes</td>
<td>The parsing offset of this file.</td>
<td>Byte</td>
</tr>
<tr>
<td>config</td>
<td>The collection configuration name, which is globally unique and consisted of ##1.##, project, $ and config.</td>
<td>None.</td>
</tr>
<tr>
<td>read_bytes</td>
<td>The number of logs read in the window.</td>
<td>Byte</td>
</tr>
<tr>
<td>parse_success_lines</td>
<td>The number of successfully parsed log lines in the window.</td>
<td>Line</td>
</tr>
<tr>
<td>parse_fail_lines</td>
<td>The number of log lines that failed to be parsed in the window.</td>
<td>Line</td>
</tr>
<tr>
<td>last_read_time</td>
<td>The last read time in the window.</td>
<td>UNIX timestamp, measured in seconds.</td>
</tr>
<tr>
<td>read_count</td>
<td>The number of times that logs are read in the window.</td>
<td>Number of times</td>
</tr>
<tr>
<td>avg_delay_bytes</td>
<td>The average of the differences between the current offset and the file size each time logs are read in the window.</td>
<td>Byte</td>
</tr>
</tbody>
</table>

**history command**

**Command format**

```
/etc/init.d/ilogtaild status history beginIndex endIndex projectName logstore-name [filePath]
```
• The history command is used to query the collection status of a Logstore or log file over a period of time.
• beginIndex and endIndex represent the start and end values for the code query window index respectively. beginIndex <= endIndex.
• If the fileFullPath is not entered in the parameter, the code queries the collection information of the Logstore. Otherwise, the collection information of the log file is queried.

Example

```
/etc/init.d/ilogtaild status history 1 3 sls-zc-test release-test-same /disk2/test/normal/access.log
begin_time status read parse_success parse_fail
last_read_time read_count avg_delay device inode file_size read_offset
17-08-29 11:26:11 ok 62.12MB 231000 0 17-08-29 11:36:11 671 0B 64800
22544459 18.22MB 18.22MB
17-08-29 11:16:11 ok 62.02MB 230615 0 17-08-29 11:26:10 667 0B 64800
22544456 16.36MB 16.36MB
17-08-29 11:06:11 ok 62.12MB 231000 0 17-08-29 11:16:11 687 0B 64800
22544452 14.46MB 14.46MB
$/etc/init.d/ilogtaild status history 2 5 sls-zc-test release-test-same
begin_time status read parse_success parse_fail
last_read_time read_count avg_delay device inode file_size read_offset
17-08-29 11:16:11 ok 62.02MB 230615 0 17-08-29 11:26:10 667 0B 0 0 0
0 303 false true 70-01-01 08:00:00 70-01-01 08:00:00 17-08-29 11:06:08
17-08-29 10:56:11 ok 62.12MB 231000 0 17-08-29 11:06:10 692 0B 0 0 0
0 302 false true 70-01-01 08:00:00 70-01-01 08:00:00 17-08-29 11:06:08
17-08-29 10:46:11 ok 62.12MB 231000 0 17-08-29 10:56:11 692 0B 0 0 0
0 302 false true 70-01-01 08:00:00 70-01-01 08:00:00 17-08-29 10:56:10
```

Output description

• This command outputs historical collection information of a Logstore or log file in the form of list, one line for each window.
• For the description of each output field, see the logstore and logfile commands.

Return values

Normal return value

0 is returned if a command input is valid (including failure to query a Logstore or log file), for example:

```
/etc/init.d/ilogtaild status logfile --format=json 1 error-project error-logstore /no/this/file
```
null
echo $? 
0
/etc/init.d/ilogtaild status all
ok
echo $?  
0

Exceptional return values

A non-zero return value indicates an exception. See the following table.

<table>
<thead>
<tr>
<th>Return value</th>
<th>Type</th>
<th>output</th>
<th>Troubleshooting</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Invalid command or missing parameters</td>
<td>invalid param, use -h for help.</td>
<td>Enter -h to view help.</td>
</tr>
<tr>
<td>1</td>
<td>The query goes beyond the 1–60 time window</td>
<td>invalid query interval</td>
<td>Enter -h to view help.</td>
</tr>
<tr>
<td>1</td>
<td>Cannot query the specified time window</td>
<td>query fail, error : $(error). For more information, see errno interpretation.</td>
<td>This issue might occur when the startup time of Logtail is less than the query time span. For other cases, open a ticket.</td>
</tr>
<tr>
<td>1</td>
<td>No matching query window time</td>
<td>no match time interval, please check logtail Status</td>
<td>Check if Logtail is running. For other cases, open a ticket.</td>
</tr>
<tr>
<td>1</td>
<td>No data in the query window</td>
<td>invalid profile, maybe logtail Restart</td>
<td>Check if Logtail is running. For other cases, open a ticket.</td>
</tr>
</tbody>
</table>

Example

```
/etc/init.d/ilogtaild status nothiscmd
invalid param, use -h for help.
echo $? 
10
/etc/init.d/ilogtaild status/all 99
invalid query interval
echo $?  
0
```
Use cases

You can obtain the overall status of Logtail by querying its health status, and obtain the related metrics during collection by querying the collection progress. With the obtained information, you can monitor log collection in a customized manner.

Monitor the running status of Logtail

Monitor the running status of Logtail by using the all command.

How it works: The current status of Logtail is queried every minute. If Logtail is under process_block, send_block, or send_error status for five successive minutes, an alarm is triggered.

The alarm duration and the status range being monitored can be adjusted according to the importance of log collection in specific scenarios.

Monitor log collection progress

Monitor the collection progress of a Logstore by using the logstore command.

How it works: The logstore command is called every ten minutes to obtain the status information of this Logstore. If the avg_delay_bytes is over 1 MB (1024*1024) or status is not ok, an alarm is triggered.

The avg_delay_bytes alarm threshold can be adjusted according to the log collection traffic.

Determine whether or not Logtail has finished collecting log files

Determine whether or not Logtail has finished collecting log files by using the logfile command.

How it works: After writing to the log file stops, the logfile command is called every ten minutes to obtain the status information of this file. If this file shows the same value for read_offset_bytes and file_size_bytes, it means that Logtail has finished collecting this log file.

Troubleshoot log collection issues

If the log collection is delayed on a server, use the history command to query related collection information on this server.
1. If the `send_block_flag` is true, it indicates that the log collection delays because of the network.
   
   • If the `send_network_quota_count` is greater than 0, you must split the Shard of the Logstore.
   • If the `send_network_error_count` is greater than 0, you must check the network connectivity.
   • ◦ If no related network error occurs, you must adjust the limit on concurrent sending and traffic limit of Logtail.

2. Sending-related parameters are normal, but the `avg_delay_bytes` is relatively high.
   
   • The average log parsing speed can be calculated by using `read_bytes` to determine if traffic generated by logs is normal.
   • Resource usage limits of Logtail can be adjusted as appropriate.

3. The `parse_fail_lines` is greater than 0.
   
   Check if the parsing configurations for log collection match with all the logs.

7.7 How do I use the Logtail automatic diagnostic tool?

If an exception occurs during log collection, you can use the Logtail automatic diagnostic tool to check whether the exception exists on the Logtail client and quickly locate and resolve errors as instructed by the tool.

Note:
This tool is currently available only to Linux servers.

Preparations

1. Download the script of the diagnostic tool.

   ```bash
   ```

   Note:
If you fail to download the tool, use the following URL and try again.

```
```

2. Install the curl tool.

The diagnostic tool uses curl to check the network connectivity. Ensure that the curl tool is installed on the server where Logtail is installed.

Startup of the diagnostic tool

1. Run the following command to run the diagnostic tool:

```
chmod 744 ./checkingtool.sh
./checkingtool.sh
sh checkingtool.sh
```

The returned information is as follows:

```
[Info]:      Logtail checking tool version : 0.3.0
[Input]:     please choose which item you want to check :
2. MachineGroup heartbeat is ok, but log files have not been collected.
   Item :
```

2. Enter 1 or 2 as prompted. The script performs different checks based on your choice.

where:

- 1 indicates the machine group heartbeat check. Select this option if the heartbeat status of the machine group is abnormal.
- 2 indicates the log collection check. Select this option if the heartbeat status of the machine group is normal but log files are not collected.

After you select the required option, the diagnostic tool automatically performs the corresponding check.
After you select the machine group heartbeat check, the diagnostic tool checks the following items:
1. Check the basic environment.

- Whether Logtail is installed.
- Whether Logtail is running.
- Whether the SSL status is normal.
- Whether the network connection with Log Service is normal.

```
[Info]:     Logtail checking tool version : 0.3.0
[Input]:  please choose which item you want to check :
         2. MachineGroup heartbeat is ok, but log files have
            not been collected.
         Item :1
[Info]:     Check logtail install files
[Info]:     Install file: ilogtail_config.json exists.
            [ OK ]
[Info]:     Install file: /etc/init.d/ilogtaild exists.
            [ OK ]
[Info]:     Install file: ilogtail exists.
            [ OK ]
            [ OK ]
[Info]:     Logtail version :
            [ OK ]
[Info]:     Check logtail running status
[Info]:     Logtail is running.
            [ OK ]
[Info]:     Check network status
[Info]:     Logtail is using ip: 11.XX.XX.187
[Info]:     Logtail is using UUID: 0DF18E97-0F2D-486F-B77F-XXXXXXXXXXXX
[Info]:     Check SSL status
[Info]:     SSL status OK.
            [ OK ]
[Info]:     Check logtail config server
[Info]:     config server address: http://config.sls.aliyun-inc.com
[Info]:     Logtail config server OK
```

If an error message appears during the check, troubleshoot the error as prompted.

2. Check whether you are the owner of the Elastic Compute Service (ECS) instance.

After checking the basic environment, check whether your server is an ECS instance and whether you buy it with your Alibaba Cloud account.

If this server is not an ECS instance or the account used to buy the server is different from that used to access Log Service, enter y. Otherwise, enter N.

```
[Input]: Is your server non-Alibaba Cloud ECS or not belong to the same account with the current Project of Log Service ? (y/N)
```

If you enter y, the diagnostic tool returns the AliUid information configured on the server where Logtail is installed. Check whether it includes your AliUid. If
not, configure an AliUid for an ECS instance under another Alibaba Cloud account or a server in an on-premises IDC.

[Input]: Is your server non-Alibaba Cloud ECS or not belong to the same account with the current Project of Log Service? (y/N) y

[Info]: Check aliyun user id(s)
[Info]: aliyun user id: 126XXXXXXXXXX79. [OK]
[Info]: aliyun user id: 165XXXXXXXXXX50. [OK]
[Info]: aliyun user id: 189XXXXXXXXXX57. [OK]

[Input]: Is your project owner account ID is the above IDs? (y/N)

3. Check the region.

Check whether the region of your project is the same as that selected during Logtail installation. If not, reinstall Logtail.

[Input]: please make sure your project is in this region: { cn-hangzhou} (y/N) :

4. Check the IP address configuration.

Check whether the IP address of the server where Logtail is installed is configured in your machine group. If not, modify the IP address configuration. For more information, see Create a machine group with IP addresses as its identifier.

If your machine group is identified by a custom ID, check whether the custom ID configured on the server where Logtail is installed is the same as that configured in your machine group. If not, modify the custom ID configuration. For more information, see Create a machine group with a custom ID as its identifier.

[Input]: please make sure your machine group's ip is same with: {11.XX.XX.187} or your machine group's userdefined-id is in: {XX-XXXXX} (y/N) :

Log collection check

After you select the log collection check, the diagnostic tool checks the following items:
1. Check the IP address configuration.

Check whether the IP address of the server where Logtail is installed is configured in your machine group, and whether the heartbeat status is normal. If not, modify the machine group.

[Input]: please make sure your machine group's ip is same with : {11.XX.XX.187} (y/N) :

2. Check the application of the Logtail configuration.

Check whether your Logtail configuration is applied to the machine group. For more information about how to check the Logtail configurations that are applied to a machine group, see Manage server groups.

[Input]: please make sure you have applied collection config to the machine group (y/N) :Y

3. Check a specified log file.

Enter the full path of the log file to be checked. Check whether this log file can be matched in the log path of your Logtail configuration.

If the Logtail configuration is incorrect, modify the configuration and save it. Run the script again one minute later to check this item for a second time.

[Input]: please input your log file's full path (eg. /var/log/nginx/access.log) :/disk2/logs/access.log
[Info]: Check specific log file
[Info]: Check if specific log file [ /disk2/logs/access.log ] is included by user config.
[Warning]: Specific log file doesn't exist.
[Warning]
[Info]: Matched config found:
[OK]
[Info]: [Project] -> sls-zc-xxxxxx
[Info]: [Logstore] -> release-xxxxxxx
[Info]: [LogPath] -> /disk2/logs
[Info]: [FilePattern] -> *.log

Persisted exception after all checks

If the Logtail client passes all the checks but still fails to collect logs, enter y for the last option of the script and press Enter.
Add the output of the script as an attachment and submit a ticket to Alibaba Cloud after-sales engineers.

[Input]: please make sure all the check items above have passed. If the problem persists, please copy all the outputs and submit a ticket in the ticket system. : (y/N)y

Quick check

The quick check runs without confirmation. You can encapsulate and customize a quick check script.

Note:
During quick check, the diagnostic tool returns the AliUid and custom ID that identifies the machine group configured on the server where Logtail is installed. If either of them does not exist, no alert is reported. If an AliUid or a custom ID that identifies the machine group is configured for the Logtail client, check whether the configuration returned by the diagnostic tool is the same as that you configured. If not, modify the configuration as required. For more information, see Configure an AliUid for an ECS instance under another Alibaba Cloud account or a server in an on-premises IDC and Create a machine group with a custom ID as its identifier.

Procedure

Run the script ./checkingtool.sh --logFile [LogFileFullPath] to perform the check. If an exception is detected, proceed as instructed by the script.

Note:
If the Logtail client passes the specified log file check and the Logtail running environment is normal, we recommend that you log on to the Alibaba Cloud console to view the exception logs of the relevant Logtail configuration. For more information, see Diagnose collection errors.
After running the Logtail automatic diagnostic tool, you can identify the causes of Logtail collection errors, and then use an appropriate solution to resolve the error accordingly. The following table describes the causes of common Logtail collection errors and their solutions.

<table>
<thead>
<tr>
<th>Error</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Installation files are incomplete.</td>
<td>Reinstall Logtail.</td>
</tr>
<tr>
<td>Logtail is not running.</td>
<td>Run the <code>/etc/init.d/ilogtaild start</code> command to start Logtail.</td>
</tr>
<tr>
<td>Multiple Logtail processes exist.</td>
<td>Run the <code>/etc/init.d/ilogtaild stop</code> command to stop Logtail, and then run the <code>/etc/init.d/ilogtaild start</code> command to restart Logtail.</td>
</tr>
<tr>
<td>Port 443 is disabled.</td>
<td>Configure the firewall to enable port 443.</td>
</tr>
<tr>
<td>The configuration server cannot be found.</td>
<td>Check whether Logtail is properly installed on a Linux server. If not, uninstall and then reinstall Logtail. For more information, see Install Logtail in Linux.</td>
</tr>
</tbody>
</table>
## Error and Solution

<table>
<thead>
<tr>
<th>Error</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>The user configuration does not exist.</td>
<td>Check whether the following operations are performed:</td>
</tr>
<tr>
<td></td>
<td>1. A Logtail configuration is created in the console.</td>
</tr>
<tr>
<td></td>
<td>2. The server is included in a machine group.</td>
</tr>
<tr>
<td></td>
<td>3. The Logtail configuration is applied to the machine group.</td>
</tr>
<tr>
<td>The specified log file cannot be matched.</td>
<td>Check whether the Logtail configuration is correct.</td>
</tr>
<tr>
<td>The specified log file is matched more than once.</td>
<td>Logtail selects a Logtail configuration randomly if you match the specified log file multiple times. We recommend that you keep only one Logtail configuration that matches the specified log file.</td>
</tr>
</tbody>
</table>

### Common parameters for the diagnostic tool

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>--help</td>
<td>Views the help documentation.</td>
</tr>
<tr>
<td>--logFile [LogFileFullPath]</td>
<td>Checks whether Logtail collects logs from LogFileFullPath and checks the basic running environment of Logtail, such as the integrity of installation files, running status, AliUid, and network connectivity.</td>
</tr>
<tr>
<td>--logFileOnly [LogFileFullPath]</td>
<td>Only checks whether Logtail collects logs from LogFileFullPath.</td>
</tr>
<tr>
<td>--envOnly</td>
<td>Only checks the running environment of Logtail.</td>
</tr>
</tbody>
</table>

### 7.8 What are the differences between LogHub and Kafka?

Kafka is a distributed messaging system that features a high throughput and horizontal scalability. It is widely used for message publishing and subscription. Serving as open-source software, Kafka helps you build a Kafka cluster as required.

Log Service is a log platform service developed based on Apsara Distributed File System. It supports real-time collection, storage, distribution, and query of various types of logs. It provides external services by using the standard Restful API.

LogHub of Log Service provides public channels for log collection and distribution. You can use LogHub if you do not want to build or maintain a Kafka cluster.
### Log Service

<table>
<thead>
<tr>
<th>Concept</th>
<th>Kafka</th>
<th>LogHub</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage object</td>
<td>Topic</td>
<td>Logstore</td>
</tr>
<tr>
<td>Horizontal partitioning</td>
<td>Partition</td>
<td>Shard</td>
</tr>
<tr>
<td>Data consumption position</td>
<td>Offset</td>
<td>Cursor</td>
</tr>
</tbody>
</table>

#### Feature comparison

<table>
<thead>
<tr>
<th>Feature</th>
<th>Kafka</th>
<th>LogHub</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependency</strong></td>
<td>On-premises or shared Kafka cluster</td>
<td>Log Service</td>
</tr>
<tr>
<td>Communications protocol</td>
<td>TCP</td>
<td>HTTP (Restful API) and port 80</td>
</tr>
<tr>
<td>Access control</td>
<td>None</td>
<td>Signature authentication and access control based on cloud accounts</td>
</tr>
<tr>
<td>Dynamic scaling</td>
<td>None</td>
<td>Auto scaling of shards, which can be dynamically merged or split without any impact on users</td>
</tr>
<tr>
<td>Multi-tenant QoS</td>
<td>None</td>
<td>Shard-based standard throttling</td>
</tr>
<tr>
<td>Number of data replicas</td>
<td>Customizable</td>
<td>Three replicas by default and not customizable</td>
</tr>
<tr>
<td>Failover or replication</td>
<td>Completed by using a tool</td>
<td>Automatically completed, which is imperceptible to users</td>
</tr>
<tr>
<td>Scaling or upgrade</td>
<td>Completed by using a tool, which affects services</td>
<td>Imperceptible to users</td>
</tr>
<tr>
<td>Write mode</td>
<td>Round robin or key hash</td>
<td>Round robin or key hash</td>
</tr>
<tr>
<td>Current consumption position</td>
<td>Stored in ZooKeeper of the Kafka cluster</td>
<td>Maintained on the server, which does not require your intervention</td>
</tr>
<tr>
<td>Data retention period</td>
<td>Specified in the configuration</td>
<td>Changed dynamically based on requirements</td>
</tr>
</tbody>
</table>
Cost comparison

For more information, see the comparison between LogHub and Kafka in #unique_272.

7.9 What are the differences among log collection agents?

Client evaluation in log collection scenarios

In the data technology (DT) era, hundreds of millions of servers, mobile terminals, and network devices generate a large number of logs every day. The centralized log processing solution effectively meets the log consumption requirements in the lifecycle of log data. Before consuming logs, you need to collect logs from devices and synchronize them to the cloud first.

Three log collection tools

- **Logstash**
  - As a part of the ELK Stack, Logstash is active in the open-source community. It can work with extensive plug-ins in the ecosystem.
  - Logstash is coded in JRuby and can run across platforms on Java virtual machines (JVMs).
  - With a modular design, Logstash features high scalability and interoperability.
- **Fluentd**
  - Fluentd is a popular log collection tool in the open-source community. Its core component, td-agent, is commercially available and maintained by Treasure Data. Fluentd is selected for evaluation in this topic.
  - Fluentd is coded in CRuby. Some key components related to its performance are re-coded in C. The overall performance of Fluentd is excellent.
  - Fluentd features a simple design and provides reliable data transmission in pipelines.
  - Compared with Logstash, Fluentd has fewer plug-ins.

- **Logtail**
  - As the producer of Alibaba Cloud Log Service, Logtail has been tested in big data scenarios for many years in Alibaba Group.
  - Logtail, which is coded in C++, delivers excellent performance in stability, resource control, and management.
  - Compared with Logstash and Fluentd, Logtail obtains less support from the open-source community and focuses more on log collection.

### Feature comparison

<table>
<thead>
<tr>
<th>Feature</th>
<th>Logstash</th>
<th>Fluentd</th>
<th>Logtail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log data read</td>
<td>Polling</td>
<td>Polling</td>
<td>Triggered by event</td>
</tr>
<tr>
<td>File rotation</td>
<td>Supported</td>
<td>Supported</td>
<td>Supported</td>
</tr>
<tr>
<td>Failover processing based on local checkpoints</td>
<td>Supported</td>
<td>Supported</td>
<td>Supported</td>
</tr>
<tr>
<td>General log parsing</td>
<td>Parsing by using Grok based on regular expressions</td>
<td>Parsing by using regular expressions</td>
<td>Parsing by using regular expressions</td>
</tr>
<tr>
<td>Specific log types</td>
<td>Mainstream formats such as delimiter, key-value, and JSON</td>
<td>Mainstream formats such as delimiter, key-value, and JSON</td>
<td>Mainstream formats such as delimiter, key-value, and JSON</td>
</tr>
<tr>
<td>Data compression for transmission</td>
<td>Supported by plug-ins</td>
<td>Supported by plug-ins</td>
<td>LZ4</td>
</tr>
<tr>
<td>Data filtering</td>
<td>Supported</td>
<td>Supported</td>
<td>Supported</td>
</tr>
<tr>
<td>Feature</td>
<td>Logstash</td>
<td>Fluentd</td>
<td>Logtail</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>---------------------------------</td>
<td>--------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Data buffer for transmission</td>
<td>Supported by plug-ins</td>
<td>Supported by plug-ins</td>
<td>Supported</td>
</tr>
<tr>
<td>Transmission exception handling</td>
<td>Supported by plug-ins</td>
<td>Supported by plug-ins</td>
<td>Supported</td>
</tr>
<tr>
<td>Runtime environment</td>
<td>Coded in JRuby and dependent on the JVM environment</td>
<td>Coded in CRuby and C and dependent on the Ruby environment</td>
<td>Coded in C++, without special requirements for the runtime environment</td>
</tr>
<tr>
<td>Thread support</td>
<td>Multithreading</td>
<td>Multithreading restricted by the global interpreter lock (GIL)</td>
<td>Multithreading</td>
</tr>
<tr>
<td>Hot upgrade</td>
<td>Not supported</td>
<td>Not supported</td>
<td>Supported</td>
</tr>
<tr>
<td>Centralized configuration management</td>
<td>Not supported</td>
<td>Not supported</td>
<td>Supported</td>
</tr>
<tr>
<td>Running status self-check</td>
<td>Not supported</td>
<td>Not supported</td>
<td>CPU or memory threshold protection supported</td>
</tr>
</tbody>
</table>

Performance comparison in log collection scenarios

For example, the following Nginx access log contains 365 bytes, from which 14 fields can be extracted:

```
42.120.74.166 37026.1 - [14/Nov/2015:17:50:05 +0800] "POST http://www.xoo.com/auction/order/Id 1051 status 0 ref 0 time 0 ip 1051 length 72 time_offset 0 off 0" 
unity_order_confirm.htm" 200 1152 "http://www.xoo.com/test_now.jhtm" "Mozilla/5.0 (Windows NT 6.1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/28.0.1500.72 Safari/537.36" *316312088* 
"78c97666db9e0b3d555e4df5a2e55" "ac1539681387814767045178.4e" center test local 28374 
```

In the simulated test scenario, this log is repeatedly written at different compression ratios. The time field of each log is set to the current system time when the log is written, and the other 13 fields are the same. Compared with the actual scenario, the simulated scenario has no difference in parsing logs. The only difference lies in that a high data compression ratio can reduce the network traffic on writing data.
Logstash

In Logstash 2.0.0, Logstash parses logs by using Grok and writes the logs to Kafka by using a built-in plug-in that enables GZIP compression.

Log parsing configuration:

```bash
grok {
  patterns_dir="/home/admin/workspace/survey/logstash/patterns"
  match=>{
    "message"=>"%{IPORHOST:ip} %{USERNAME:rt} - \[%{HTTPDATE:
time}\] "%{WORD:method} %{DATA:url}\n"%{NUMBER:status} %{NUMBER:size }
  }"
  remove_field=>["message"]
}
```

The following table lists test results.

<table>
<thead>
<tr>
<th>Write transactions per second (TPS)</th>
<th>Write traffic (Unit: KB/s)</th>
<th>CPU usage (Unit: %)</th>
<th>Memory usage (Unit: MB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>178.22</td>
<td>22.4</td>
<td>427</td>
</tr>
<tr>
<td>1,000</td>
<td>356.45</td>
<td>46.6</td>
<td>431</td>
</tr>
<tr>
<td>5,000</td>
<td>1,782.23</td>
<td>221.1</td>
<td>440</td>
</tr>
<tr>
<td>10,000</td>
<td>3,564.45</td>
<td>483.7</td>
<td>450</td>
</tr>
</tbody>
</table>

Fluentd

In td-agent 2.2.1, Fluentd parses logs by using regular expressions and writes the logs to Kafka by using the third-party plug-in fluent-plugin-kafka that enables GZIP compression.

Log parsing configuration:

```bash
<source>
  type tail
  format /^(? <ip>\S+)\s(? <rt>\d+)\s-\s\[(? <time>\[^\]\]*)\]\s"(? <url>\[^"\]+)"\s(? <status>\d+)\s"(? <ref>\[^"\]+)"\s"(? <agent>\[^"\]+)"\s"(? <cookie_unb>\d+)"\s"(? <cookie_cookie2>\w+)"\s"(? <monitor_traceid>\w+)"\s\s(? <cell>\w+)\s\s(? <ups>\w+)\s\s(? <remote_port>\d+)\s/^
  time_format %d/%b/%Y:%H:%M:%S %z
  path /home/admin/workspace/temp/mock_log/access.log
  pos_file /home/admin/workspace/temp/mock_log/nginx_access.pos
  tag nginx.access
</source>
```

The following table lists test results.
Due to the restrictions of the GIL, a single process of Fluentd uses only one CPU core. You can install the multiprocess plug-in to use multiple processes for achieving a higher log throughput.

Logtail

In Logtail 0.9.4, Logtail uses regular expressions to extract log fields, compresses data by using the LZ4 compression algorithm, and then writes the data to Alibaba Cloud Log Service in compliance with HTTP. The batch_size parameter is set to 4000.

Log parsing configuration:

```plaintext
logRegex : (\S+)\s(\d+)\s-\s\[(\[[^\]]+\]\)]\s"([^"]+)"\s(\d+)\s(\d+)\s"([^"]+)"\s"([^"]+)"\s(\d+)\s(\d+)\s"([^"]+)"\s(\d+)\s(\d+)\s(\d+).* keys : ip,rt,time,url,status,size,ref,agent,cookie_unb,cookie_cookie2,monitor_traceid,cell,ups,remote_port timeformat : %d/%b/%Y:%H:%M:%S
```

The following table lists test results.

<table>
<thead>
<tr>
<th>Write TPS</th>
<th>Write traffic (Unit: KB/s)</th>
<th>CPU usage (Unit: %)</th>
<th>Memory usage (Unit: MB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>178.22</td>
<td>1.7</td>
<td>13</td>
</tr>
<tr>
<td>1,000</td>
<td>356.45</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>5,000</td>
<td>1,782.23</td>
<td>15.3</td>
<td>23</td>
</tr>
<tr>
<td>10,000</td>
<td>3,564.45</td>
<td>31.6</td>
<td>25</td>
</tr>
</tbody>
</table>

Comparison of single-core CPU processing capabilities
The three log collection tools have their own advantages and disadvantages:

- Logstash supports all mainstream log types, the most abundant plug-ins, and flexible customization. However, its performance on log collection is relatively poor, and it requires high memory usage when running in the JVM environment.
- Fluentd supports all mainstream log types and many plug-ins. Its performance on log collection is excellent.
- Logtail occupies the fewest CPU and memory resources of machines, achieves a high performance throughput, and provides comprehensive support for common log collection scenarios. However, it lacks the support of plug-ins, so it is less flexible and scalable than the preceding two tools.

<table>
<thead>
<tr>
<th>Tool</th>
<th>tps</th>
<th>inflow (KB/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>logtail</td>
<td>31600</td>
<td>11263.66</td>
</tr>
<tr>
<td>logstash</td>
<td>2070</td>
<td>737.98</td>
</tr>
<tr>
<td>fluentd</td>
<td>5320</td>
<td>1889.96</td>
</tr>
</tbody>
</table>
7.10 Collect logs in complete regular mode

7.10.1 How do I modify a regular expression?

When configuring Logtail to collect text logs, you must specify a regular expression based on your log sample if you parse and collect logs in full regex mode. This topic describes how to modify a regular expression.

To modify a regular expression that you specified in the Log Service console, you can click Validate to check the following items:

- For the regular expression used to specify the starting header of a cross-line log, check whether the current regular expression can correctly match the expected number of log entries.
- For the regular expression used to extract fields, check whether the value of each field meets your requirements.

If you need to verify more items and modify a regular expression, you can use online tools such as Regex101 and RegexTester. You can copy and paste the regular expression automatically generated in the console to a tool, and then enter actual logs for further verification and modification.

In full regex mode, Log Service automatically generates regular expressions, which may not be suitable for the message field of cross-line logs. This topic describes how to use Regex101 to verify a regular expression.

Procedure

1. Copy the regular expression automatically generated by Log Service based on the log sample.

2. Go to the website of Regex101.
3. In the REGULAR EXPRESSION field, paste the automatically generated regular expression:

```
\[\([(\^[^\]]+)]\s\[(\w+)\s(\[^:\]+:\s\w+\s\w+\s\[^:\]+:\S+\s\[^:\]+:\S+\s\S+)\].
```

The meaning of the regular expression appears in the right pane of the page.

4. In the TEST STRING field, paste part of the log sample.

In the following figure, some content after `at` is not included in the `message` field. The log entries included and not included in the message field are highlighted in orange and blue, respectively. Therefore, this regular expression does not fully match the log sample. That is, this regular expression is incorrect for the log sample and cannot be used to collect all required log data.
5. Verify another error: The entered log sample contains only two colons.

In the following figure, the regular expression fails to match the log sample.

![Image of a regular expression editor showing a fails case](image1)

6. Replace the last element in the regular expression with `\[\S\s]+` and check whether the regular expression matches the log sample.

In the following figure, the regular expression matches the content after `at`.

![Image of a regular expression editor showing a successful match](image2)

In the following figure, the regular expression matches the log sample that contains only two colons.

![Image of a regular expression editor showing a successful match](image3)

You can use the preceding method to specify and modify your regular expression and apply it to a Logtail configuration.
7.10.2 How do I optimize regular expressions?

You can optimize regular expressions to improve the Logtail collection performance.

The following describes some suggestions about how to optimize regular expressions:

- Use precise characters.

  Do not arbitrarily use .* to match fields because this regular expression can match with a wide range of search results. Such actions can lead to mismatches occurring or a decrease in matching performance. For example, to return results of fields that consist only of letters, use [A-Za-z].

- Use correct measure words.

  Do not arbitrarily use plus signs (+), commas (,), or asterisks. For example, to match target IP addresses, use \d instead of \d+ or \d{1,3} because of its higher efficiency.

- Debug multiple times.

  Debugging is similar to troubleshooting. You can debug the time consumed by your regular expressions at the Regex101 website, and promptly optimize them if there is a large amount of backtracking.

7.10.3 How do I collect various formats of logs in complete regular mode?

The complete regular mode requires format consistency among all logs. However, some logs may contain content in multiple formats. In this case, you can use the Schema-On-Write or Schema-On-Read mode to process the logs.

For example, a Java log is a program log that contains both correct information and error information (such as information about abnormal stacks), including:

- Multi-line WARNING logs
- Simple text INFO logs
- Key-value DEBUG logs

```java
[2018-10-01T10:30:31,000] [WARNING] java.lang.Exception: another exception happened
  at TestPrintStackTrace.f(TestPrintStackTrace.java:3)
  at TestPrintStackTrace.g(TestPrintStackTrace.java:7)
  at TestPrintStackTrace.main(TestPrintStackTrace.java:16)
[2018-10-01T10:32:000] [INFO] info something
```
You can use the following modes to process such logs:

- **Schema-On-Write**: In this mode, Logtail applies multiple Logtail Configs with different regular settings to a log so that correct fields can be extracted.

  **Note:**
  Logtail cannot apply multiple Logtail Configs to a file. Therefore, you need to set up multiple symbolic links for the directory where the file is stored. Then, each Logtail Config works on a symbolic link, thereby allowing you to aggregate multiple Logtail Configs to collect the file at the same time.

- **Schema-on-read**: In this mode, you need to use the common regular expressions of the multi-format logs.

  For example, for collection of a multi-line log, you can use the time and log level as line start regular expressions and the residual parts as the message field. If you want to analyze the message field, you can set up an index for it, extract the required content, and then analyze the content based on query and analysis functions provided by Log Service, such as regular expression extraction.

  **Note:**
  This mode is recommended only when you need to analyze at a small-scale (for example, tens of millions) of logs.

### 7.11 Why am I unable to collect SLB access logs?

This topic describes how to troubleshoot in cases where you are unable to collect SLB access logs.

1. Check whether the access log collection function has been activated for SLB instances.

   **Activate the access log collection function for each SLB instance separately.** Then, the generated access logs can be written into your Logstore in real time.

   **To do so,** log on to the SLB console, and choose Logs > Access Logs to view the Access Logs (Layer-7) list.

   - **Verify that the specified SLB instance exists.**
• Confirm the Storage Path of the SLB instance.

This column displays information about the project and Logstore. In this case, make sure that you check whether SLB logs exist in the correct location in the console.

2. Check whether RAM users are correctly authorized.

During activation of the access log collection function, the system guides you through RAM user authorization. The function can be successfully activated only after RAM users are successfully authorized. If RAM users are incorrectly created or deleted, the collected logs cannot be delivered to your Logstore.

Troubleshooting

Log on to the RAM Console. On the RAM Roles page, check whether the AliyunLogArchiveRole role exists.

• If AliyunLogArchiveRole does not exist, use your Alibaba Cloud account to log on to the RAM console and click the quick authorization link to create the RAM users required for authorization.
• If AliyunLogArchiveRole exists, click the role name and check whether the role is correctly authorized.

The following shows the default policy. If your policy has been modified, we recommend that you replace the current policy with the default policy.

```json
{
  "Version": "1",
  "Statement": [
    {
      "Action": [
        "log:PostLogStoreLogs"
      ],
      "Resource": "*",
      "Effect": "Allow"
    }
  ]
}
```

3. Check whether any log is generated.

If you do not find any SLB access log in the Log Service console, it is likely that no log has been generated. Possible causes include:
Layer-7 listening is not configured for the current instance.

Currently, only instances configured with layer-7 listening are supported. Common layer-7 listening protocols include HTTP and HTTPS. For more information, see #unique_278.

Historical logs that were generated before activation of the access log collection function are not collected.

Instead, only logs that are generated after activation of the access log collection function are collected.

The specified instance did not receive a request.

Logs are generated only when you request access to the listener of the instance.

7.12 How do I set the time format?

This topic describes how to set the time format for Logtail Configs and the precautions you need be aware of first.

- The minimum granularity that you can configure for timestamps in Log Service is seconds.
- In the time field, only the front part that contributes to time parsing is required.

The following shows a setting example:

<table>
<thead>
<tr>
<th>Custom1</th>
<th>2017-12-11 15:05:07</th>
</tr>
</thead>
<tbody>
<tr>
<td>%Y-%m-%d %H:%M:%S</td>
<td></td>
</tr>
<tr>
<td>Custom2</td>
<td>[2017-12-11 15:05:07.012]</td>
</tr>
<tr>
<td>[%Y-%m-%d %H:%M:%S</td>
<td></td>
</tr>
<tr>
<td>RFC822</td>
<td>02 Jan 06 15:04 MST</td>
</tr>
<tr>
<td>%d %b %y %H:%M</td>
<td></td>
</tr>
<tr>
<td>RFC822Z</td>
<td>02 Jan 06 15:04 -0700</td>
</tr>
<tr>
<td>%d %b %y %H:%M</td>
<td></td>
</tr>
<tr>
<td>RFC850</td>
<td>Monday, 02-Jan-06 15:04:05 MST</td>
</tr>
<tr>
<td>%A, %d-%b-%y %H:%M:%S</td>
<td></td>
</tr>
<tr>
<td>RFC1123</td>
<td>Mon, 02 Jan 2006 15:04:05 MST</td>
</tr>
<tr>
<td>%A, %d-%b-%y %H:%M:%S</td>
<td></td>
</tr>
<tr>
<td>RFC3339</td>
<td>2006-01-02T15:04:05Z07:00</td>
</tr>
<tr>
<td>%Y-%m-%dT%H:%M:%S</td>
<td></td>
</tr>
<tr>
<td>RFC3339Nano</td>
<td>2006-01-02T15:04:05.999999999Z07:00</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
7.13 Troubleshoot log collection exceptions in containers

This topic provides solutions to exceptions that may occur when you use a Logtail container (a common container or Kubernetes) to collect logs.

Troubleshooting operations:

- Troubleshoot heartbeat exceptions in a machine group
- Troubleshoot log collection exceptions in a container

Other O&M operations:

- Log on to the Logtail container
- View Logtail operational logs
- View Logtail standard output (stdout)
- View the status of log-related components in a Kubernetes cluster
- View the version information, IP address, and time of Logtail

What do I do if I mistakenly delete a Logstore that is created through CRD?

Troubleshoot heartbeat exceptions in a machine group

You can determine whether the Logtail on a container is correctly installed by checking the heartbeat status of a machine group.

1. Check the heartbeat status of the machine group.

   a. Log on to the Log Service console, and then click the target project name.
   b. In the left-side navigation pane, click Logtail Machine Group.
   c. Find the target machine group and click Status.

   Record the number of nodes for which heartbeat status is OK.

2. Check the number of Worker nodes in the cluster.

   Run `kubectl get node | grep -v master` to view the number of Worker nodes.

<table>
<thead>
<tr>
<th>NAME</th>
<th>STATUS</th>
<th>ROLES</th>
<th>AGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>cn-hangzhou.i-bp17enxc2us3624wexh2v1.10.4</td>
<td>Ready</td>
<td>&lt;none&gt;</td>
<td>238d</td>
</tr>
</tbody>
</table>
3. Compare whether the number of the nodes with heartbeat status of OK is the same as the number of Worker nodes. Then, use an appropriate troubleshooting method according to the following possible comparison results:

- The heartbeat status of all nodes is Failed.
  - If you use standard Docker logs, check whether `${your_region_name}`, `${your_aliyun_user_id}`, and `${your_machine_group_user_defined_id}` are correct by following the instructions provided in parameter description.
  - If you use installation for Kubernetes on Alibaba Cloud Container Service, open a ticket.
  - If you use self-built Kubernetes installation, check whether `{your-project-suffix}`, `{regionId}`, `{aliuid}`, `{access-key-id}`, and `{access-key-secret}` are correct by following the instructions provided in parameter description. If the parameters are incorrect, run `helm del --purge alibaba-log-controller` to delete the installation package and reinstall Kubernetes.

- The number of nodes for which the heartbeat status is OK is smaller than the number of Worker nodes.
  a. Determine whether to use the yaml file to manually deploy DaemonSet.
   
   Run `kubectl get po -n kube-system -l k8s-app=logtail`. If any result is returned, you have manually deployed DaemonSet by using the yaml file.
  b. Download the latest DaemonSet template.
  c. Set `${your_region_name}`, `${your_aliyun_user_id}`, and `${your_machine_group_name}` as needed.
  d. Run `kubectl apply -f ./logtail-daemonset.yaml` to update the DaemonSet yaml file.

For other comparison results, open a ticket.

Troubleshoot log collection exceptions in a container

If you cannot find any log on the preview or query page in the console, Log Service has not collected any log from your container. In this case, check the container status and perform the following steps:

1. Check whether the machine group status is normal.
2. Check whether the Config identifier is correct.

**Check whether** IncludeLabel, ExcludeLabel, IncludeEnv, and ExcludeEnv in the Config match the configurations of the target container.

**Note:**
Label indicates the container label (label information in `docker inspect`) instead of the one defined in Kubernetes. You can temporarily remove the parameters and check whether any log can be collected. If yes, the exception is caused by an incorrect Config identifier.

3. Check other items.

If you want to collect files from your container, note that:

- Logtail does not collect any file if there are no modified files in your container.
- Only the files that are stored by default or mounted to your local PC can be collected.

**Log on to the Logtail container**

- **Common Docker**
  
  1. **On the host**, run `docker ps | grep logtail` to search for the Logtail container.
  2. **Run** `docker exec -it ****** bash` to log on to the Logtail container.

  ```
  docker ps | grep logtail
  223fbd3ed2a6e  registry.cn-hangzhou.aliyuncs.com/log-service/logtail /usr/local/ilogtail 8 days ago          Up 8 days                               logtail-iba
  docker exec -it 223fbd3ed2a6e bash
  ```

  - **Kubernetes**
  
  1. **Run** `kubectl get po -n kube-system | grep logtail` to search for the Logtail Pod.
  2. **Run** `kubectl exec -it -n kube-system ****** bash` to log on to the Pod.

  ```
  kubectl get po -n kube-system | grep logtail
  logtail-ds-g5wgd             1/1
  logtail-ds-slpn8             1/1
```
$kubectl exec -it -n kube-system logtail-ds-g5wgd bash

View Logtail operational logs

**Logtail logs named** ilogtail.LOG and logtail_plugin.LOG are stored in the /usr/local/ilogtail/ directory.

1. **Log on to the Logtail container.**

2. **Open the /usr/local/ilogtail/ directory.**

```
cd /usr/local/ilogtail
```

3. **View the ilogtail.LOG and logtail_plugin.LOG files.**

```
cat ilogtail.LOG
cat logtail_plugin.LOG
```

View Logtail standard output (stdout)

You can ignore the following stdout because the container stdout has no reference for application.

```
start umount useless mount points, /shm$|/merged$|/mqueue$
umount: /logtail_host/var/lib/docker/overlay2/3fd0043af174cb0273c3176950f8e2b2dd95d130e110172ef57fe840c82155/merged: must be superuser to unmount
umount: /logtail_host/var/lib/docker/overlay2/1f85d25009528d9a749c1bf8c16edff44beab6e69718/merged: must be superuser to unmount
umount: /logtail_host/var/lib/docker/overlay2/5c3125daddacedec29df72ad0c52fac800cd56c6e880dc4e8a640b1e16c22db8/merged: must be superuser to umount
......
xargs: umount: exited with status 255; aborting
umount done
start logtail
ilogtail is running
logtail status:
ilogtail is running
```

View the status of log-related components in a Kubernetes cluster

To view the status of log-related components in a Kubernetes cluster, you can run

```
helm status alibaba-log-controller.
```

View the version information, IP address, and time of Logtail

The related information is stored in the app_info.json file under the /usr/local/ilogtail/ directory in the Logtail container. The following is an example:

```
kubectl exec logtail-ds-gb92k -n kube-system cat /usr/local/ilogtail/app_info.json
```
What do I do if I mistakenly delete a Logstore that is created through CRD?

If you delete a Logstore that is automatically created through CRD, the collected data cannot be recovered, and the CRD configurations of the Logstore become invalid. In this case, you can use either of the following methods to prevent possible log collection exceptions:

- Use another CRD-created Logstore and take care to name the Logstore with a different name to the Logstore that was mistakenly deleted.

- Restart the alibaba-log-controller Pod. You can run `kubectl get po -n kube-system | grep alibaba-log-controller` to search for the Pod.