Alibaba Cloud Log Service

Data Collection

Issue: 20190816

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, consequential, 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 content of the Alibaba Cloud website, including but not limited to works, products, images, archives, information, materials, website architecture, website graphic layout, and webpage design, 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 the Alibaba Cloud website, product programs, or content 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.

Generic conventions

Table -1: Style conventions

Style	Description	Example
	This warning information indicates a situation that will cause major system changes, faults, physical injuries, and other adverse results.	 Danger: Resetting will result in the loss of user configuration data.
	This warning information indicates a situation that may cause major system changes, faults, physical injuries, and other adverse results.	 Warning: Restarting will cause business interruption. About 10 minutes are required to restore business.
	This indicates warning information, supplementary instructions, and other content that the user must understand.	 Notice: Take the necessary precautions to save exported data containing sensitive information.
	This indicates supplemental instructions, best practices, tips, and other content that is good to know for the user.	 Note: You can use Ctrl + A to select all files.
>	Multi-level menu cascade.	Settings > Network > Set network type
Bold	It is used for buttons, menus , page names, and other UI elements.	Click OK.
`Courier font`	It is used for commands.	Run the `cd / d C :/ windows` command to enter the Windows system folder.
Italics	It is used for parameters and variables.	`bae log list --instanceid` *Instance_ID*
[] or [a|b]	It indicates that it is a optional value, and only one item can be selected.	`ipconfig` *[-all|-t]*

Style	Description	Example
{} or {a|b}	It indicates that it is a required value, and only one item can be selected.	`swich {stand | slave}`

Contents

1 Collection methods

LogHub supports a variety of RESTful APIs that provide different log collection methods, for example, log collection through one or more clients, websites, protocols , SDKs, and APIs.

Data sources

Log Service can collect logs from the following sources:

Type	Source	Access method	Details
Application	Program output	Logtail	-
	Access logs	Logtail	#unique_5
	Link track	Jaeger Collector and Logtail	-
Programming language	Java	SDK and Java Producer Library	-
	Log4J Appender	1.x and 2.x	-
	LogBack Appender	LogBack	-
	C	Native	-
	Python	Python	-
	Python Logging	Python Logging Handler	-
	PHP	PHP	-
	C#	C#	-
	C++	C++ SDK	-
	Go	Go	-
	NodeJS	NodeJs	-
	JS	JS/Web Tracking	-
OS	Linux	Logtail	-
	Windows	Logtail	-
	Mac/Unix	Native C	-
	Docker files	Logtail file collection	-

Type	Source	Access method	Details
	Docker output	Logtail container stdout	-
Mobile client	iOS/Android	#unique_11 and #unique_12	-
	Websites	JS/Web Tracking	-
	Intelligent IoT	C Producer Library	-
Cloud product	Various products, such as ECS and OSS For more information, see Cloud product logs	Cloud product console	Cloud product logs
	Import MaxCompute data	Use Dataworks to export MaxCompute data	#unique_15
Third-party software	Logstash	Logstash	-

The following table lists the cloud products from which Log Service can collect logs:

Network and access point selection

Log Service provides service endpoints in each region, and the following types of network access methods are supported:

- (Recommended) Intranet (classic networks) and private networks (VPCs): are applicable to regions with smooth service access and high-quality bandwidth links.
- Internet (classic networks): can be used without any limits. The access speed depends on the link quality. HTTPS is recommended to maintain access security.

FAQ

- Q: Which type of network applies to physical connections?

  A: Intranet/private networks
- Q: Can Internet IP addresses be collected during Internet data collection?

  A: Yes. You can follow the instructions provided in #unique_18 and enable the Internet IP address recording function.

- Q: Which type of network can I use if I want to collect logs from an ECS server located in region A and send them to a project on a Log Service server located in region B?

  A: Use the Internet to transfer logs after install the Internet-version Logtail on the ECS server. As for other scenarios, follow the instructions provided in #unique_19.
- Q: How can I determine whether access is established successfully?

  A: Access is established successfully if information is returned after you run the following command:

  ```
  curl $myproject.cn-hangzhou.log.aliyuncs.com
  ```

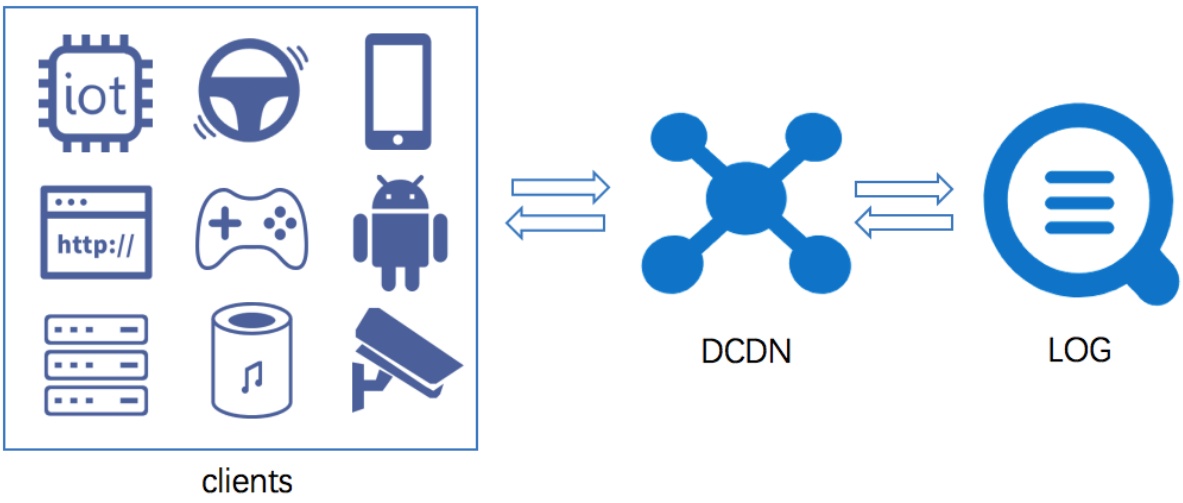
  In this command, `$myproject` indicates the project name, and `cn-hangzhou.log.aliuncs.com` indicates the access point.

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.

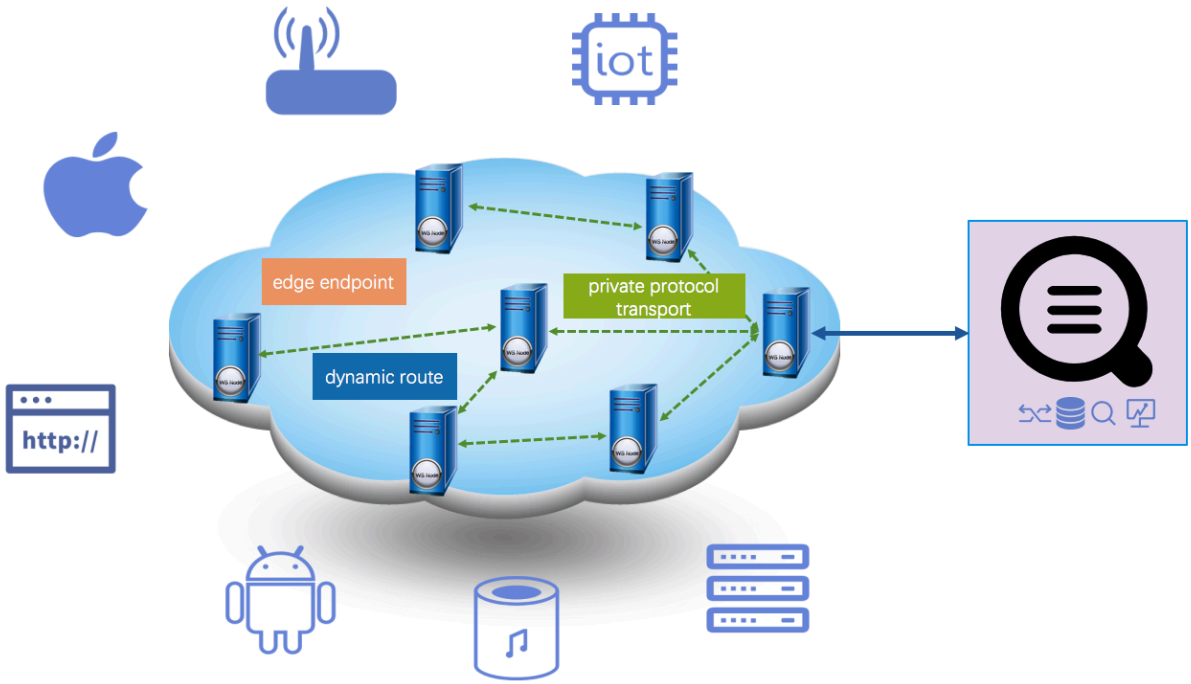
Global Acceleration for Log Service is based on Alibaba Cloud Content Delivery Network (CDN) hardware resources. Alibaba Cloud has optimized the stability of log collection and data transmission from various forms of data sources, such as mobile phones, Internet of Things (IoT) devices, smart devices,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

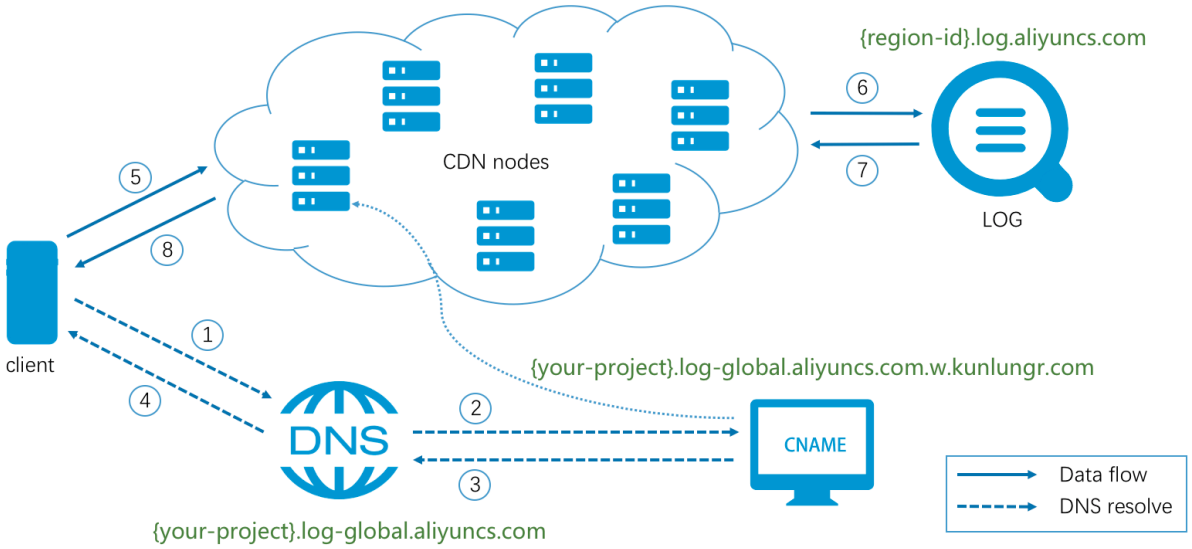
premises IDCs, and other cloud servers) can access the nearest edge node of Alibaba Cloud CDN all over the world and be routed to Log Service through the inner high-speed channels of CDN. Compared with data transmission on the Internet, this method can greatly reduce the network latency and jitter.



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_22.

- 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 Accelerati on 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

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

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
- 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 Global Acceleration for Log Service.

Prerequisites

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

Procedure

After you enable HTTP Global Acceleration for the target project, you can also configure the Logtail, SDK, and other methods as required to collect logs by using Global Acceleration.

1. Enable HTTP acceleration.
2. (Optional) Enable HTTPS acceleration.

   If you use HTTPS to access Log Service, ensure that HTTPS acceleration is enabled. For more information about how to configure HTTPS acceleration, see Enable HTTPS acceleration.
3. Collect logs by using Global Acceleration.
   - Logtail
     - To install Logtail after Global Acceleration is enabled, you can follow the installation procedure for Global Acceleration in #unique_26. Then, Global

Acceleration is automatically enabled when Log Service collects logs through Logtail.

- If Logtail is installed before you enable Global Acceleration, you need to manually switch the Logtail collection mode to global acceleration. For more information, see #unique_27.

· SDK, Producer, or Consumer

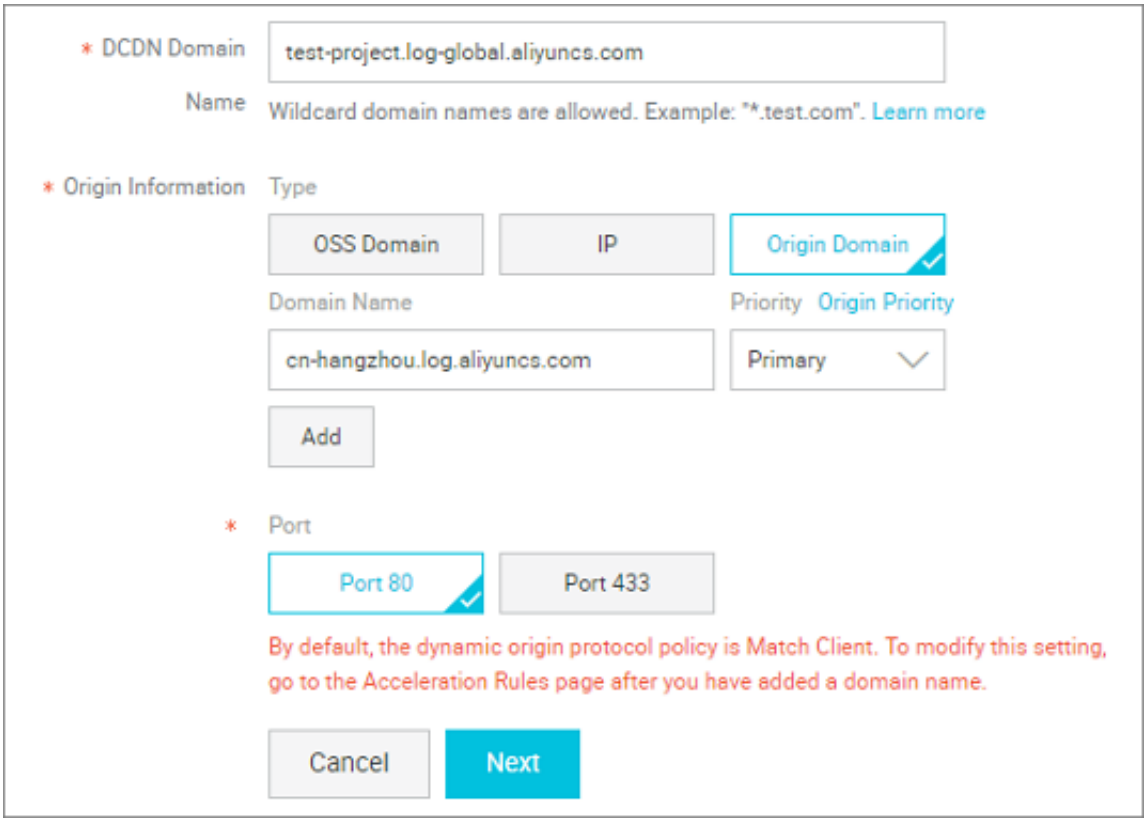
If you use other methods such as the SDK, Producer, and Consumer to access Log Service, you can replace the configured endpoint with `log-global.aliyuncs.com` to achieve global acceleration.

Enable HTTP acceleration

1. Log on to the Dynamic Route for CDN console. In the left-side navigation pane, click Domain Names to go to the Domain Names page.
2. Click Add Domain Name in the upper-left corner to go to the Add Domain Name page.
3. Set `DCDN Domain Name`, enter other information as required, and then click Next.

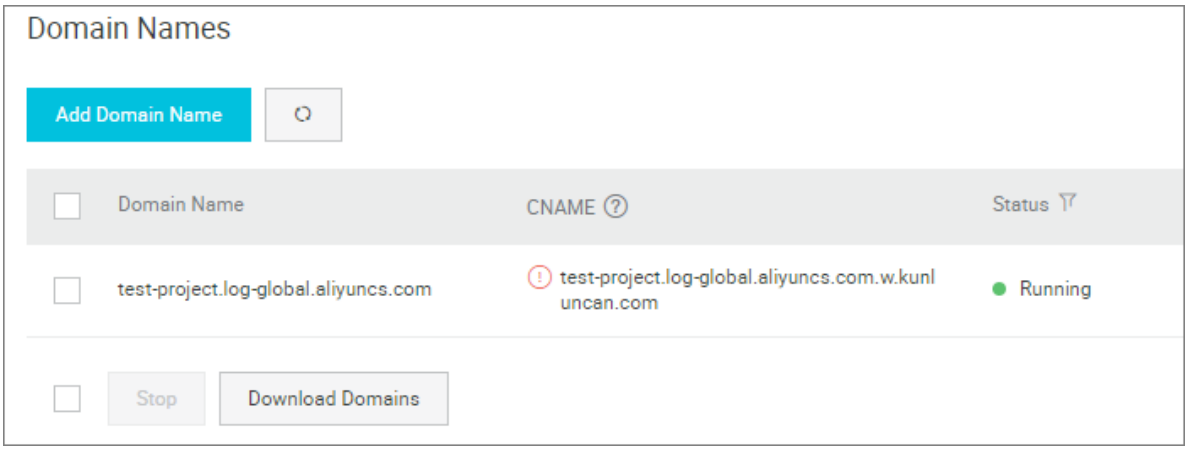
Parameter	Description
DCDN Domain Name	Enter *project_name*`.log-global.aliyuncs.com`, where *project_name* is replaced with your project name.
Origin Type	Select `Origin Domain`.
Domain Name	Enter the Internet service endpoint for the region of your project. For more information about endpoints, see #unique_28.
Port	Select Port 80. If you need HTTPS acceleration, you can configure HTTPS separately. For more information, see Enable HTTPS acceleration in this topic.

Parameter	Description
Acceleration Region	`Mainland China` is selected by default. If you need to use Global Acceleration, open a ticket to apply for a whitelist from Dynamic Route for CDN. After your application is approved, you can select an acceleration region based on your needs.

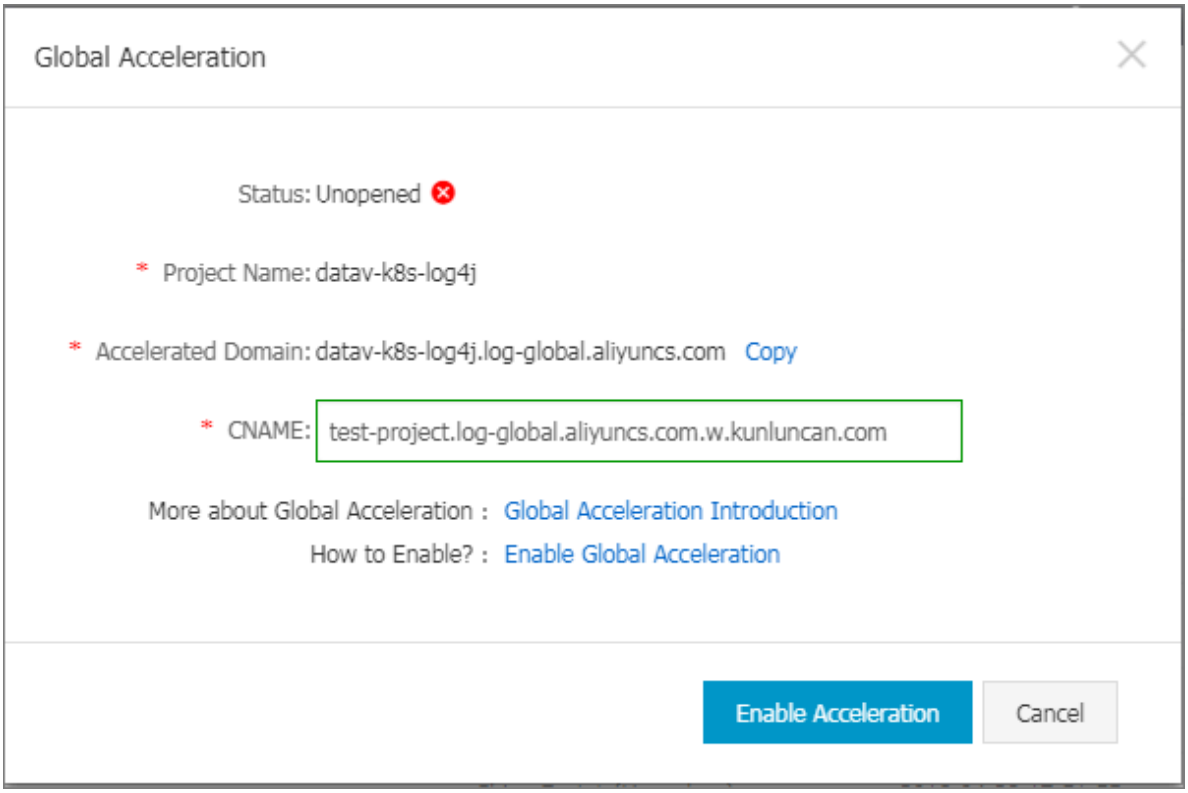


4. Go to the Domain Names page as prompted.

   You can view the CNAME of the added accelerating domain name on the Domain Names page.



5. Log on to the Log Service console. On the Projects page, click Global Acceleration in the Actions column of the target project.
6. In the dialog box that appears, enter the CNAME of the accelerating domain name. Click Enable Acceleration.



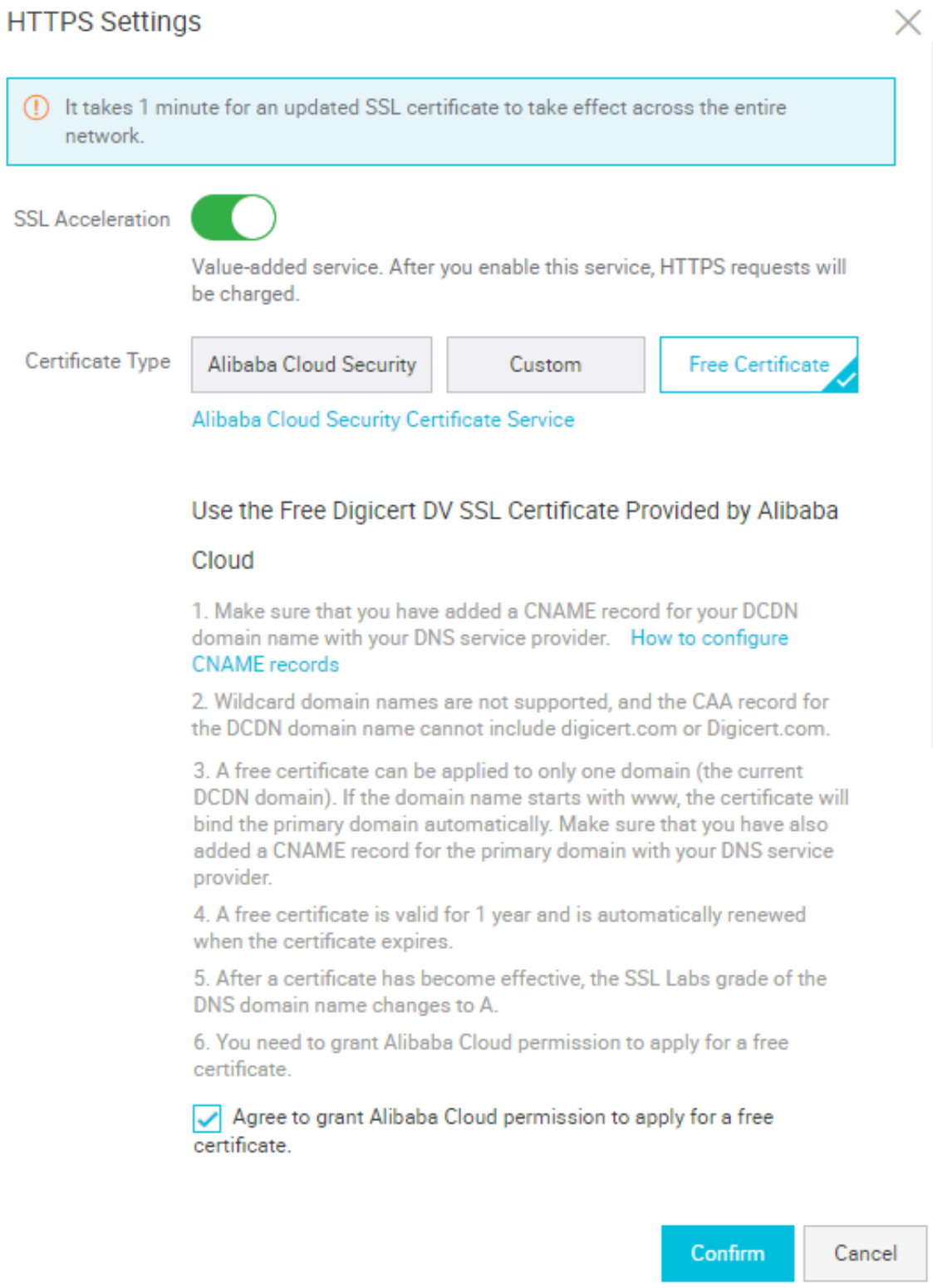
After you complete the preceding steps, Global Acceleration is enabled for Log Service.

Enable HTTPS acceleration

After HTTP acceleration is enabled, if you need to use HTTPS to access Log Service, you can enable HTTPS acceleration. The procedure is as follows:

1. Log on to the Dynamic Route for CDN console. In the left-side navigation pane, click Domain Names to go to the Domain Names page.
2. Click Configure in the Actions column of 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.
   - Enable SSL Acceleration.
   - Select Free Certificate for Certificate Type.



After the configuration is completed, 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 your accelerating domain name to check whether the acceleration configuration takes effect.

  For example, if Global Acceleration is enabled for the `test-project` project, you can run a `curl` command to send a request to the accelerating domain name. The following response indicates that the acceleration configuration takes effect:

```
$ curl test-project.log-global.aliyuncs.com
{"Error":{"Code":"OLSInvalid Method","Message":"The
 script name is invalid: /","RequestId":"5B55386A2C
 E41D1F4FBC F7E7"}}
```

- What can I do if the `project not exist` error is reported for access to an accelerating domain name?

  This error is caused usually by an invalid origin domain name. You need to log on to the Dynamic Route for CDN console and change the origin domain name of the accelerating domain name to the Internet service endpoint for the region of your project. For more information, see #unique_28.

  Note:
  The change of the origin domain name takes several minutes. You can wait patiently.

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. #unique_30/unique_30_Connect_42_section_sst_dsz_q2b.

2. (Optional) #unique_30/unique_30_Connect_42_section_sst_dsz_q2b.

   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 #unique_30/unique_30_Connect_42_section_sst_dsz_q2b.
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 #unique_31. 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:
     a. In Control Panel, choose System and Security > Administrative Tools.
     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_c onfig . json* .

   Change the endpoint in `data_serve r_list` to `log - global . aliyuncs . com` by following the instructions provided in #unique_32/unique_32_Connect_42_section_jh3_dpk_2fb.

3. Start the Logtail.
   - In Linux, run `/etc/init.d/ilogtaild start` as the admin user.
   - In Windows:
     a. In Control Panel, choose System and Security > Administrative Tools.
     b. Open the Services program and locate the LogtailWorker file.
     c. Right-click the file and click Start in the shortcut menu.

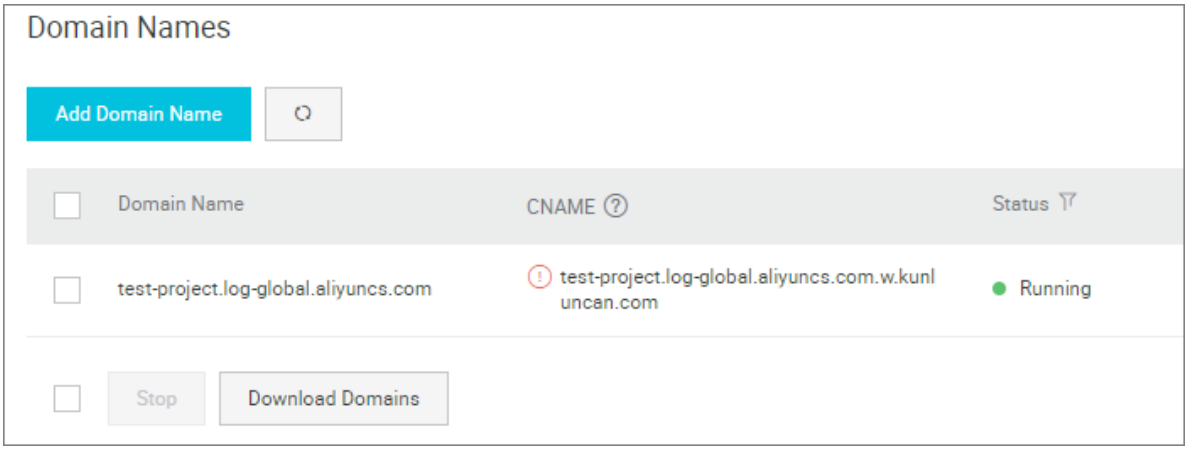
2.4 Disable Global Acceleration

To disable Global Acceleration for Log Service, perform the following operations.

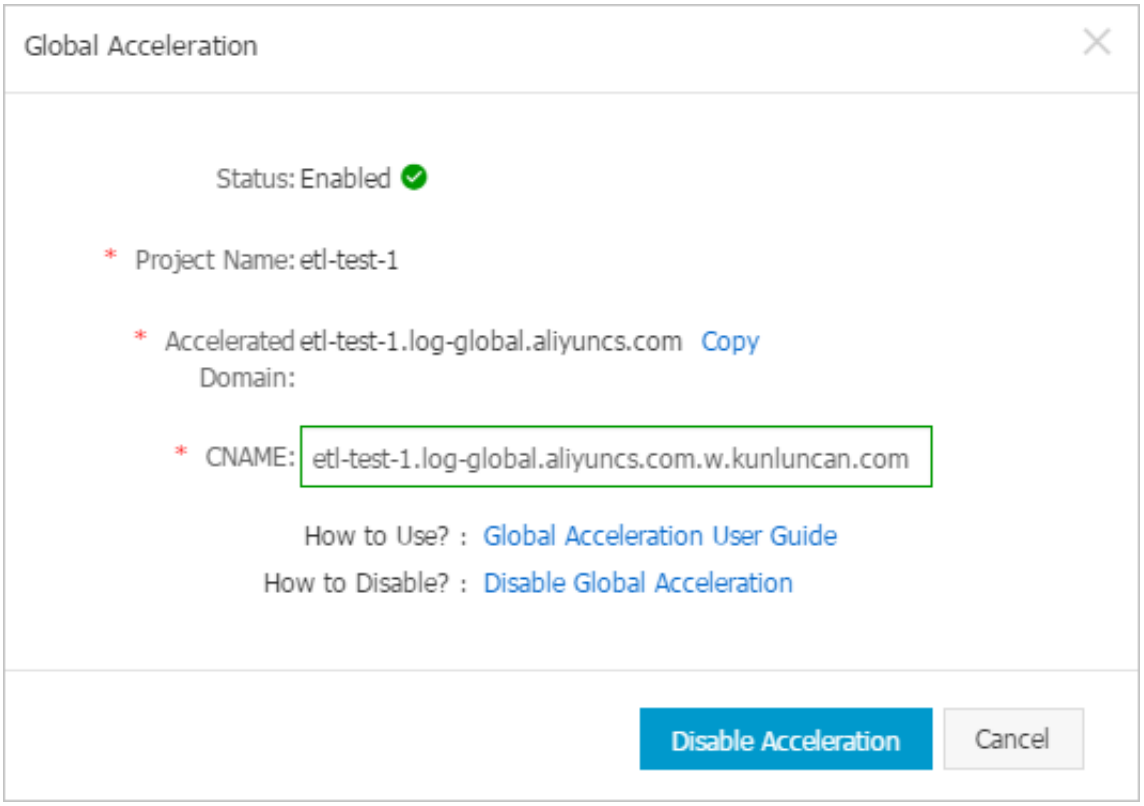
> **Note:**
> When you disable Global Acceleration, the accelerated domain name configured during provisioning becomes unavailable. Make sure that all of your clients do not upload or request data through the domain name before you disable Global Acceleration.

Disable Global Acceleration

1. Log on to the Dynamic Route for CDN Console. Click Domain name management in the left-side navigation pane to enter the Domain name management page.
2. View the CNAME corresponding to the domain name that is to be disabled .



3. Log on to the Log Service console. On the Project list page, click Global Acceleration at the right of a specified project.
4. Enter CNAME and click Disable acceleration.


Global Acceleration
Status: Enabled
* Project Name: etl-test-1
* Accelerated Domain: etl-test-1.log-global.aliyuncs.com Copy
* CNAME: etl-test-1.log-global.aliyuncs.com.w.kunluncan.com
How to Use? : Global Acceleration User Guide
How to Disable? : Disable Global Acceleration
Disable Acceleration
Cancel

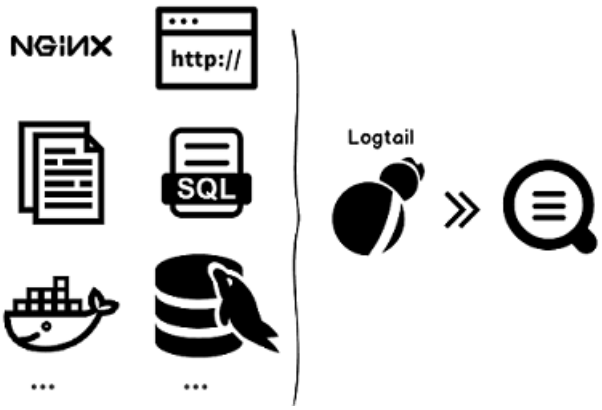
3 Logtail collection

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 #unique_37 and #unique_31.

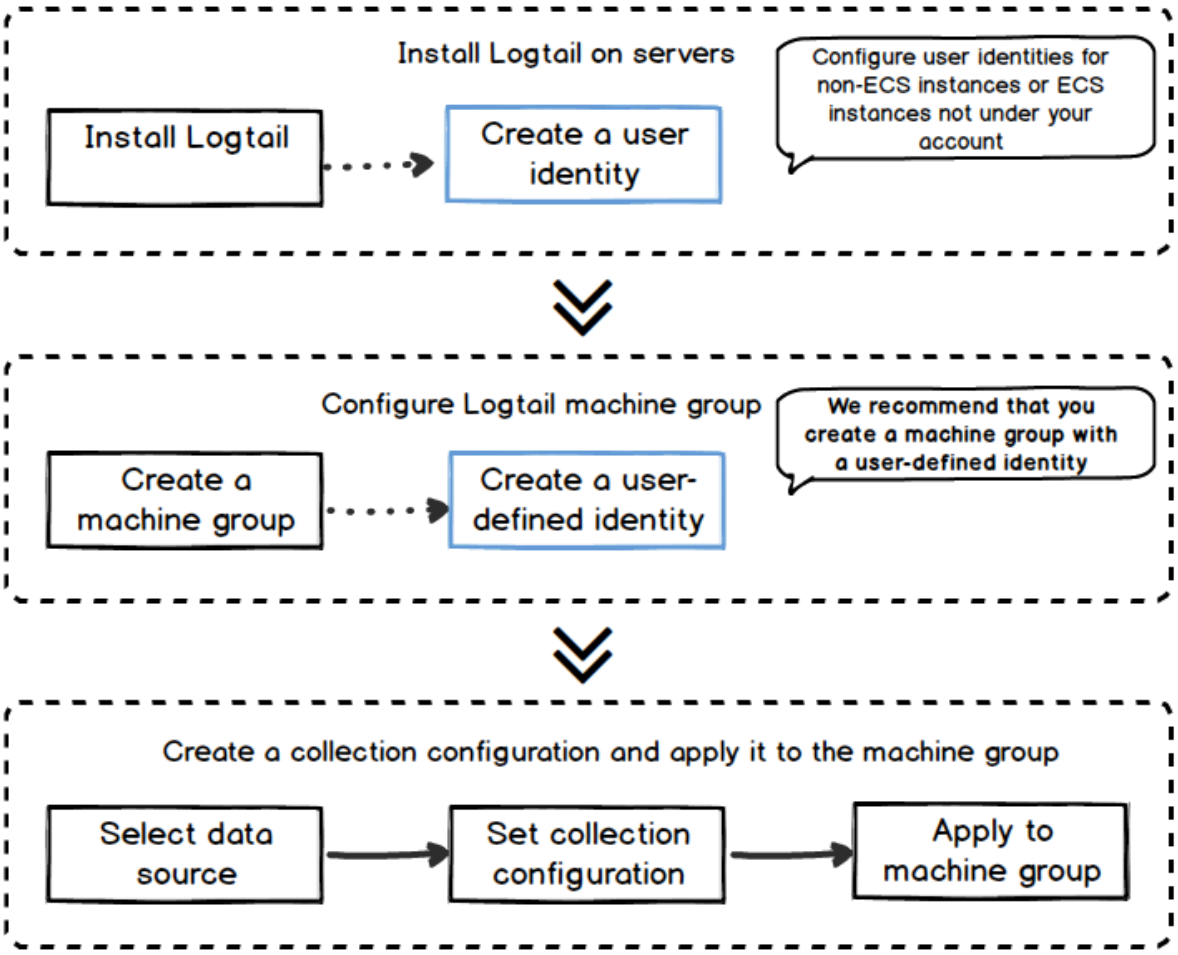
- Comprehensive self-protection mechanism. To make sure that the collection agent running on your machine does not significantly affect the performance of your services, the Logtail client strictly protects and limits the usage of CPU, memory, and network resources.

Processing capabilities and limits

See #unique_38.

Procedure

Figure 3-2: Configuration process



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 #unique_37 and #unique_31

2. #unique_39. Skip this step if you are about to collect logs from Alibaba Cloud ECS instances
3. #unique_40. 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 Collect text logs and #unique_42 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_43.
- Alibaba Cloud Container Service Kubernetes cluster: see #unique_44.
- Self-built Kubernetes: see#unique_45/unique_45_Connect_42_section_kdx_bqr_zdb
- Other self-built Docker clusters: see#unique_46

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 #unique_37 and #unique_31. 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 exceptionsWhen 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 #unique_55.

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.

 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_reque st_concurr ency` by following the instructions provided in #unique_56 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.

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

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_ser ver_addres s` and `data_serve r_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 #unique_47.

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

Configuration item	Description
config_server_address	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.

Configuration item	Description
data_server_list	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.
cluster	Region name
endpoint	Service endpoint
cpu_usage_limit	CPU usage threshold, which is calculated by core
mem_usage_limit	Resident memory usage threshold
max_bytes_per_sec	Maximum amount of raw data Logtail can send. The amount will not be limited if the data sending rate exceeds 20 Mbit/s.
process_thread_count	Number of threads Logtail uses to write data to log files
send_request_concurrency	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.

File address:

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

File example:

```
$ cat  / usr / local / ilogtail / ilogtail_c  onfig . json
{
    " config_ser  ver_addres  s " : " http :// logtail . cn - hangzhou
 - intranet . log . aliyuncs . com ",
    " data_serve  r_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_th  read_count " :  1 ,
    " send_reque  st_concurr  ency " :  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 #unique_65.

 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_LOG  TAIL_USER_  ID` in the Logtail container. You can view the file through `docker  inspect  ${ logtail_co  ntainer_na  me } |  grep  ALIYUN_LOG  TAIL_USER_  ID`.

- Windows: `C :\ LogtailDat a \ users \`

File example

```
$ ls / etc / ilogtail / users /
 1559122535 02 **** 1329232535 02 ****
```

User-defined identity file (user_defined_id)

The file is used to configure machine groups with custom identifiers. For more information, see #unique_39.

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_defin ed_id*
- Container: The file is directly configured through the environment variable `ALIYUN_LOG TAIL_USER_ DEFINED_ID` in the Logtail container. You can view the file through *docker inspect ${ logtail_co ntainer_na me } | grep ALIYUN_LOG TAIL_USER_ DEFINED_ID .*
- Windows: *C :\ LogtailDat a \ user_defin ed_id*

File example

```
$ cat / etc / ilogtail / user_defin ed_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_c onfig . json*
- **Container:** */ usr / local / ilogtail / user_log_c onfig . json*
- **Windows**
  - **x64:** *C :\ Program Files (x86)\ Alibaba \ Logtail \ user_log_c onfig . json*
  - **x32:** *C :\ Program Files \ Alibaba \ Logtail \ user_log_c onfig . json*

File example

```
$ cat  / usr / local / ilogtail / user_log_c  onfig . json
{
   " metrics " : {
      "## 1 . 0 ## k8s - log - c12ba2028 ***** 939f0b $ app - java " :
 {
         " aliuid " : " 16542189 ***** 50 ",
         " category " : " app - java ",
         " create_tim  e " :  1534739165 ,
         " defaultEnd  point " : " cn - hangzhou - intranet . log .
 aliyuncs . com ",
         " delay_alar  m_bytes " :  0 ,
         " enable " :  true ,
         " enable_tag " :  true ,
         " filter_key  s " : [],
         " filter_reg  s " : [],
         " group_topi  c " : "",
         " local_stor  age " :  true ,
         " log_type " : " plugin ",
         " log_tz " : "",
         " max_send_r  ate " : - 1 ,
         " merge_type " : " topic ",
         " plugin " : {
            " inputs " : [
               {
                  " detail " : {
                     " IncludeEnv " : {
                        " aliyun_log  s_app - java " : " stdout "
                     },
                     " IncludeLab  le " : {
                        " io . kubernetes . container . name " : "
 java - log - demo - 2 ",
                        " io . kubernetes . pod . namespace " : "
 default "
                     },
                     " Stderr " :  true ,
                     " Stdout " :  true
                  },
                  " type " : " service_do  cker_stdou  t "
               }
            ]
         },
```

```
        " priority " :  0 ,
        " project_na  me " : " k8s - log - c12ba2028c *****
 ac1286939f  0b ",
        " raw_log " :  false ,
        " region " : " cn - hangzhou ",
        " send_rate_  expire " :  0 ,
        " sensitive_  keys " : [],
        " tz_adjust " :  false ,
        " version " :  1
      }
   }
}
```

AppInfo recording file (app_info.json)

The file contains various time information, such as the Logtail startup time and the time when Logtail obtains the IP address and host name. The IP address is needed when you configure machine groups with IP addresses as identifiers.

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 3-2: Field description

Field	Description
UUID	Server serial number
hostname	Host name
instance_id	Randomly generated identifier for uniquely indicating Logtail

Field	Description
ip	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. **Note:** If the target machine group uses an IP address as an identifier, the IP address configured in the machine group must be the same as the one specified by this field. If an incorrect IP address is configured on your server, you need to modify the IP address within the machine group, wait one minute, and then check again.
logtail_version	Version of the Logtail client
os	OS version
update_time	Time when Logtail is last started

File 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_i  d " : " E5F93BC6 - B024 - 11E8 - 8831 - 0A58AC1403
 9E_172 . 20 . 3 . 158_153605  3315 ",
   " ip " : " 172 . 20 . 3 . 158 ",
   " logtail_ve  rsion " : " 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_tim  e " : " 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_c onfig . json
[ 2018 - 09 - 13 01 : 13 : 59 . 025460 ] [ INFO ] [ 3155 ]
  [ build / release64 / sls / ilogtail / AppConfig . cpp : 176 ]
 load logtail config file , detail :{
   " config_ser ver_addres s " : " http :// logtail . ap - southeast
 - 2 - intranet . log . aliyuncs . com ",
   " data_serve r_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_pl ugin . LOG*
- **Container:** */ usr / local / ilogtail / logtail_pl ugin . LOG*
- **Windows:** plug-in logs are not supported.

File example

```
$ tail / usr / local / ilogtail / logtail_pl ugin . LOG
2018 - 09 - 13 02 : 55 : 30 [ INF ] [ docker_cen ter . go : 525 ]
[ func1 ] docker fetch all : start
2018 - 09 - 13 02 : 55 : 30 [ INF ] [ docker_cen ter . go : 529 ]
[ func1 ] docker fetch all : stop
2018 - 09 - 13 03 : 00 : 30 [ INF ] [ docker_cen ter . go : 525 ]
[ func1 ] docker fetch all : start
2018 - 09 - 13 03 : 00 : 30 [ INF ] [ docker_cen ter . go : 529 ]
[ func1 ] docker fetch all : stop
2018 - 09 - 13 03 : 03 : 26 [ INF ] [ log_file_r eader . go :
221 ] [ ReadOpen ] [## 1 . 0 ## sls - zc - test - hz - pub $ docker -
stdout - config , k8s - stdout ] open file for read , file
:/ logtail_ho st / var / lib / docker / containers / 7f46afec6a
14de39b59e e9cdfbfa8a 70c2fa26f1 148b2e2f31 bd3410f5b2 d624
/ 7f46afec6a 14de39b59e e9cdfbfa8a 70c2fa26f1 148b2e2f31
bd3410f5b2 d624 - json . log offset : 40379573 status :
794354 - 64769 - 40379963
2018 - 09 - 13 03 : 03 : 26 [ INF ] [ log_file_r eader . go :
221 ] [ ReadOpen ] [## 1 . 0 ## k8s - log - c12ba2028c fb444238cd
9ac1286939 f0b $ docker - stdout - config , k8s - stdout ]
open file for read , file :/ logtail_ho st / var / lib /
docker / containers / 7f46afec6a 14de39b59e e9cdfbfa8a 70c2fa26f1
148b2e2f31 bd3410f5b2 d624 / 7f46afec6a 14de39b59e e9cdfbfa8a
70c2fa26f1 148b2e2f31 bd3410f5b2 d624 - json . log offset :
40379573 status : 794354 - 64769 - 40379963
2018 - 09 - 13 03 : 04 : 26 [ INF ] [ log_file_r eader . go :
308 ] [ CloseFile ] [## 1 . 0 ## sls - zc - test - hz - pub $ docker
- stdout - config , k8s - stdout ] close file , reason : no
read timeout file :/ logtail_ho st / var / lib / docker
/ containers / 7f46afec6a 14de39b59e e9cdfbfa8a 70c2fa26f1
```

```
148b2e2f31  bd3410f5b2  d624 / 7f46afec6a  14de39b59e  e9cdfbfa8a
70c2fa26f1  148b2e2f31  bd3410f5b2  d624 - json . log      offset :
40379963      status : 794354 - 64769 - 40379963
2018 - 09 - 13   03 : 04 : 27  [ INF ] [ log_file_r  eader . go :
308 ] [ CloseFile ] [## 1 . 0 ## k8s - log - c12ba2028c  fb444238cd
9ac1286939  f0b $ docker - stdout - config , k8s - stdout ]     close
  file ,  reason : no   read   timeout      file :/ logtail_ho  st /
var / lib / docker / containers / 7f46afec6a  14de39b59e  e9cdfbfa8a
70c2fa26f1  148b2e2f31  bd3410f5b2  d624 / 7f46afec6a  14de39b59e
e9cdfbfa8a  70c2fa26f1  148b2e2f31  bd3410f5b2  d624 - json . log
  offset : 40379963      status : 794354 - 64769 - 40379963
2018 - 09 - 13   03 : 05 : 30  [ INF ] [ docker_cen  ter . go : 525 ]
[ func1 ]  docker   fetch   all : start
2018 - 09 - 13   03 : 05 : 30  [ INF ] [ docker_cen  ter . go : 529 ]
[ func1 ]  docker   fetch   all : stop
```

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 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_pat  h_config . json
```

File example

```
$ cat  / usr / local / ilogtail / docker_pat  h_config . json
{
   " detail " : [
      {
         " config_nam  e " : "## 1 . 0 ## k8s - log - c12ba2028c
 fb444238cd  9ac1286939  f0b $ nginx ",
         " container_  id " : " df19c06e85  4a0725ea7f  ca7e0378b0
 450f7bd312  2f94fe3e75  4d8483fd33  0d10 ",
         " params " : "{\ n    \" ID \" : \" df19c06e85  4a0725ea7f
 ca7e0378b0  450f7bd312  2f94fe3e75  4d8483fd33  0d10 \",\ n
 \" Path \" : \"/ logtail_ho  st / var / lib / docker / overlay2
 / 947db34669  5a1f65e63e  582ecfd10a  e1f57019a1  b99260b6c8
 3d00fcd189  2874 / diff / var / log \",\ n    \" Tags \" : [\ n
```

```
      \" nginx - type \",\ n       \" access - log \",\ n       \"
 _image_nam  e_ \",\ n       \" registry . cn - hangzhou . aliyuncs
 . com / log - service / docker - log - test : latest \",\ n       \"
 _container  _name_ \",\ n       \" nginx - log - demo \",\ n       \"
 _pod_name_ \",\ n       \" nginx - log - demo - h2lzc \",\ n       \"
 _namespace  _ \",\ n       \" default \",\ n       \" _pod_uid_ \",\
 n       \" 87e56ac3 - b65b - 11e8 - b172 - 00163f0086  85 \",\ n
   \" _container  _ip_ \",\ n       \" 172 . 20 . 4 . 224 \",\ n
 \" purpose \",\ n       \" test \"\ n    ]\ n }\ n "
     }
   ],
   " 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.

Server type	Reside in the same region as the Project	Configure an AliUid	Network type
ECS instances under your account	Yes	Not required	Alibaba Cloud intranet

Server type	Reside in the same region as the Project	Configure an AliUid	Network type
	No	Not required	Internet or Global Acceleration
ECS instances of other accounts	Yes	Required	Alibaba Cloud intranet
	No	Required	Internet or Global Acceleration
Cloud vendor servers or your own IDC servers	-	Required	Internet or Global Acceleration

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 #unique_65.

Examples of selecting a network type

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

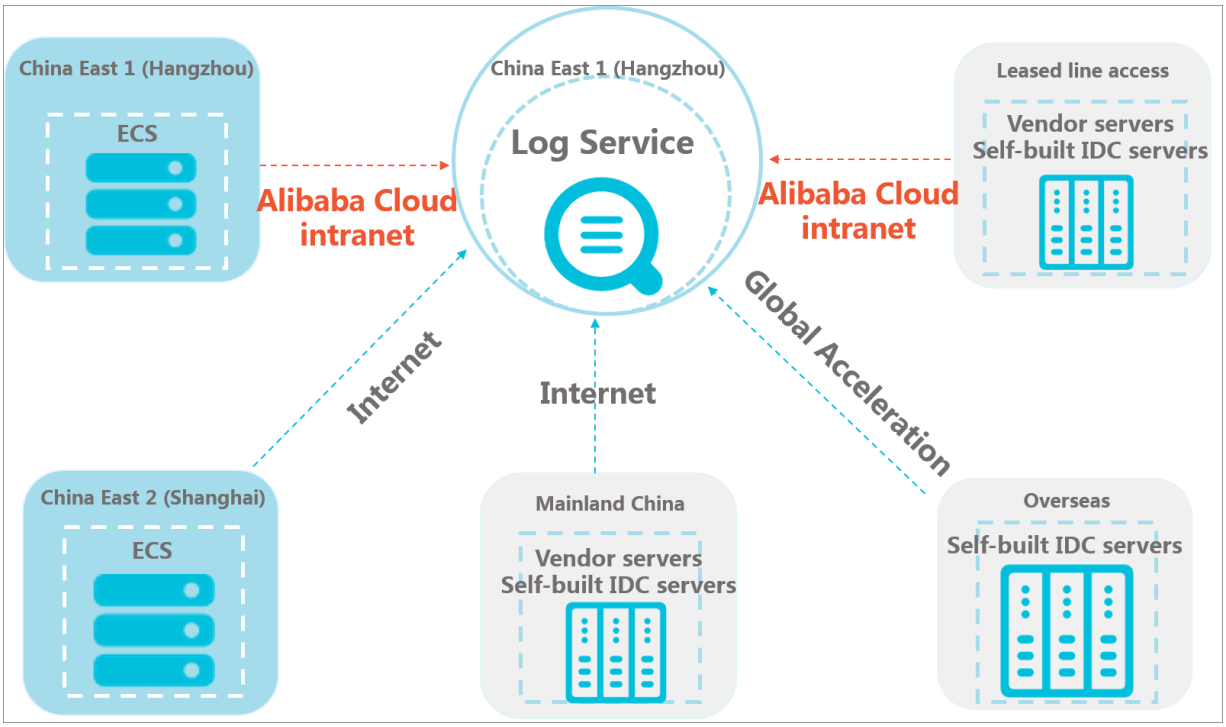
Note:

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.

Scenario	Region of the Log Service Project	Server type	Region of the ECS instance	Selected region for installing a Logtail client	Network type	Configure an AliUid
ECS and the Project are in the same region.	China East 1 (Hangzhou)	ECS of your current account	China East 1 (Hangzhou)	China East 1 (Hangzhou)	intranet	Not required
ECS and the Project are in different regions.	China East 2 (Shanghai)	ECS of your current account	China North 1 (Beijing)	China North 1 (Beijing)	Internet	Not required
Other accounts	China East 2 (Shanghai)	ECS belongs to other accounts.	China North 1 (Beijing)	China North 1 (Beijing)	Internet	Required
Server is in the local IDC.	China East 5 (Shenzhen)	Self-built IDC	-	China East 5 (Shenzhen)	Internet	Required

Scenario	Region of the Log Service Project	Server type	Region of the ECS instance	Selected region for installing a Logtail client	Network type	Configure an AliUid
Global Acceleration	Hong Kong	Self-built IDC	-	Hong Kong	Global Acceleration	Required

Figure 3-3: Examples of selecting a network type



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:

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

   - Windows:

     Open Management Tool in Control Panel, open Service, right-click *LogtailWor ker* , 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 / 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`

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:

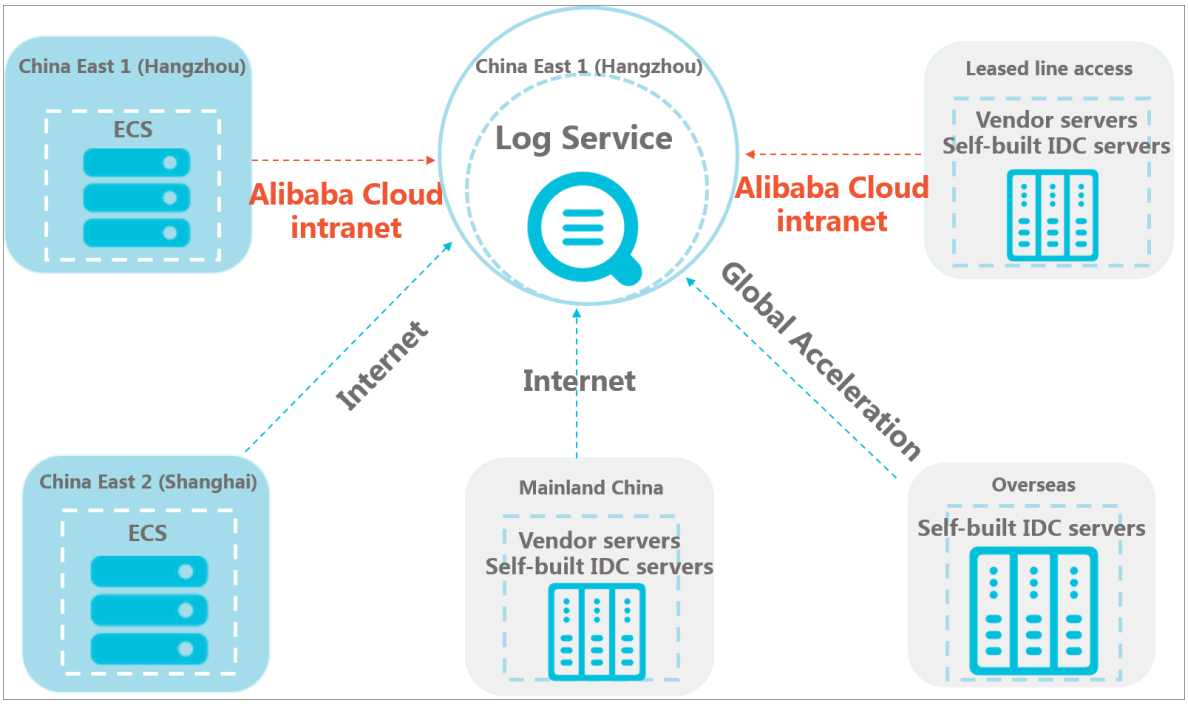
- Aliyun Linux
- Ubuntu
- Debian
- CentOS
- OpenSUSE
- Red Hat

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_regio n_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

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

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.

     ```
     wget http :// logtail - release - cn - hangzhou . oss - cn -
     hangzhou . aliyuncs . com / linux64 / logtail . sh - O logtail
     . sh ; chmod 755 logtail . sh
     ```

  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_regio n_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_regio n_name }` with the actual region name, and then run the installation command.

     ```
     wget http://logtail-
     release-${your_region_name}.oss-${your_region_name}-
     ```

```
internal.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755
 logtail.sh; ./logtail.sh install ${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.

Region	Installation command
China (Hangzhou)	`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`
China (Shanghai)	`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`
China (Qingdao)	`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`
China (Beijing)	`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`
China (Zhangjiakou)	`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`
China (Hohhot)	`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`
China (Shenzhen)	`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`
China (Chengdu)	`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`

 **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 #unique_65.

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

   In the installation command, `${ your_regio n_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_regio n_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/
   ```

```
linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh
 install ${your_region_name}-internet
```

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.

Region	Installation command
China (Hangzhou)	`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`
China (Shanghai)	`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`
China (Qingdao)	`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`
China (Beijing)	`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`
China (Zhangjiakou)	`wget http://logtail-release-cn-zhangjiako-u.oss-cn-zhangjiako-u.aliyuncs.com/linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh install cn-zhangjiako-u-internet`
China (Hohhot)	`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`
China (Shenzhen)	`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`
China (Chengdu)	`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`

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_regio n_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/
   ```

```
linux64/logtail.sh -O logtail.sh; chmod 755 logtail.sh; ./logtail.sh
 install ${your_region_name}-acceleration
```

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.

China (Beijing)	`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`
China (Qingdao)	`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`
China (Hangzhou)	`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`
China (Shanghai)	`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`
China (Shenzhen)	`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`
China (Zhangjiakou)	`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`
China (Hohhot)	`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`
China (Chengdu)	`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`

```
{
   " UUID " : " 0DF18E97 - 0F2D - 486F - B77F -*********",
   " hostname " : " david *******",
   " instance_i  d " : " F4FAFADA - F1D7 - 11E7 - 846C - 00163E3034
 9E_ ********* _151512954  8 ",
   " ip " : "**********",
   " logtail_ve  rsion " : " 0 . 16 . 0 ",
   " os " : " Linux ;  2 . 6 . 32 - 220 . 23 . 2 . ali1113 . el5 .
 x86_64 ; # 1   SMP   Thu   Jul   4   20 : 09 : 15   CST   2013 ;
 x86_64 ",
   " update_tim  e " : " 2018 - 01 - 05   13 : 19 : 08 "
}
```

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 .
 wget   http :// logtail - release - cn - hangzhou . oss - cn -
 hangzhou . aliyuncs . com / linux64 / logtail . sh  - O   logtail .
 sh ;  chmod   755   logtail . sh
#  Upgrade   Logtail .
 sudo  ./ logtail . sh   upgrade
```

Response:

```
#  The   upgrade   is   successful .
 Stop   logtail   successful  ly .
 ilogtail   is   running
 Upgrade   logtail   success
{
   " UUID " : "***",
   " hostname " : "***",
   " instance_i  d " : "***",
   " ip " : "***",
   " logtail_ve  rsion " : " 0 . 16 . 11 ",
   " os " : " Linux ;  3 . 10 . 0 - 693 . 2 . 2 . el7 . x86_64 ; # 1
 SMP   Tue   Sep   12   22 : 26 : 13   UTC   2017 ;  x86_64 ",
   " update_tim  e " : " 2018 - 08 - 29   15 : 01 : 36 "
}

#  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:

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

- **Stop Logtail**

  Run the following command as an administrator to stop Logtail:

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

Uninstall Logtail

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

```
wget   http :// logtail - release - cn - hangzhou . oss - cn -
hangzhou . aliyuncs . com / linux64 / logtail . sh  - O   logtail .
sh
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:

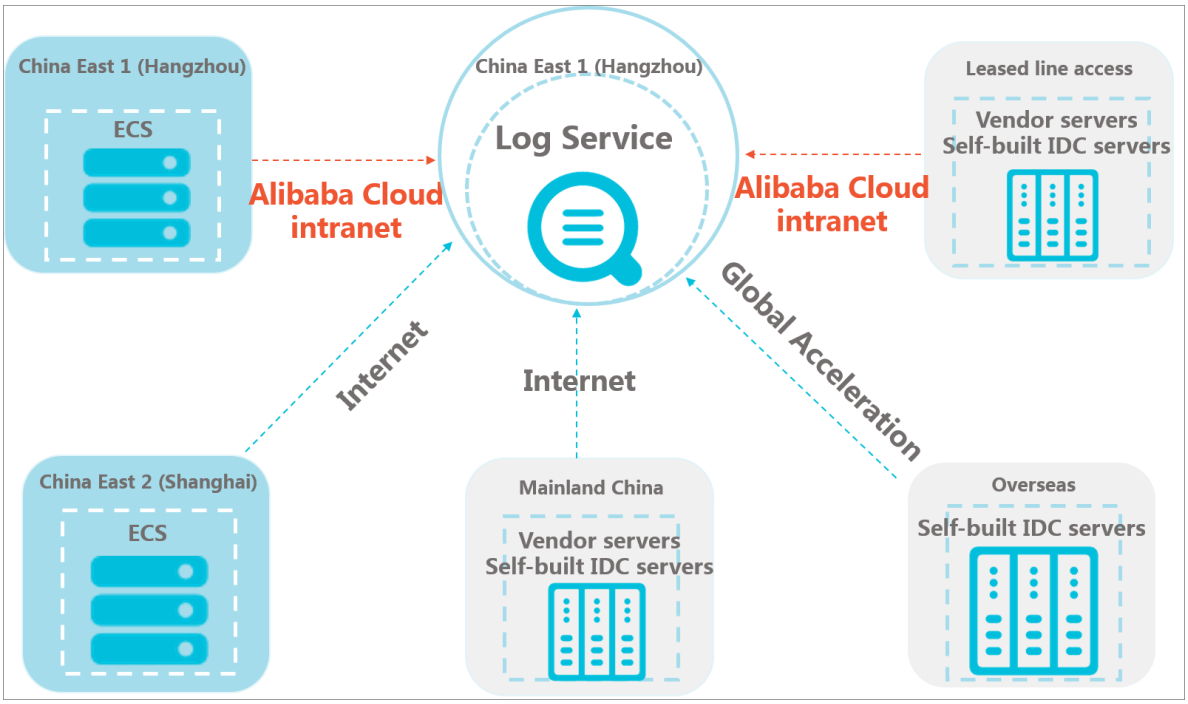
- **Windows 7 (Client) 32-bit**
- **Windows 7 (Client) 64-bit**
- **Windows Server 2008 32-bit**
- **Windows Server 2008 64-bit**
- **Windows Server 2012 64-bit**
- **Windows Server 2016 64-bit**

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_install.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_in staller` 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.

Region	Alibaba Cloud internal network (classic network or VPC)	Internet	Global Accelerati on
China (Qingdao)	`.\ logtail_in staller . exe install cn - qingdao`	`.\ logtail_in staller . exe install cn - qingdao - internet`	`.\ logtail_in staller . exe install cn - qingdao - accelerati on`
China (Beijing)	`.\ logtail_in staller . exe install cn - beijing`	`.\ logtail_in staller . exe install cn - beijing - internet`	`.\ logtail_in staller . exe install cn - beijing - accelerati on`
China (Zhangjiakou)	`.\ logtail_in staller . exe install cn - zhangjiako u`	`.\ logtail_in staller . exe install cn - zhangjiako u - internet`	`.\ logtail_in staller . exe install cn - zhangjiako u - accelerati on`
China (Hohhot)	`.\ logtail_in staller . exe install cn - huhehaote`	`.\ logtail_in staller . exe install cn - huhehaote - internet`	`.\ logtail_in staller . exe install cn - huhehaote - accelerati on`
China (Hangzhou)	`.\ logtail_in staller . exe install cn - hangzhou`	`.\ logtail_in staller . exe install cn - hangzhou - internet`	`.\ logtail_in staller . exe install cn - hangzhou - accelerati on`

Region	Alibaba Cloud internal network (classic network or VPC)	Internet	Global Accelerati on
China (Shanghai)	`.\ logtail_in staller . exe install cn - shanghai`	`.\ logtail_in staller . exe install cn - shanghai - internet`	`.\ logtail_in staller . exe install cn - shanghai - accelerati on`
China (Shenzhen)	`.\ logtail_in staller . exe install cn - shenzhen`	`.\ logtail_in staller . exe install cn - shenzhen - internet`	`.\ logtail_in staller . exe install cn - shenzhen - accelerati on`
China (Chengdu)	`.\ logtail_in staller . exe install cn - chengdu`	`.\ logtail_in staller . exe install cn - chengdu - internet`	`.\ logtail_in staller . exe install cn - chengdu - accelerati on`
Hong Kong	`.\ logtail_in staller . exe install cn - hongkong`	`.\ logtail_in staller . exe install cn - hongkong - internet`	`.\ logtail_in staller . exe install cn - hongkong - accelerati on`
US (Silicon Valley)	`.\ logtail_in staller . exe install us - west - 1`	`.\ logtail_in staller . exe install us - west - 1 - internet`	`.\ logtail_in staller . exe install us - west - 1 - accelerati on`
US (Virginia)	`.\ logtail_in staller . exe install us - east - 1`	`.\ logtail_in staller . exe install us - east - 1 - internet`	`.\ logtail_in staller . exe install us - east - 1 - accelerati on`

Region	Alibaba Cloud internal network (classic network or VPC)	Internet	Global Accelerati on
Singapore	`.\ logtail_in staller . exe install ap - southeast - 1`	`.\ logtail_in staller . exe install ap - southeast - 1 - internet`	`.\ logtail_in staller . exe install ap - southeast - 1 - accelerati on`
Australia (Sydney)	`.\ logtail_in staller . exe install ap - southeast - 2`	`.\ logtail_in staller . exe install ap - southeast - 2 - internet`	`.\ logtail_in staller . exe install ap - southeast - 2 - accelerati on`
Malaysia (Kuala Lumpur)	`.\ logtail_in staller . exe install ap - southeast - 3`	`.\ logtail_in staller . exe install ap - southeast - 3 - internet`	`.\ logtail_in staller . exe install ap - southeast - 3 - accelerati on`
Indonesia (Jakarta)	`.\ logtail_in staller . exe install ap - southeast - 5`	`.\ logtail_in staller . exe install ap - southeast - 5 - internet`	`.\ logtail_in staller . exe install ap - southeast - 5 - accelerati on`
India (Mumbai)	`.\ logtail_in staller . exe install ap - south - 1`	`.\ logtail_in staller . exe install ap - south - 1 - internet`	`.\ logtail_in staller . exe install ap - south - 1 - accelerati on`
Japan (Tokyo)	`.\ logtail_in staller . exe install ap - northeast - 1`	`.\ logtail_in staller . exe install ap - northeast - 1 - internet`	`.\ logtail_in staller . exe install ap - northeast - 1 - accelerati on`

Region	Alibaba Cloud internal network (classic network or VPC)	Internet	Global Accelerati on
Germany (Frankfurt)	`.\ logtail_in staller . exe install eu - central - 1`	`.\ logtail_in staller . exe install eu - central - 1 - internet`	`.\ logtail_in staller . exe install eu - central - 1 - accelerati on`
UAE (Dubai)	`.\ logtail_in staller . exe install me - east - 1`	`.\ logtail_in staller . exe install me - east - 1 - internet`	`.\ logtail_in staller . exe install me - east - 1 - accelerati on`
UK (London)	`.\ logtail_in staller . exe install eu - west - 1`	`.\ logtail_in staller . exe install eu - west - 1 - internet`	`.\ logtail_in staller . exe install eu - west - 1 - accelerati on`

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 #unique_65.

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_in staller` directory where you decompress the Logtail installation package. Then, run the following command to uninstall Logtail:

```
.\ logtail_in staller . exe uninstall
```

After Logtail is uninstalled, the installation directory of Logtail will be deleted. However, some residual configuration information is kept in the *C :\ LogtailDat a* 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_ch eck_point* : contains major checkpoint information of Logtail.
- *users* : contains configured AliUids.

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_c  onfig . json
```

- File format:

  JSON

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

```
{
    ...
    " cpu_usage_  limit " :  0 . 4 ,
    " mem_usage_  limit " :  100 ,
    " max_bytes_  per_sec " :  2097152 ,
    " process_th  read_count " :  1 ,
    " send_reque  st_concurr  ency " :  4 ,
    " buffer_fil  e_num " :  25 ,
    " buffer_fil  e_size " :  20971520 ,
    " buffer_fil  e_path " : "",
    ...
```

```
}
```

Common configuration parameters

Parameter	Description	Value
cpu_usage_limit	The CPU usage threshold, which is calculated by core. In most cases, the single-core processing capability is about 24 Mbit/s in simple mode and about 12 Mbit/s in full mode. For more information, see	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.
mem_usage_limit	The usage threshold of resident memory. To collect more than 1,000 distinct files, increase the threshold value properly.	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.

Parameter	Description	Value
max_bytes_per_sec	The traffic limit on the raw data sent by Logtail. Traffic exceeding 20 MB/s is not throttled.	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.
process_thread_count	The number of threads with which Logtail processes written data of log files. 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.	The value is of the Int type. Valid values: [1, 64]. Default value: 1.
send_request_concurrency	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. Note: A large value will lead to excessive network port usage. In this case, you need to modify relevant TCP parameters.	The value is of the Int type. Valid values: [1, 1000]. Default value: 20.

Parameter	Description	Value
buffer_file_num	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.	The value is of the Int type. Valid values: [1, 100]. Default value: 25.
buffer_file_size	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.	The value is of the Int type. Unit: Byte. Valid values: [1048576, 104857600]. Default value: 20971520, which is 20 MB.
buffer_file_path	The directory that stores cached files. After modifying this parameter, you need to move the files named in the format of *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.	By default, the value is an empty string. In this case, the cached files are stored in the Logtail installation directory */ usr / local / ilogtail .*
bind_interface	The name of the NIC bound to the local machine. For example, eth1. This parameter is valid only for Logtail for Linux.	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.

Parameter	Description	Value
check_point_filename	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.	Default value: /tmp/logtail_check_point.

Parameter	Description	Value
`user_config_file_path`	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 `user_config_file_path` to */data/logtail/user_log_config.json* 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.	By default, the *user_log_config.json* file is stored in the directory where the binary process is located.
`discard_old_data`	Specifies whether to discard historical logs. A value of true indicates that logs generated more than 12 hours ago will be discarded.	The value is of the Boolean type. Default value: true.
`working_ip`	The local IP address reported by Logtail. If the value is an empty string, Logtail automatically obtains the IP address of the local machine.	The value is an IP address. By default, the value is an empty string.
`working_hostname`	The local hostname reported by Logtail. If the value is an empty string, Logtail automatically obtains the hostname of the local machine.	The value is of the String type. By default, the value is an empty string.

Parameter	Description	Value
max_read_buffer_size	The maximum size of a log, in Bytes. If the size of a single log exceeds 512 KB, you can adjust the parameter value.	The value is of the Long type. Default value: 524288, which is 512 KB.
oas_connect_timeout	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.	The value is of the Long type. Unit: second. Default value: 5.
oas_request_timeout	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.	The value is of the Long type. Unit: second. Default value: 10.

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
```

```
/ etc / init . d / ilogtaild   status
```

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 #unique_65.
> - 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 #unique_40.

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_39.

3.4.2 Create a machine group with an IP address as its identifier

Log Service now supports machine groups with IP addresses as identifiers. After adding the server IP addresses obtained by Logtail to the target machine groups, you can use the same Logtail Config to collect logs from the servers.

Prerequisites

- You have created a project and a Logstore.
- You own at least one server. If the server is an Alibaba Cloud ECS server, make sure that the server is deployed in the region to which the project belongs.
- You have installed Logtail on the server. For details about how to install Logtail on a server, see #unique_31 and #unique_37.
- For servers of other cloud providers, on-premises IDCs, and ECS servers under other Alibaba Cloud accounts, confirm that you have configured AliUids for them. For details about how to configure AliUids, see #unique_65.

Context

You can collect logs through the Alibaba Cloud intranet even if you have not configured an intranet IP address. Server logs can be collected and sent to Log Service through the Alibaba Cloud intranet only when you use an Alibaba Cloud ECS server, the ECS server is deployed in the region to which the project belongs, and you select Alibaba Cloud intranet (Classic Network and VPC) when installing Logtail.

Procedure

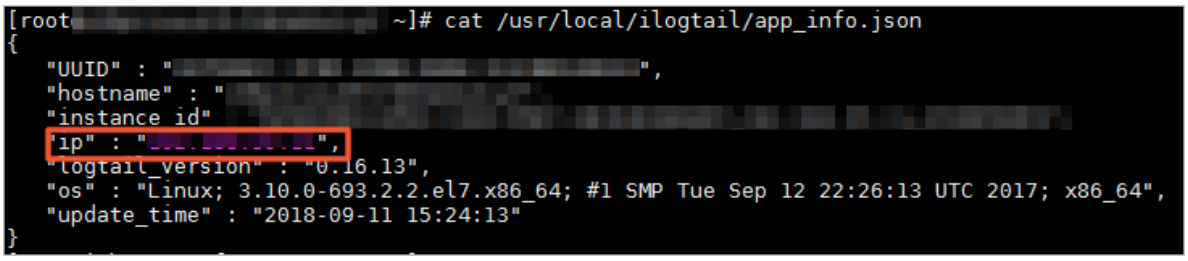
1. View the IP address of the server. The IP address is automatically generated by Logtail.

   The IP address is recorded in the ip field in the *app_info.json* file.

   You can view the file in the server with Logtail installed. The file path is:

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

   The following figure shows an example IP address of a Linux server.

2. Log on to the Log Service console, and then click the target project name.
3. Choose LogHub - Collect > Logtail Machine Group to view the Machine Groups list.
4. In the upper-right corner, click Create Machine Group.

   Alternatively, after creating a Logtail Config in the data access wizard, click Create Machine Group on the Apply to Machine Group page.

5. Create a machine group.

   a) Enter a Name.

   The name must be 3 to 128 characters in length and can contain lowercase letters, numbers, hyphens (-), and underscores (_). It must start and end with a lowercase letter or number.

   b) invalid content

   Note:

Exercise caution when you set a name for the machine group because the name cannot be changed once it is set.

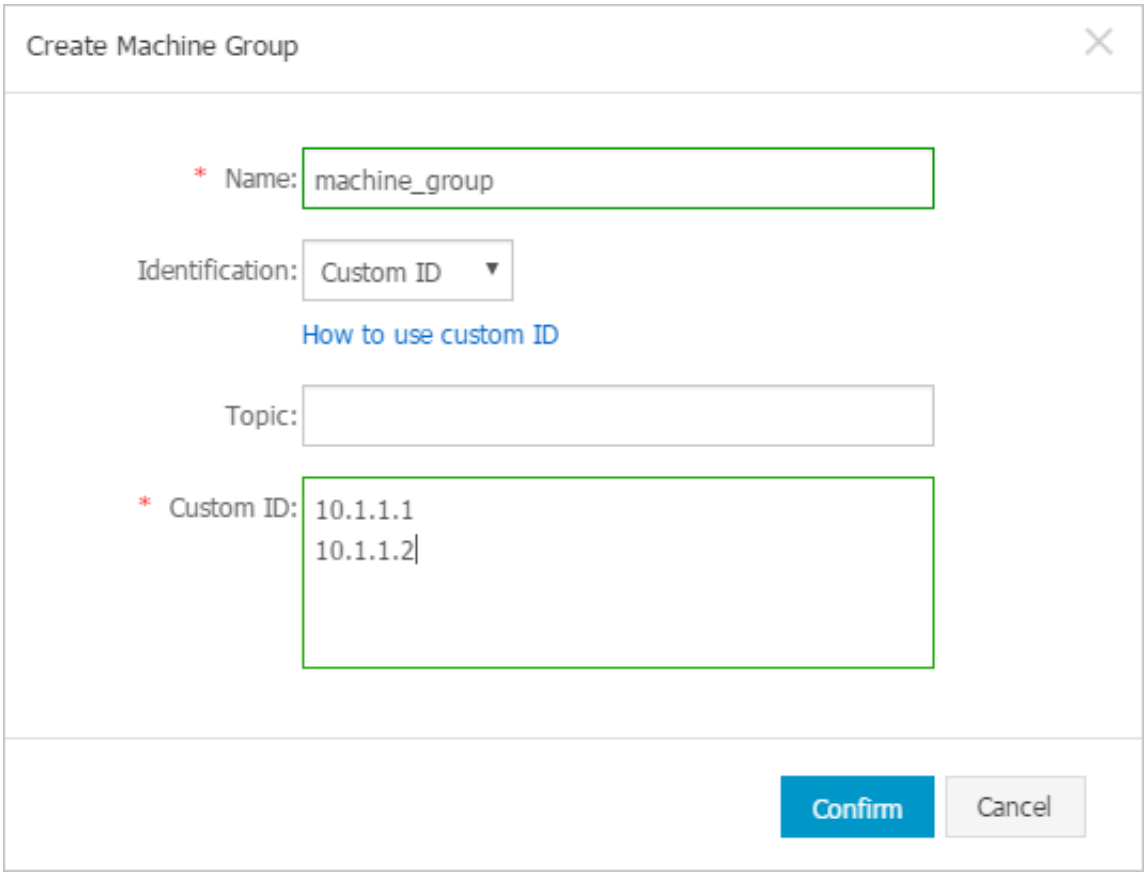
c) Set Identification to IP Address.

d) Enter an IP address.

The IP address is the server IP address you obtained from 1.

Note:

- You need to obtain the server IP address by following the instructions provided in 1.
- When multiple servers exist in the machine group, you need to use line breaks to separate the IP addresses of the servers.
- You cannot add both Windows servers and Linux servers to the machine group.



6. Optional: Enter a Topic.

For more information about machine group topics, see #unique_88.

7. Click Confirm.

Result

You can view the newly created machine group in the Machine Groups list.

3.4.3 Create an ID to identify a machine group

In addition to IP addresses, you can also use custom IDs as identifiers of machine groups.

The advantages of using custom IDs as identifiers for machine groups are described in the following scenarios:

- In a custom network, such as a VPC, IP addresses conflicts may occur among server. As a result, Log Service cannot manage Logtail. In this case, custom IDs can be added to the servers to prevent IP address conflicts.
- A custom ID can be configured for multiple servers to achieve elastic scaling of a machine group. Specifically, you can add the same custom ID to new servers so that Log Service can automatically identify the servers and add them to the machine group.

Procedure

1. Set custom IDs on the servers.
   - For Logtail in Linux:

     Set custom IDs by using the `/ etc / ilogtail / user_defin ed_id` file.

     The following is an example:

     ```
     # vim / etc / ilogtail / user_defin ed_id
     ```

     Enter `userdefine d` in this file.
   - For Logtail in Windows:

     Set custom IDs by using the `C :\ LogtailDat a \ user_defin ed_id` file.

     The following is an example:

     ```
     C :\ LogtailDat a > more user_defin ed_id
     userdefine d_windows
     ```

   Note:

- You cannot add Linux and Windows servers into the same machine group. Therefore, you must set different custom IDs for Linux and Windows servers.
- When multiple custom IDs are configured for a server, you need to use line breaks to separate them.
- If the `/etc/ilogtail/` or `C:\LogtailData\a` directory or the `/etc/ilogtail/user_defined_id` or `C:\LogtailData\a\user_defined_id` file does not exist, manually create one.

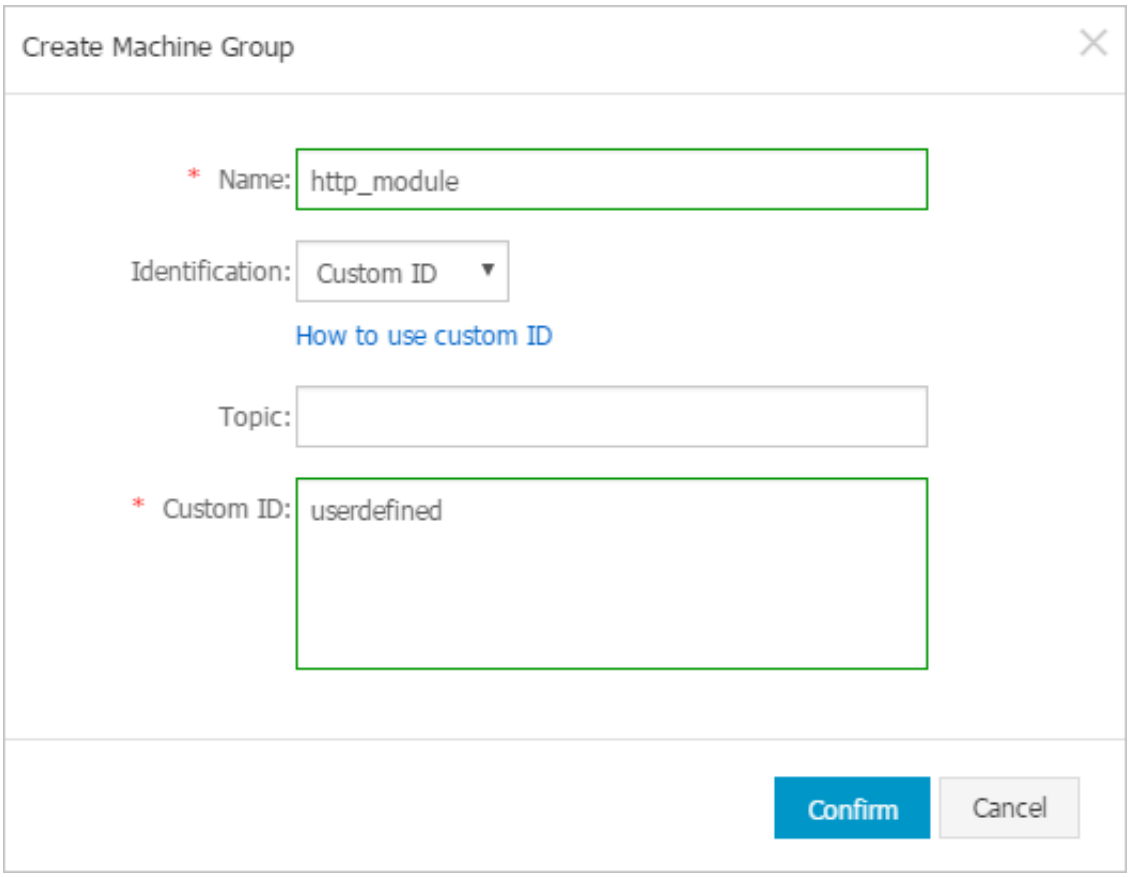
2. Create a machine group.
   a. Log on to the Log Service console, and then click the target project name.
   b. In the left-side navigation pane, click Logtail Machine Group.
   c. On the Machine Groups page, click Create Machine Group the upper-right corner.
   d. Set the machine group configurations.
      - Name: Enter a name.

        The name must be 3 to 128 characters in length and can contain lowercase letters, numbers, hyphens (-), and underscores (_). It must start and end with a lowercase letter or number.

         Note:

Exercise caution when you set a name for the machine group because the name cannot be changed once it is set.

- Identification: Select Custom ID.
- (Optional) Topic: Enter a topic. For more information about machine group topics, see #unique_88.
- Custom ID: Enter the custom ID obtained from step 1.

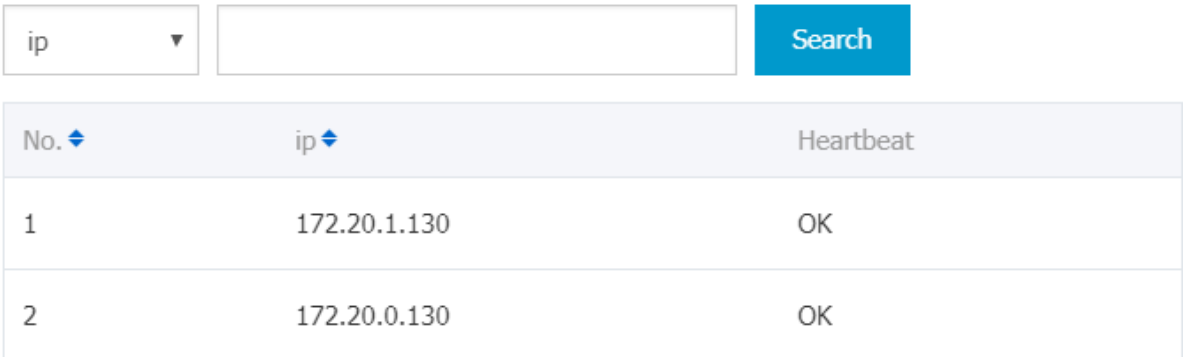


e. Click Confirm.

Note:
To scale out a server group, you only need to add the custom ID to the new server.

3. Check the status of the machine group.

   On the Machine Groups page, locate the target machine group, and click Status in the Actions column. Then, you can view the list containing all servers using the same custom ID and their heartbeat information.



Disable custom IDs

If you want use server IP addresses as machine group identifiers, delete the user_defined_id file. The setting takes effect within one minute.

- In Linux:

```
rm  - f  / etc / ilogtail / user_defin  ed_id
```

- In Windows:

```
del   C :\ LogtailDat  a \ user_defin  ed_id
```

Effective time

By default, addition, deletion, and modification of the user_defined_id file take effect within one minute.

If you want the setting to take effect immediately, run the following command and restart Logtail:

- In Linux:

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

- In Windows:

  Choose Control Panel > Administrative Tools > Services. Then, right-click the LogtailWorker service and choose Restart from the shortcut menu.

Example

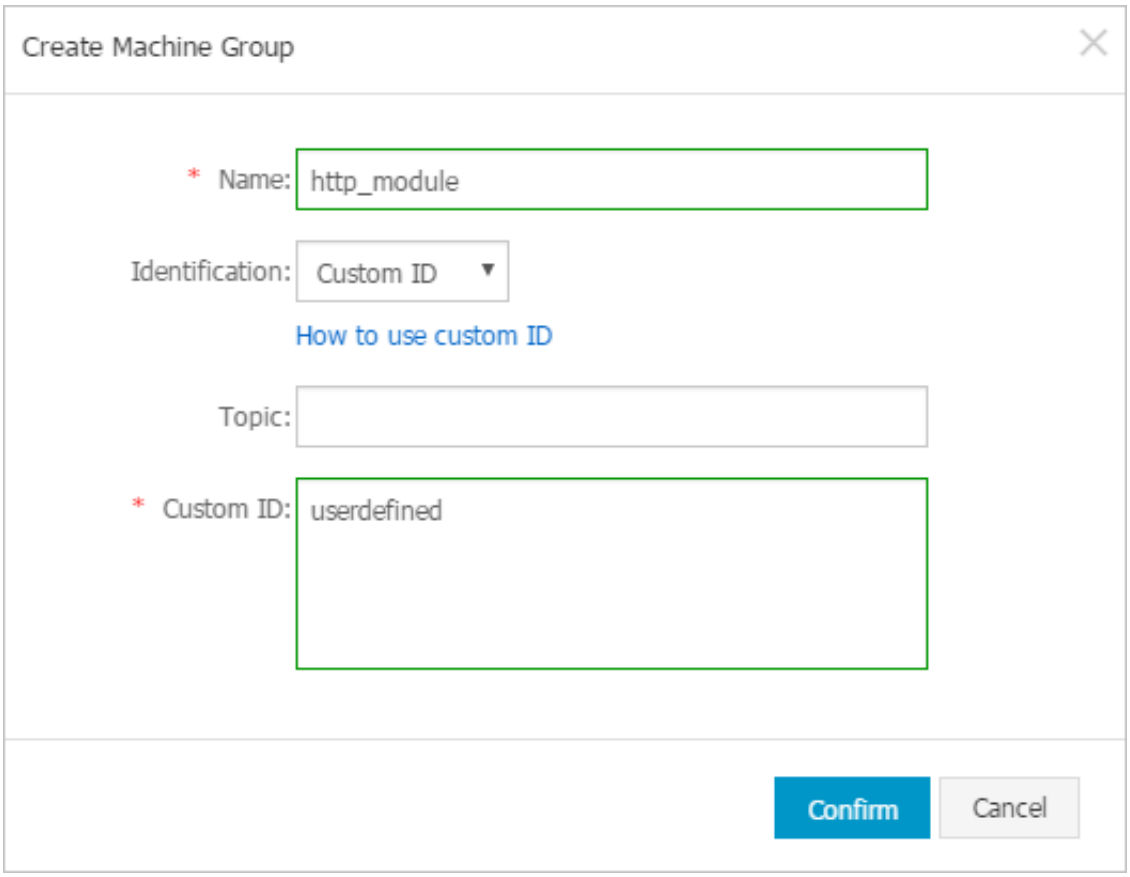
The system is composed of multiple modules, each of which can contain multiple servers. For example, a common website can be divided into a frontend HTTP request processing module, a cache module, a logic processing module, and a storage module. Each module can be individually expanded. Therefore, you need to enable Log Service to collect logs from new servers in real time.

1. Create a custom ID.

   After installing the Logtail client, use custom IDs as machine group identifiers. In this example, there are four identifiers (indicating the four modules): http_module , cache_module, logic_module, and store_module

2. Create a machine group.

   Set Identification to the actual custom ID of the machine group. The following figure uses the http_module machine group as an example.



3. In the Machine Group Status dialog box, view the list containing all servers using the same custom ID and their heartbeat information.
4. Add the custom ID to the 10.1.1.3 server if the server is added to the machine group. Then, you can view the new server in the Machine Group Status dialog box.

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.

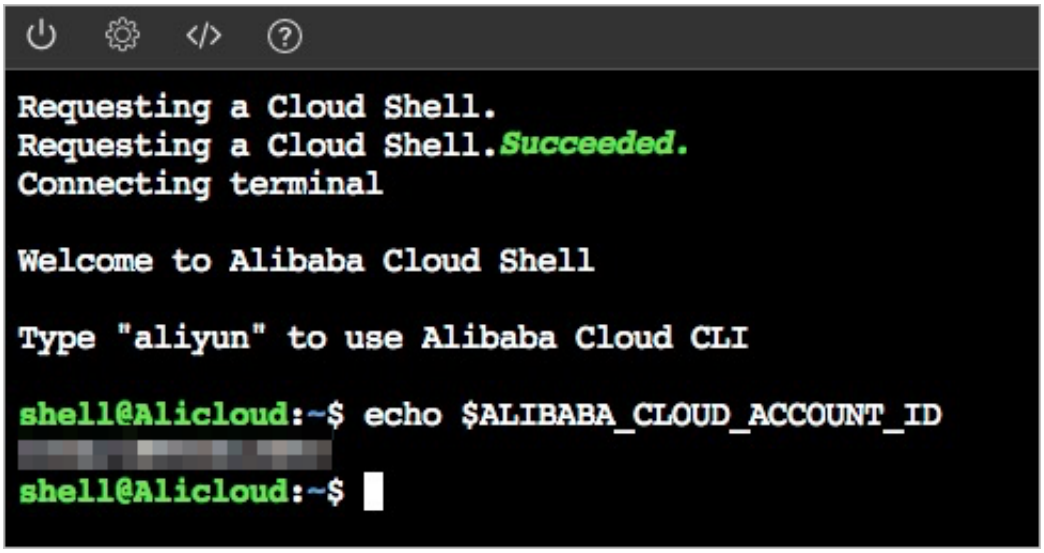
Prerequisites

- 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 #unique_31 and #unique_37 as needed.

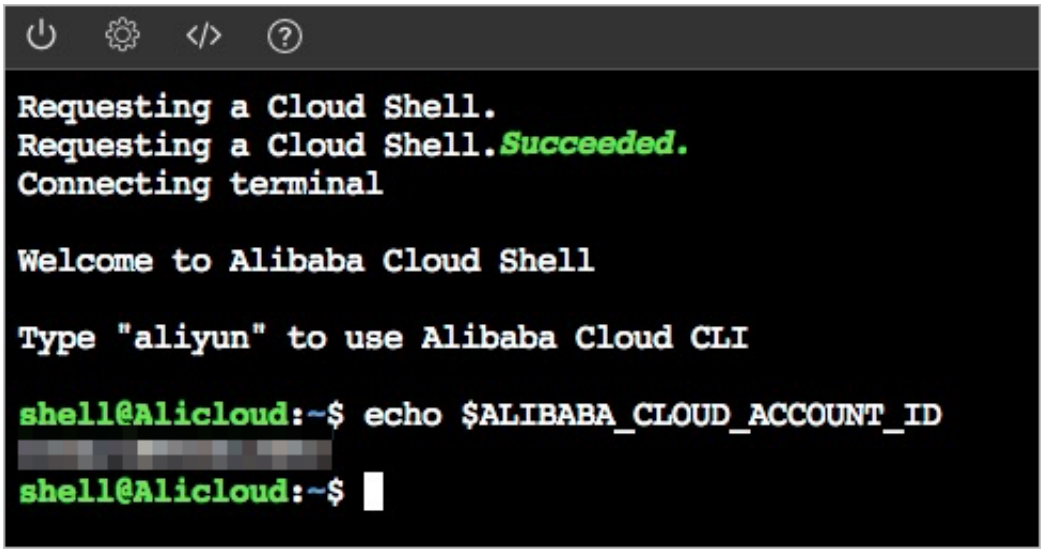
Procedure

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.



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



2. Configure an AliUid for the server.

   - In Linux:

     Create a file named after the AliUid in the `/ etc / ilogtail / 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 / ilogtail / users / 1 ***************
     touch  / etc / ilogtail / users / 1 ***************
     ```

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

     ```
     rm  / etc / ilogtail / users / 1 ***************
     ```

   - In Windows:

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

     If you want to delete the AliUid, you can simply delete this file. (`C :\ LogtailDat  a \ 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 Create a Logtail configuration

The Logtail client provides an easy way to collect logs from Elastic Compute Service (ECS) instances in the Log Service console. After installing the Logtail client, you must create a log collection configuration for the Logtail client. For how to install Logtail, see #unique_31 and #unique_37. You can create and modify the Logtail configurations of LogStores in the LogStore list.

Create a Logtail configuration

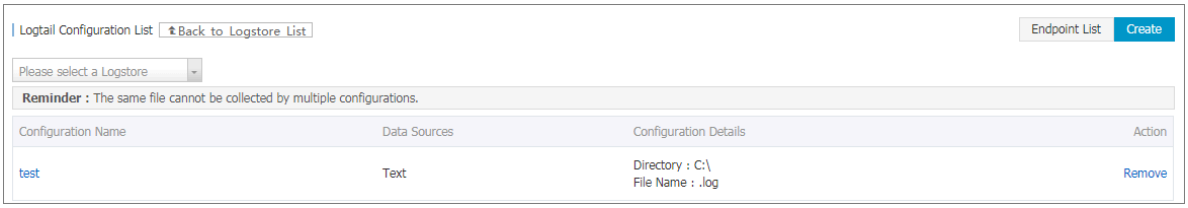
For how to create a Logtail configuration in the Log Service console, see Collect text logs and #unique_42.

View Logtail configuration list

1. Log on to the Log Service console.
2. On the Project List page, click the project name, to enter the Logstore List page.
3. On the Logstore List page, click Manage at the right of the Logstore. Logtail Configuration List page appears.

   All the configurations of this Logstore are displayed on the page, including the configuration name, data sources, and configuration details. When the data source is Text, the file path and file name are displayed under Configuration Details.

   Figure 3-7: Logtail configuration list



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

Modify a Logtail configuration

1. Log on to the Log Service console.
2. On the Project List page, click the project name.
3. On the Logstore List page, click Manage at the right of the Logstore. The Logtail Configuration List page appears.
4. Click the name of the Logtail configuration to be modified.

   You can modify the log collection mode and specify the machine group to which the modified mode is applied. The configuration modification process is the same as the configuration creation process.

Delete a Logtail configuration

1. Log on to the Log Service console.
2. On the Project List page, click the project name.
3. On the Logstore List page, click Manage at the right of the Logstore. The Logtail Configuration List page appears.

4. Click Remove at the right of the Logtail configuration to be deleted.

   After the configuration is deleted successfully, it is unbound from the machine groups that applied this configuration and Logtail stops collecting the log files of the deleted configuration.

    Note:
   You must delete all the Logtail configurations in a Logstore before deleting the Logstore.

3.4.6 Manage a machine group

Log Service manages all the Elastic Compute Service (ECS) instances whose logs need to be collected by using the Logtail client in the form of machine groups. You can go to the Machine Groups page by clicking a project name on the Project List page and then clicking LogHub - Collect > Logtail Machine Group in the left-side navigation pane on the Logstore List page. You can create, modify, and delete a machine group, view the machine group list and status, manage the configurations, and use the machine group identification in the Log Service console.

Create a machine group

You can define a machine group by using:

- IP: Define the machine group name and add the intranet IP addresses of a group of machines. For more information, see #unique_40.
- User-defined identity: Define an identity for the machine group and configure the identity on the corresponding machine for association. For more information, see #unique_39.

For how to create a machine group, see #unique_40.

View machine group list

1. Log on to the Log Service console.

2. On the Logstore List page, click Logtail Machine Group in the left-side navigation pane. The Machine Groups page appears.

   View all of the machine groups in the project.

   Figure 3-8: View a list of machine groups

Modify a machine group

After creating a machine group, you can adjust the ECS instances in the machine group as per your needs.

> **Note:**
> The machine group name cannot be modified after the machine group is created.

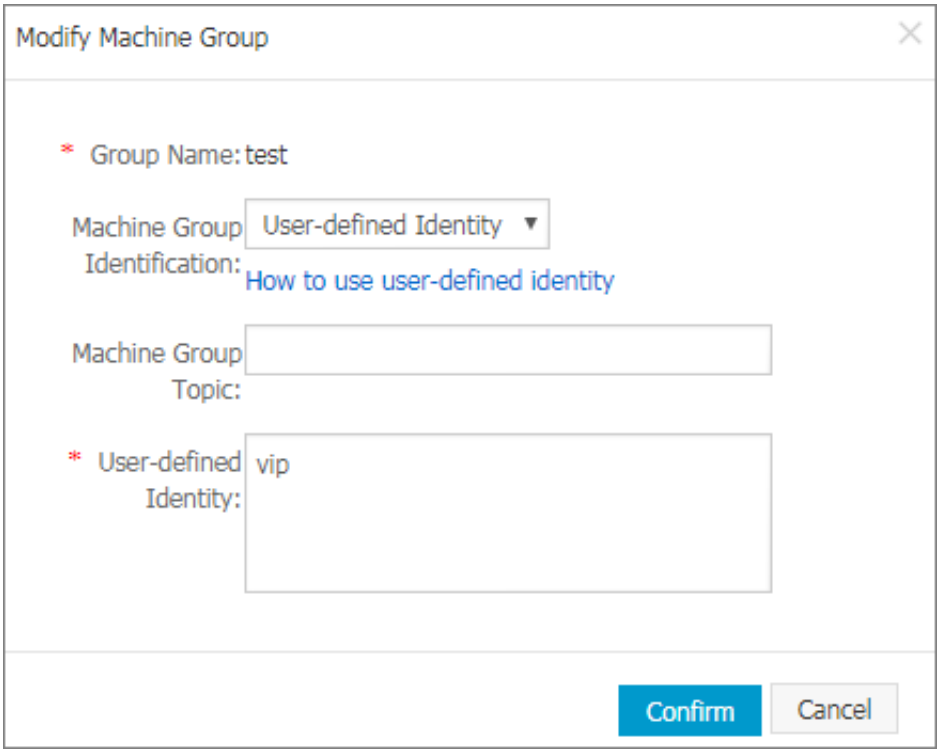
1. Log on to the Log Service console.
2. On the Logstore List page, click Logtail Machine Group in the left-side navigation pane. The Machine Groups page appears.

   All machine groups under the project are displayed.
3. Click Modify at the right of the machine group.
4. Modify the configurations and then click Confirm.

> **Note:**

The machine group name cannot be modified.

Figure 3-9: Modify a Machine Group



View status

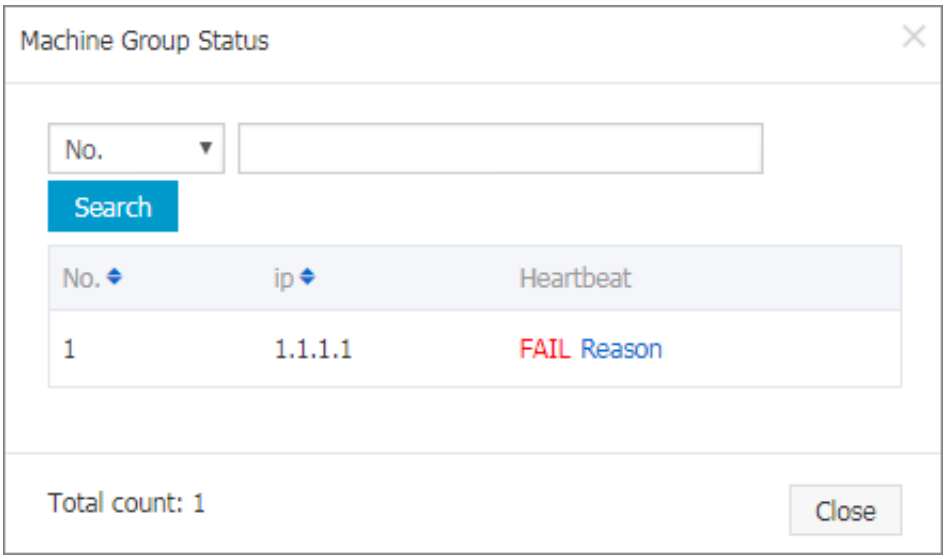
To verify that the Logtail client is successfully installed on all ECS instances in a machine group, view the heartbeat status of the Logtail client.

Note:
After a machine is added to a machine group, the change to the hearbeat status of the machine group takes about two minutes to take effect. Then, you can view the hearbeat status.

1. Log on to the Log Service console.
2. On the Project List page, click the project name. On the Logstore List page, click LogHub - Collect > Logtail Machine Group in the left-side navigation pane. The Machine Groups page appears.

3. Find the target machine group, and then click Machine Status.
   - If the Logtail client is installed on each ECS instance, and Logtail can normally communicate with Log Service, the heartbeat status is displayed as `OK`.
   - If Logtail cannot communicate with Log Service, the heartbeat status is displayed as `FAIL`. Furthermore, if the heartbeat status retains as `FAIL`, troubleshoot the exception by following the prompted messages on the page and the instructions provided by #unique_93. If the problem persists, open a tick for support.

Figure 3-10: View the machine group status



Managing configurations

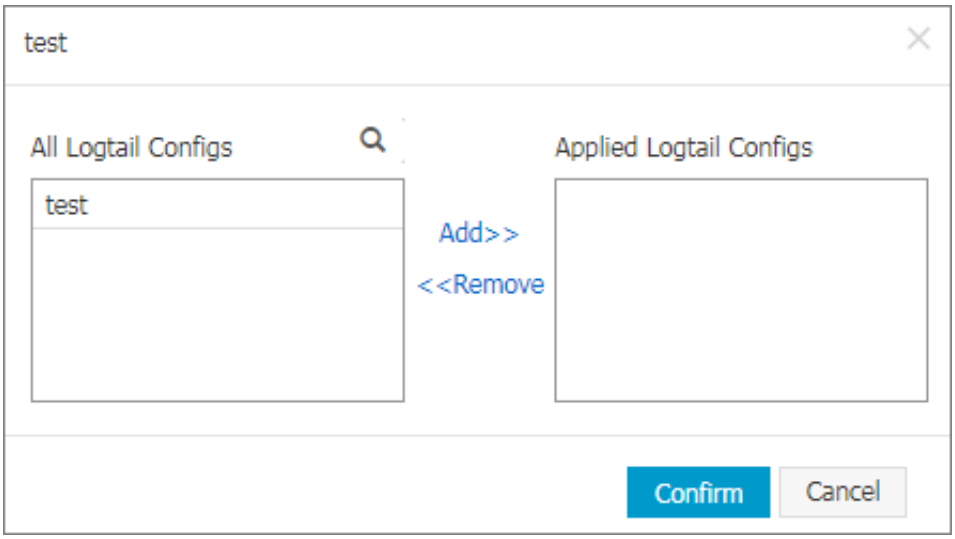
Log Service manages all the servers whose logs need to be collected by using machine groups. One important management item is the collection configuration of the Logtail client. For more information, see Collect text logs and #unique_42. You can apply or delete a Logtail configuration to/from a machine group to decide what logs are collected, how the logs are parsed, and to which Logstore the logs are sent by the Logtail on each ECS instance.

1. Log on to the Log Service console.
2. On the Logstore List page, click Logtail Machine Group in the left-side navigation pane. The Machine Groups page appears.
3. Click Config at the right of the machine group.

4. Select the Logtail configuration and click Add or Remove to add or remove the configuration to/from the machine group.

   After a Logtail configuration is added, it is issued to the Logtail client on each ECS instance in the machine group. After a Logtail configuration is removed, it is removed from the Logtail client.

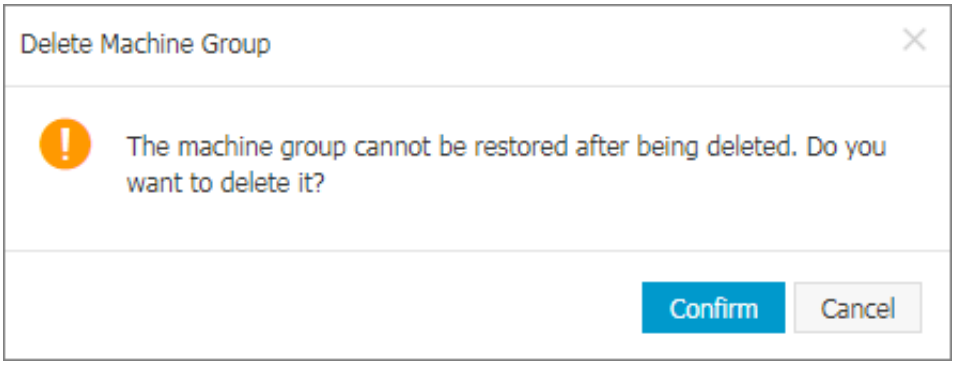
Figure 3-11: Managing Machine Group configurations



Delete a machine group

1. Log on to the Log Service console.
2. On the Project List page, click the project name. On the Logstore List page, click LogHub - Collect > Logtail Machine Group in the left-side navigation pane. The Machine Groups page appears.
3. Click Delete at the right of the machine group.
4. Click Confirm in the appeared dialog box.

Figure 3-12: Delete a machine group



3.5 Text logs

3.5.1 Collect text logs

The Logtail client helps Log Service users collect text logs from Elastic Compute Service (ECS) instances or local servers in the console.

Prerequisites

- You must install Logtail before collecting logs. For installation methods, see #unique_31 and #unique_37.
- To collect logs from ECS instances or local servers, you must open ports 80 and 443.

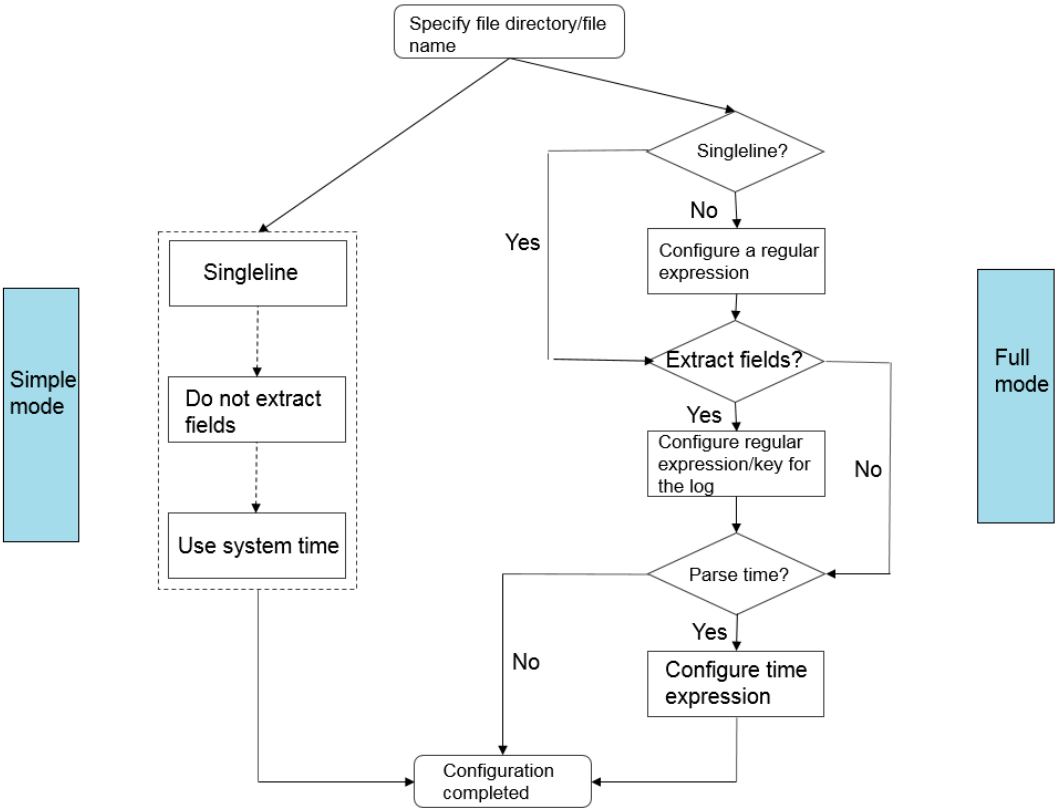
Limits

- A file can only be collected using one configuration. To collect a file with multiple configurations, we recommend you use the soft link. For example, to collect files under `/ home / log / nginx / log` with two configurations, you can use the original path for one configuration, and run the command `ln - s / home / log / nginx / log / home / log / nginx / link_log` to create a soft link of this folder, and then use the soft link path for the other configuration.
- For more information about operating systems supported by the Logtail client, see #unique_4.
- The ECS instances of the classic network or Virtual Private Cloud (VPC) and the Log Service project must belong to the same region. If your source data is transmitted by Internet (similar to IDC), you can select the region that the Log Service resides in based on the region description.

Configuration process of log collection

The following are simple mode and full mode examples. The configuration process as follows:

Figure 3-13: Log collection configuration process



Log collection modes

Logtail supports simple mode, delimiter mode, JSON mode, full mode, and other log collection methods.

- Simple mode

  Currently, simple mode is the single-line mode. By default, one line of data is a log, and two logs are separated by a line break in the log file. The system does not extract log fields (that is, the regular expression (.*) by default), and uses the current server system time as the generated log time. To configure more settings, you can change the configuration to full mode to adjust the settings. For more information on how to change the Logtail configuration, see #unique_96.

  In the simple mode, specify the file directory and file name. Then, the Logtail collects logs by each line and uses the system time.

- Delimiter mode

  Logtail can collect delimiter logs through the delimiter mode. For more information, see #unique_97.

- JSON mode

  You can select JSON mode to collect JSON logs.

- Full mode

  To configure more personalized field extraction settings for log contents (such as cross-line logs and field extraction), select Full Mode.

  Log Service provides a log sample-based regular expression generation function in the data collection wizard. However, multiple manual tests to fit the log samples are required because of the different log samples. For more information about how to test the regular expressions, see How do I test regular expressions?

Procedure

1. Click Project name to enter the Logstore List.
2. Select Logstore, and click the Wizard at the right side of the Logstore.
3. Select the data source.

   Select Text under Other Sources and then click Next to go to the Configure Data Source step.
4. Specify the Configuration Name.

   The configuration name can be 3–63 characters long, contain lowercase letters , numbers, hyphens (-), and underscores (_), and must begin and end with a lowercase letter or number.

    Note:
   The configuration name cannot be modified after the configuration is created.
5. Specify the log directory and the file name.

   The directory structure must be a full path or a path that contains wildcards.

    Note:

Only `*` and `?`can be used as wildcards in the directory.

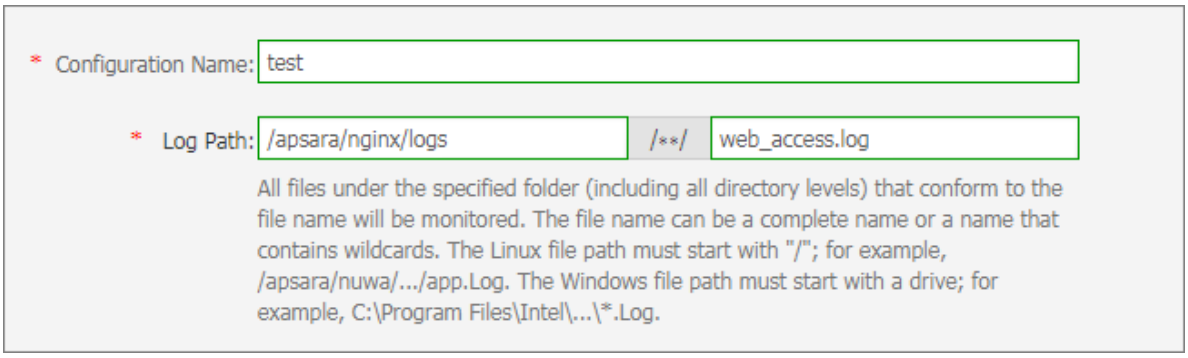
The log file name must be a complete file name or a name that contains wildcards. For the rules of file names, see Wildcard matching.

The search mode of log files is the multi-level directory matching mode, namely , under the specified folder (including directories of all levels), all the files that conform to the file name can be monitored.

- `/ apsara / nuwa / … /*. log` means the files whose suffix is `. log` and exist in the `/ apsara / nuwa` directory (including its recursive subdirectories).
- `/ var / logs / app_ * … /*. log *` means the files whose file name contains `. log` and exist in all of the directories that conform to the `app_ *` mode (including their recursive subdirectories) under the `/ var / logs` directory.

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

Figure 3-14: Specify the Directory and file name



6. Set collection mode. The following uses the full mode as an example.
   a. Enter the Log Sample.

      The purpose of providing a log sample is facilitating the Log Service console in automatically extracting the regex matching mode in logs. Be sure to use a log from the actual environment.

   b. Disable Singleline.

      By default, the single-line mode is used, that is, two logs are separated by a line break. To collect cross-line logs (such as Java program logs), you must disable Singleline and then configure the Regular Expression.

   c. Modify. the Regular Expression.

      You can select to automatically generate the regular expression or manually enter the regular expression. After entering the log sample, click Auto Generate to automatically generate the regular expression. If failed, you can switch to the manual mode to enter the regular expression for verification.

   d. Enable Extract Field.

      To analyze and process fields separately in the log content, use the Extract Field function to convert the specified field to a key-value pair before sending it to Log Service. Therefore, you must specify a method for parsing the log content, that is, a regular expression.

      The Log Service console allows you to specify a regular expression for parsing the log content in two ways. The first option is to automatically generate a regular expression through simple interactions. You can select the field to be extracted in the log sample and then click Generate RegEx to automatically generate the regular expression in the Log Service console.

      In this way, you can generate the regular expression without writing it on your own. You can also manually enter a regular expression. Click Manually Input Regular Expression to switch to the manual input mode. After entering the regular expression, click Validate to validate whether or not the entered regular

expression can parse and extract the log sample. For more information, see How do I test regular expressions?

No matter the regular expression for parsing the log content is automatically generated or manually entered, you must name each extracted field, that is, set keys for the fields.

Figure 3-15:

e. Set Use System Time.

Default settings Use System Time is set by default. If disabled, you must specify a certain field (value) as the time field during field extraction and name this field `time`. After selecting a `time` field, you can click Auto Generate

in Time Format to generate a method to parse this time field. For more information on log time format, see #unique_100.

f. Enable Drop Failed to Parse Logs as needed.

This option specifies whether to upload the logs with parsing failure to Log Service.

When this option is enabled, the logs with parsing failure will not be uploaded to Log Service. When the option is disabled, the raw log will be uploaded to Log Service when log parsing fails. The key of the raw log is `__raw_log_ _` and the value is the log content.

7. (Optional) Set Advanced Options as needed and click Next.

If you have no special requirements, retain the default settings.

Configuration item	Desceiption
Upload Original Log	Select whether or not to upload the original log. If enabled, the new field is added by default to upload the original log.
Topic Generation Mode	· Null - Do not generate topic: The default option, which indicates to set the topic as a null string and you can query logs without entering the topic. · Machine Group Topic Attributes: Used to clearly differentiate log data generated in different frontend servers. · File Path Regular: With this option selected, you must enter the Custom RegEx to use the regular expression to extract contents from the path as the topic. Used to differentiate log data generated by users and instances. Used to differentiate log data generated by users and instances.
Custom RegEx	After selecting File Path Regular as Topic Generation Mode, you must enter your custom regular expression.
Log File Encoding	· utf8: Use UTF-8 encoding. · gbk: Use GBK encoding.
Maximum Monitor Directory Depth	Specify the maximum depth of the monitored directory when logs are collected from the log source, that is, at most how many levels of logs can be monitored. The range is 0–1000, and 0 indicates to only monitor the current directory level.

Configuration item	Desceiption
Timeout	A log file has timed out if it does not have any update within a specified time. You can configure the following settings for Timeout. · Never Time out: Specify to monitor all log files persistent ly and the log files never time out. · 30 minute timeout: A log file has timed out and is not monitored if it does not have any update within 30 minutes.
Filter Configurat ion	Only logs that completely conform to the filter conditions can be collected. For example: · collect logs that conform to a condition: `Key : level Regex : WARNING | ERROR` indicates to only collect logs that are in the WARNING or ERROR level. · filter logs that do not conform to a condition : - `Key : level Regex :^(?!. *( INFO | DEBUG ))`, indicates to not collect logs whose level is INFO or DEBUG. - `Key : url Regex :. *^(?!.*( healthchec k )). *`, indicates to filter logs with healthcheck in the url. Such as logs in which key is url and value is `/ inner / healthchec k / jiankong . html` will not be collected. For similar examples, seeregex-exclude-word and regex-exclude-pattern.

8. Click Next after completing the configurations.

   If you have not created a machine group, you must create one first. For how to create a machine group, see Create a machine #unique_40group.

   Note:

   · It takes up to three minutes for the Logtail configuration to take effect, so be patient.

- To collect IIS access logs, see #unique_101.
- After creating the Logtail configuration, you can view the Logtail configuration list, modify the Logtail configuration, or delete the Logtail configuration. For more information, see #unique_96.

Figure 3-16: Applying the configuration to the machine group

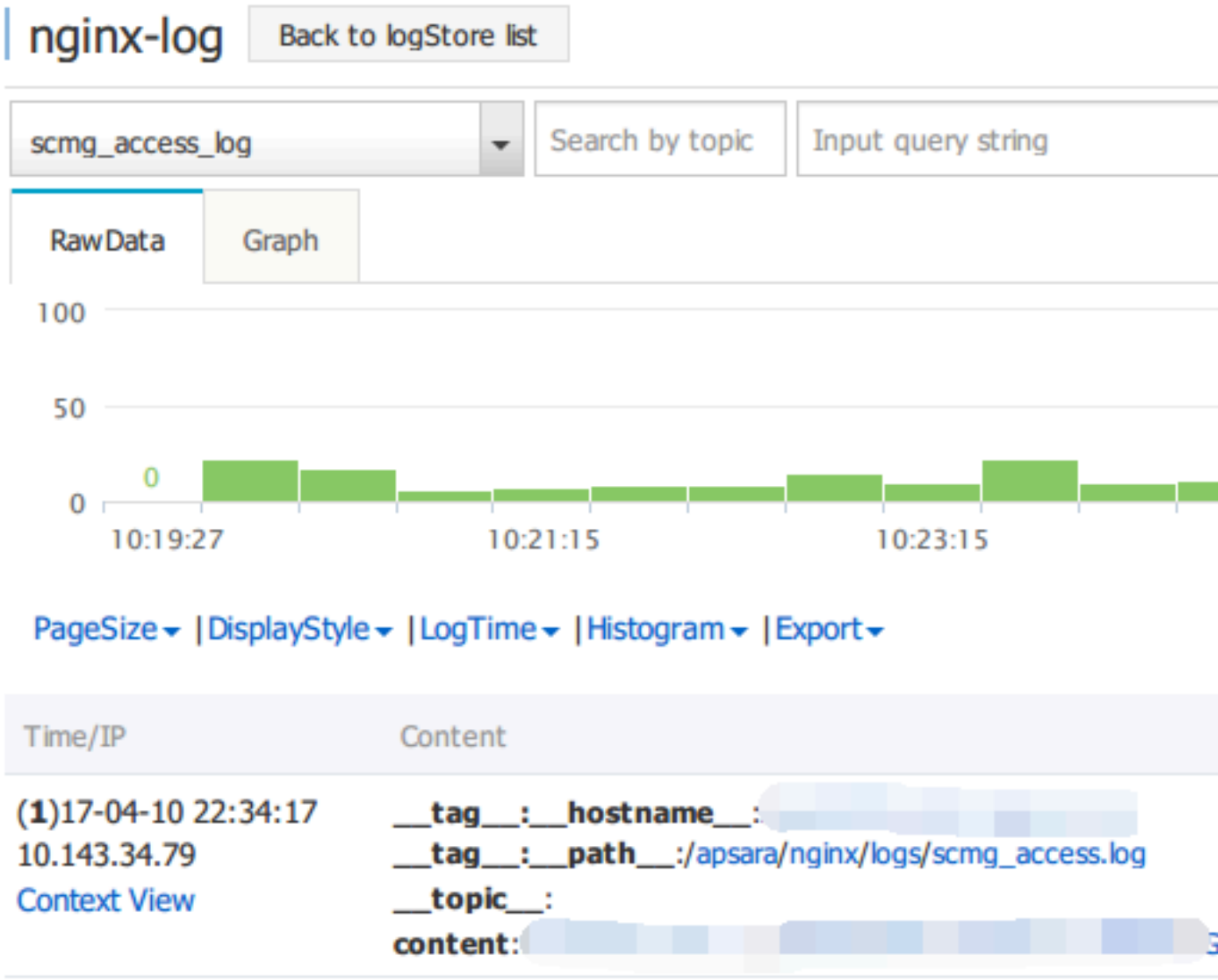
Log Service starts to collect logs after completing the configurations.

Subsequent operations

After completing the preceding configurations, you can configure the Search, Analysis, and Visualization and Shipper & ETL as instructed on the page.

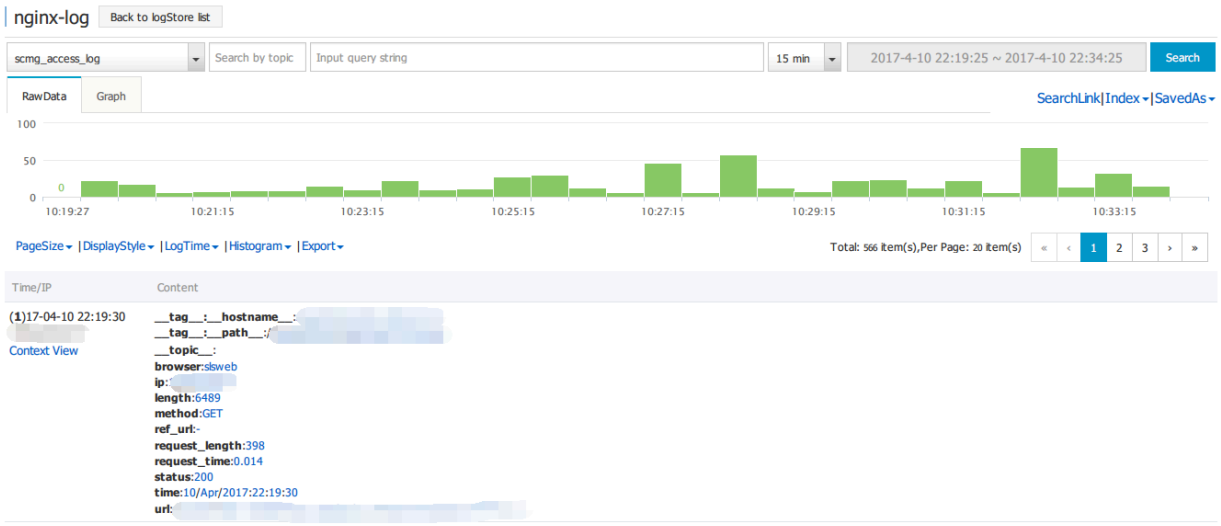
Logs collected to Log Service in the simple mode are as follows. All the contents of each log are displayed under the key named content.

Figure 3-17: Preview



Logs collected to Log Service in the full mode are as follows. The contents of each log are collected to Log Service according to the configured key-value.

Figure 3-18: Preview



Logtail configuration items

You must complete the configuration items when configuring Logtail. The descriptions and limits of the commonly used configuration items are as follows.

Configuration item	Description
Log path	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. The path that contains wildcards can match with directories of multiple levels, that is, under the specified folder (including directories of all levels), all the files that conform to the file name can be monitored.
Log file name	The name of the file from which logs are collected, which is case-sensitive and can contain wildcards, for example, `*.log`. The file name wildcards in Linux include *, “?”, and [···].

Local Storage	Whether or not to enable the local cache to temporarily store logs that cannot be sent because of short-term network interruption.
First-line log header	Specifies the starting header of a multiline log by specifying a regular expression. Lines cannot be used to separate individual logs when multiline log is collected (such as the stack information in application logs). In this case, you must specify the start line of a multi-line log. When this line is discovered, this indicates the last log has ended and a new log has begun. Therefore, you must specify a matching rule for the starting header, that is, a regular expression here.
Log parsing expression	Defines how to extract a piece of log information and convert it to a log format supported by Log Service. The user must specify a regular expression to extract the required log field informatio n and define the name of each field to be extracted.
Log time format	Defines how to parse the time format of the timestamp string in log data. For more information, see #unique_100.

Writing method of logs

In addition to using Logtail to collect logs, Log Service also provides APIs and SDKs to help you write logs.

- APIs to write logs

  Log Service provides RESTful APIs to help you write logs. You can use the #unique_102 API to write data. For more information on a complete API reference, see #unique_103.

- Use SDKs to write logs

  In addition to APIs, Log Service also provides SDKs in multiple languages (Java, .NET, PHP, and Python) to help you write logs. For more information on a complete SDK reference, see SDK reference #unique_6.

3.5.2 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 )"
10 . 1 . 1 . 1  - - [ 13 / Mar / 2016 : 10 : 00 : 11  + 0800 ] " GET
 /  HTTP / 1 . 1 "  0 . 011   180   404   570  "-" " Mozilla / 4 . 0
( compatible ;  MSIE   6 . 0 ;  Windows   NT   5 . 1 ;  360se )"
```

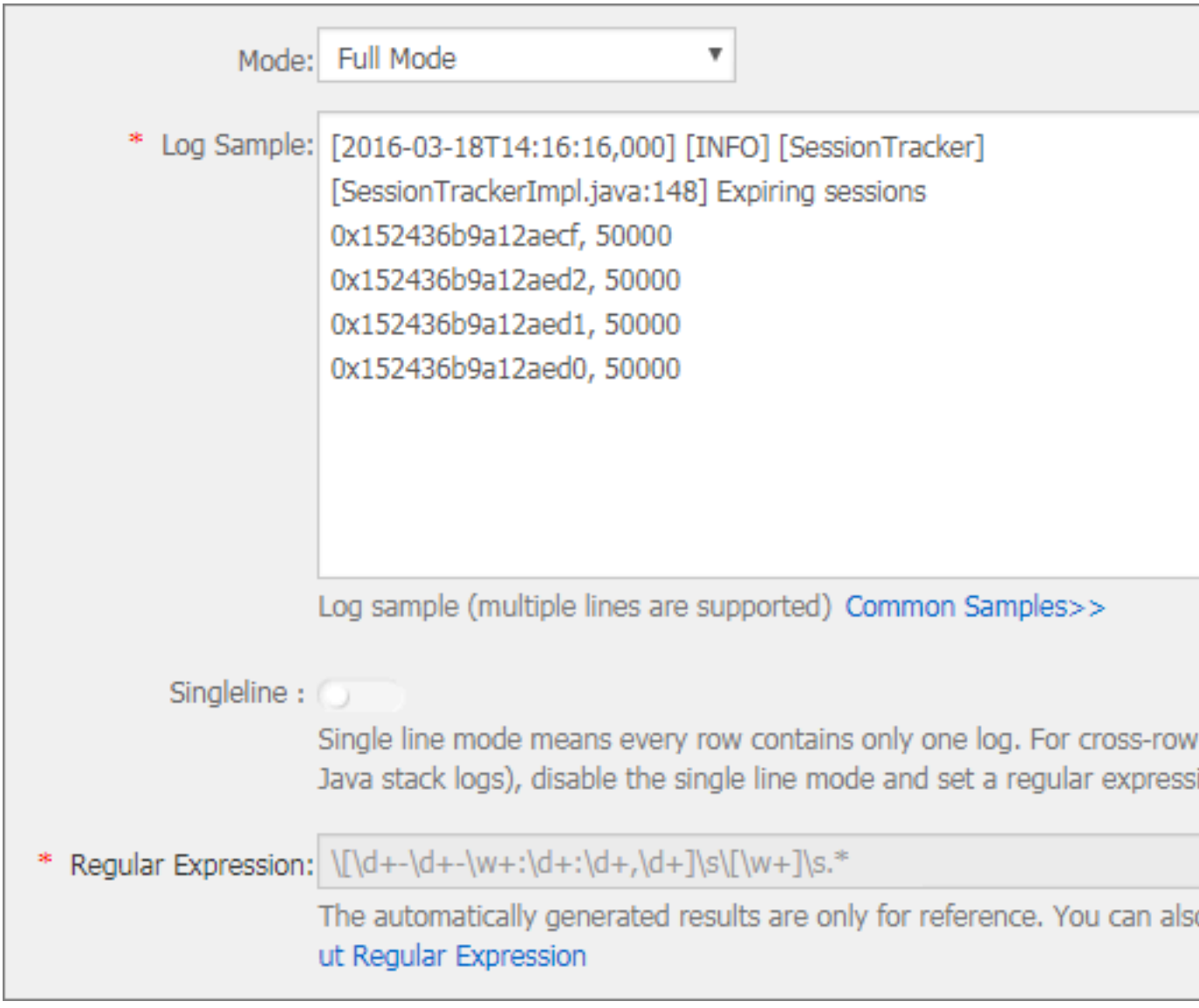
For Java applications, a program log usually spans several lines. The characteristic log header is used to separate two logs. For example, see the following Java program log:

```
[ 2016 - 03 - 18T14 : 16 : 16 , 000 ] [ INFO ] [ SessionTra  cker ] [
 SessionTra  ckerImpl . java : 148 ]  Expiring   sessions
 0x152436b9  a12aecf ,  50000
 0x152436b9  a12aed2 ,  50000
 0x152436b9  a12aed1 ,  50000
```

```
0x152436b9  a12aed0 ,  50000
```

The preceding Java log has a starting field in the time format. The regular expression is \[\d+-\d+-\w+:\d+:\d+,\d+]\s.* *. You can complete the configurations in the console as follows.

Figure 3-19: 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.

- When fields are extracted

  Regular expression: `(\ S +)\ s -\ s -\ s \[(\ S +)\ s [^]]+]\ s "(\ w +).*`, 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 : 13 / Mar / 2016 : 10 : 00 : 10` Time expression: `% d /% b /% Y :% H :% M :% S`
- The system time when the log is collected Time: Timestamp when the log is collected.

3.5.3 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.

Format	Description	Example
%a	The abbreviation of a day in a week.	Fri
%A	The full name of a day in a week.	Friday
%b	The abbreviation of a month.	Jan
%B	The full name of a month.	January
%d	The day in a month, in decimal format. Valid values: [01, 31].	07 or 31
%h	The abbreviation of a month, which is the same as `% b` .	Jan
%H	The hour in 24-hour format.	22
%I	The hour in 12-hour format.	11
%m	The month in decimal format.	08
%M	The minutes in decimal format. Valid values: [00, 59].	59
%n	The line break.	A line break
%p	The abbreviation of a period in 12-hour format.	AM or PM
%r	The time in 12-hour format, which is equivalent to `% I :% M :% S % p` .	11:59:59 AM

Format	Description	Example
%R	The time expressed in hour and minutes, which is equivalent to `% H :% M`.	23:59
%S	The seconds in decimal format. Valid values: [00, 59].	59
%t	The tab character.	The tab character
%y	The year (excluding the century) in decimal format. Valid values: [00, 99].	04 or 98
%Y	The year in decimal format.	2004 or 1998
%C	The century in decimal format. Valid values: [00, 99].	16
%e	The day in a month, in decimal format. Valid values: [1, 31]. Prefix a space to a single-digit number.	7 or 31
%j	The day in a year, in decimal format. Valid values: [00, 366].	365
%u	The day in a week, in decimal format. Valid values: [1, 7]. A value of 1 indicates Monday.	2
%U	The week in a year, where Sunday is the first day of each week. Valid values: [00, 53].	23

Format	Description	Example
%V	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].	24
%w	The day in a week, in decimal format. Valid values: [0, 6]. A value of 0 indicates Sunday.	5
%W	The week in a year, where Monday is the first day of each week. Valid values: [00, 53].	23
%c	The standard date and time.	To specify more information such as the long date or short date, you can use the preceding supported formats for more precise expression.
%x	The standard date.	To specify more information such as the long date or short date, you can use the preceding supported formats for more precise expression.
%X	The standard time.	To specify more information such as the long date or short date, you can use the preceding supported formats for more precise expression.
%s	The Unix timestamp.	1476187251

Example

The following table lists the common log time formats, examples, and corresponding time expressions.

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

3.5.4 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_ALA RM` 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 configurat ions, 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 Collect text logs.

2. Gets the configuration unique identity.

   Obtain a unique identifier for the configuration from the `user_log_c onfig . 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:

```
grep "##" / usr / local / ilogtail / user_log_c onfig . json |
awk '{ print $ 1 }'
                    "## 1 . 0 ## log - config - test $ multi "
                    "## 1 . 0 ## log - config - test $ ecs - test "
                    "## 1 . 0 ## log - config - test $ metric_sys
tem_test "
                    "## 1 . 0 ## log - config - test $ redis -
status "
```

3. Add local events.

   Save local events to JSON file `local_even t . 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 " : "${ your_confi  g_unique_i  d }",
    " dir " : "${ your_log_d  ir }",
    " name " : "${ your_log_f  ile_name }"
   },
  {
   ...
   }
   ...
]
```

   · Configuration items

Configuration items	Description:	Exam
Config	Indicates the configuration unique identifier that is obtained in step 2.	##1.0#
dir	Indicates the folder where logs are located. Note: The folder cannot end in /.	/ dat
name	Indicates a log name that support wildcards.	acce log

   Note:
   To prevent Logtail from loading invalid JSON files, we recommend that you save local event configurations to a temporary file, and after editing the temporary file, copy the content to `local_even t . json` .

   · Configuration example

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

```
$  cat  / usr / local / ilogtail / local_even  t . json
[
  {
```

```
    " config ": "## 1 . 0 ## log - config - test $ ecs - test ",
    " dir ": "/ data / log ",
    " name ": " access . log *"
  },
  {
    " config ": "## 1 . 0 ## log - config - test $ tmp - test ",
    " dir ": "/ tmp ",
    " name ": " access . log . 2017 - 08 - 09 "
  }
]
```

- How can I check whether Logtail has loaded the configuration?

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

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

  - Check whether the content in `local_even t . 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_even t . json` has been cleared, but these keywords cannot be found, the local configuration file may be invalid and has been filtered out.
  - Search the #unique_66 result for the `LOAD_LOCAL _EVENT_ALA RM` 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.

3.5.5 Set a log topic

Note:

No topic can be set for a syslog.

Topic generation modes

You can set a topic when using Logtail to collect logs or using APIs or SDKs to upload data. Currently, the following topic generation modes are supported in the console: Null - Do not generate topic, Machine Group Topic Attributes, and File Path RegEx.

- Null - Do not generate topic

  The default log topic generation mode is Null - Do not generate topic when you configure Logtail to collect text files in the console. The topic is an empty string, and you can query logs without entering a topic.

- Machine Group Topic Attributes

  The Machine Group Topic Attributes mode is used to differentiate log data of different servers. If the log data of different servers is stored in the same file or file path, you can divide the servers into different machine groups to differentiate the log data of different servers by topic. To do this, set different topics for different machine groups when creating machine groups, and select Machine Group Topic Attributes as the topic generation mode. Apply the previously created Logtail configuration to those machine groups to complete the configuration.

  With the Machine Group Topic Attributes mode selected, Logtail uploads the topic attribute of the machine group where the current server belongs as the topic name to Log Service when reporting data. When querying logs, you need to specify a topic, that is, you need to specify the topic attribute of the target machine group as the query condition.

- File Path RegEx

  - The File Path RegEx mode is used to differentiate log data generated by users and instances. If logs are stored under different directories based on different users or instances but their subdirectories and log file names are the same, Log Service cannot differentiate which user or instance generates the logs when collecting log files. To resolve this problem, you can select File Path RegEx as the topic generation mode, enter the regular expression for exact match of the file path, and set the topic name as the instance name.
  - With the File Path RegEx mode selected, Logtail uploads the instance name as the topic name to Log Service when reporting data. The topic generated varies depending on your directory structure and configuration. You need to specify the topic name as the instance name when querying logs. For example, the

following directory setting stores logs in different directories according to the users to which the logs belong:

```
/ 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, that is, `? P < key >`. The value of key can contain only lowercase letters and digits. An example is as follows:

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

The following custom tags are created for the logs:

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

Note:

Logtail 0.16.19 and later are supported.

- For the *service . log* files in */ logs* directory on the server, Log Service cannot differentiate which user or instance generates the specified logs. In this case, you can select File Path RegEx as the topic generation mode, and enter the regular expression `\/(.*)\/ serviceA \/. *` to extract the instance names. After configuration, different topics are generated for logs in different directories, including userA, userB, and userC. Then, you can specify the topic to query logs.

Note:

In the regular expression of the file path, you need to escape backslashes (/).

- Static topic generation

Select File Path RegEx as the topic generation mode, and enter `customized :// + user - defined topic name` in the Custom RegEx field.

Note:

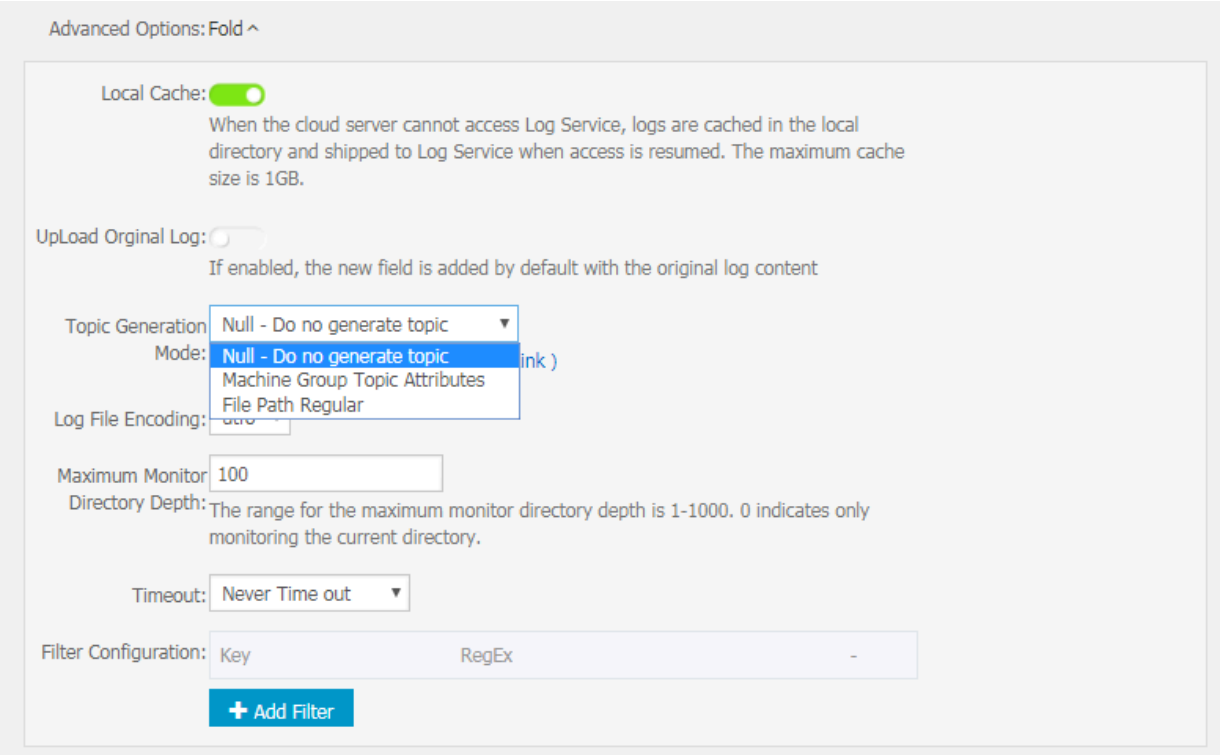
Logtail 0.16.21 and later are supported.

Set a log topic

1. Configure Logtail in the Log Service console. For more information, see Collect text logs.

   To select Machine Group Topic Attributes as the topic generation mode, set the machine group topic when creating or modifying a machine group.
2. On the Logtail configuration page, unfold Advanced Options, and select a topic generation mode from the Topic Generation Mode drop-down list.

Figure 3-20: Set the log topic



Modify a log topic

To change the log topic generation mode, modify the Topic Generation Mode option on the Logtail configuration page.

> **Note:**
> The modified configuration only applies to the data collected after the modification takes effect.

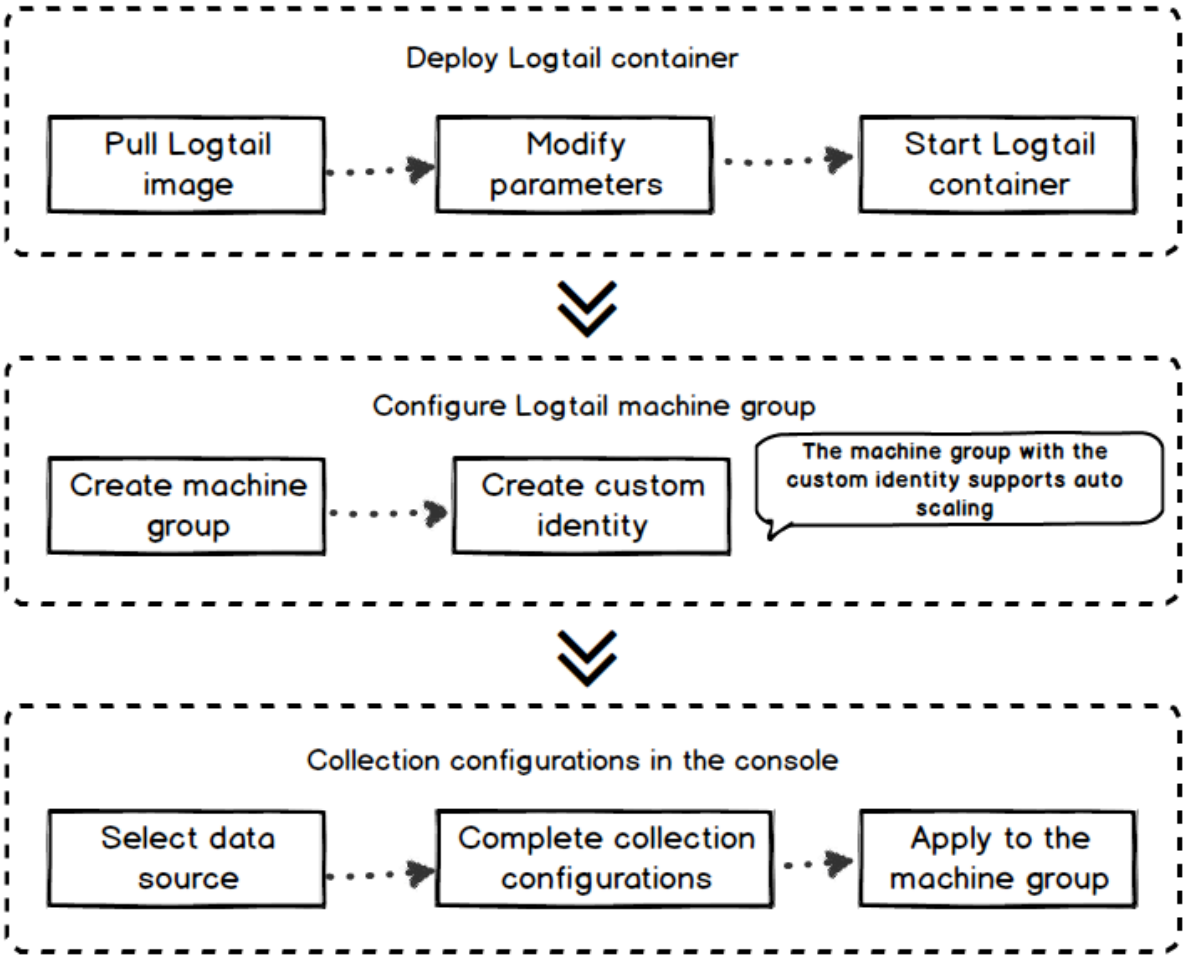
3.6 Container log collection

3.6.1 Collect standard Docker logs

Logtail supports collecting standard Docker logs and uploading these logs together with the container-related metadata information to Log Service.

Configuration process

Figure 3-21: Configuration process



1. Deploy a Logtail container.
2. Configure a Logtail machine group.

   Create a machine group with a custom ID in the Log Service console. No additional O&M is needed for further container cluster scaling.
3. Create collection configurations for the server side.

   Create collection configurations in the Log Service console. All the collection configurations are for the server side. No local configuration is needed.

Step 1. Deploy a Logtail container

1. Pull the Logtail image.

```
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.

```
docker run -d -v /:/logtail_host:ro -v /var/run/docker.sock:/var/run/docker.sock --env ALIYUN_LOGTAIL_CONFIG=/etc/ilogtail/conf/${your_region_name}/ilogtail_config.json --env ALIYUN_LOGTAIL_USER_ID=${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:**

Take either of the following actions before setting the parameters. Otherwise, the `container text file busy` error may occur when you remove another container.

- For CentOS 7.4 and later, set fs.may_detach_mounts to 1. For more information, see bug 1468249, bug 1441737, and issue 34538.

- Add the `-- privileged` flag to the startup parameters. For more information, see Docker run reference.

Parameter	Description
`${ your_regio n_name }`	The region where the Log Service project is located. Set this parameter to an appropriate value according to the network type. Valid values: - For the Internet, specify the region in the *region*-internet format. For example, the value for the China (Hangzhou) region is `cn - hangzhou - internet`. - For the Alibaba Cloud internal network, specify the region in the *region* format. For example, the value for the China (Hangzhou) region is `cn - hangzhou`. For more information about installation parameters in each region, see Table 1 Logtail installation parameters. Set this parameter according to the region of the project.
`${ your_aliyu n_user_id }`	The user ID. Set this parameter to the ID of your Alibaba Cloud account, which is of the String type. For more information about how to view the ID, see step 1 in #unique_65.
`${ your_machi ne_group_u ser_define d_id }`	The custom ID of your cluster machine group. For more information about how to set the custom ID, see step 1 in #unique_39.

```
 docker   run  - d  - v  /:/ logtail_ho  st : ro  - v  / var / run /
 docker . sock :/ var / run / docker . sock
-- env   ALIYUN_LOG  TAIL_CONFI  G =/ etc / ilogtail / conf /
 cn_hangzho  u / ilogtail_c  onfig . json  -- env
```

```
 ALIYUN_LOG  TAIL_USER_  ID = 1654218 ******-- env   ALIYUN_LOG
TAIL_USER_  DEFINED_ID = log - docker - demo   registry . cn -
hangzhou . aliyuncs . com / log - service / logtail
```

Note:

You can customize the startup parameter configurations of the Logtail container if the following conditions are met:

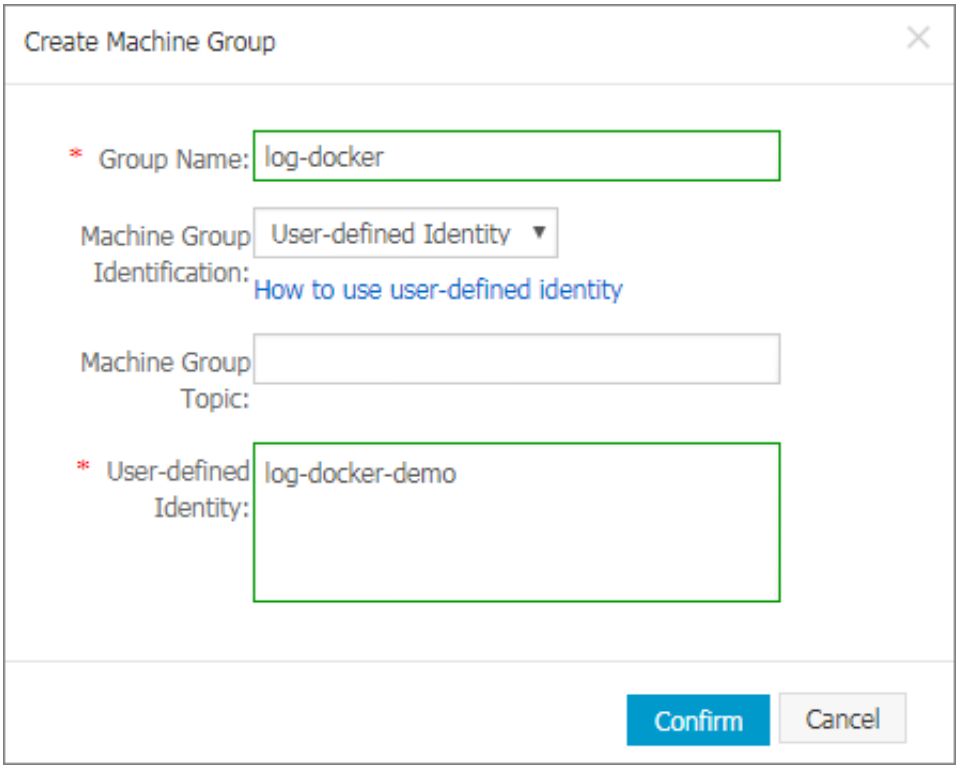
1. You have the following three environment variables before starting the Logtail container: `ALIYUN_LOG TAIL_USER_ DEFINED_ID`, `ALIYUN_LOG TAIL_USER_ ID`, and `ALIYUN_LOG TAIL_CONFI G`.
2. The domain socket of Docker is mounted to `/ var / run / docker . sock`.
3. To collect standard container output, container logs, or host files, mount the root directory to the `/ logtail_ho st` directory of the Logtail container.
4. If there is an error log `The parameter is invalid : uuid = none` in the Logtail log file */ usr / local / ilogtail / ilogtail . LOG*, create a *product_uu id* file on the host. Write any legal UUID, for example, `169E98C9 - ABC0 - 4A92 - B1D2 - AA6239C0D2 61`, to the file and mount the file to the */ sys / class / dmi / id / product_uu id* directory of the Logtail container.
5. If the live restore setting is enabled for your Docker Engine, run the `curl --unix - socket / var / run / docker . sock http :/ x > / dev / null 2 >& 1` command before the Docker Engine is restarted to verify that the domain socket used by Logtail is valid.

Step 2. Configure a Logtail machine group

1. Activate Log Service, and then create a project and a Logstore. For more information, see #unique_112.
2. Click #unique_40 on the Machine Groups page in the Log Service console.

3. Select Custom ID from the Identification drop-down list. Enter the value of `ALIYUN_LOG_TAIL_USER_DEFINED_ID` set in the previous step in the Custom ID field.

Figure 3-22: Configure the Logtail machine group



Click Confirm to create the machine group. One minute later, click Status on the right of the Machine Groups page to view the heartbeat status of the deployed Logtail container. For more information, see the View status section in #unique_113.

Step 3. Create collection configurations

Create Logtail collection configurations in the console as needed. For more information, see:

- Container text logs (recommended)
- Container standard output (recommended)
- Host text files

  By default, the root directory of a host is mounted to the `/logtail_host` directory of the Logtail container. You need to prefix the configuration path with /logtail_host. For example, to collect data in the `/home/logs/app_log/`

directory of the host, set the log path on the configuration page as `/ logtail_ho st / home / logs / app_log /`.

- #unique_42

Other operations

- Check the running status of the Logtail container

  You can run the `docker exec ${ logtail_co ntainer_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_co ntainer_id } cat / usr / local / ilogtail / app_info . json` command to view relevant information about Logtail.

- View the operational logs of Logtail

  Logtail operational logs 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 `ilogtail . LOG . x . gz`.

  Example:

```
[ root @ iZbp17enxc 2us3624wex h2Z ilogtail ]# docker exec
 a287de895e 40 tail - n 5 / usr / local / ilogtail / ilogtail
 . LOG
[ 2018 - 02 - 06 08 : 13 : 35 . 721864 ] [ INFO ] [ 8 ] [
 build / release64 / sls / ilogtail / LogtailPlu gin . cpp : 104 ]
    logtail plugin Resume : start
[ 2018 - 02 - 06 08 : 13 : 35 . 722135 ] [ INFO ] [ 8 ] [
 build / release64 / sls / ilogtail / LogtailPlu gin . cpp : 106 ]
    logtail plugin Resume : success
[ 2018 - 02 - 06 08 : 13 : 35 . 722149 ] [ INFO ] [ 8 ] [
 build / release64 / sls / ilogtail / EventDispa tcher . cpp : 369
 ] start add existed check point events , size : 0
[ 2018 - 02 - 06 08 : 13 : 35 . 722155 ] [ INFO ] [ 8 ] [
 build / release64 / sls / ilogtail / EventDispa tcher . cpp : 511
 ] add existed check point events , size : 0 cache
   size : 0 event size : 0 success count : 0
[ 2018 - 02 - 06 08 : 13 : 39 . 725417 ] [ INFO ] [ 8 ] [
 build / release64 / sls / ilogtail / ConfigMana ger . cpp : 3776 ]
    check container path update flag : 0 size : 1
```

  The container standard output is not for reference. Ignore the following standard output:

```
start umount useless mount points , / shm $|/ merged $|/
mqueue $
```

```
 umount : / logtail_ho  st / var / lib / docker / overlay2 /
 3fd0043af1  74cb0273c3  c7869500fb  e2bdb95d13  b1e110172e
 f57fe840c8  2155 / merged :  must   be   superuser   to   unmount
 umount : / logtail_ho  st / var / lib / docker / overlay2 /
 d5b10aa193  99992755de  1f85d25009  528daa749c  1bf8c16edf
 f44beab6e6  9718 / merged :  must   be   superuser   to   unmount
 umount : / logtail_ho  st / var / lib / docker / overlay2 /
 5c3125dadd  acedec29df  72ad0c52fa  c800cd56c6  e880dc4e8a
 640b1e16c2  2dbe / merged :  must   be   superuser   to   unmount
......
 xargs :  umount :  exited   with   status   255 ;  aborting
 umount   done
 start   logtail
 ilogtail   is   running
 logtail   status :
 ilogtail   is   running
```

- **Restart Logtail**

  The following sample code shows how to restart Logtail:

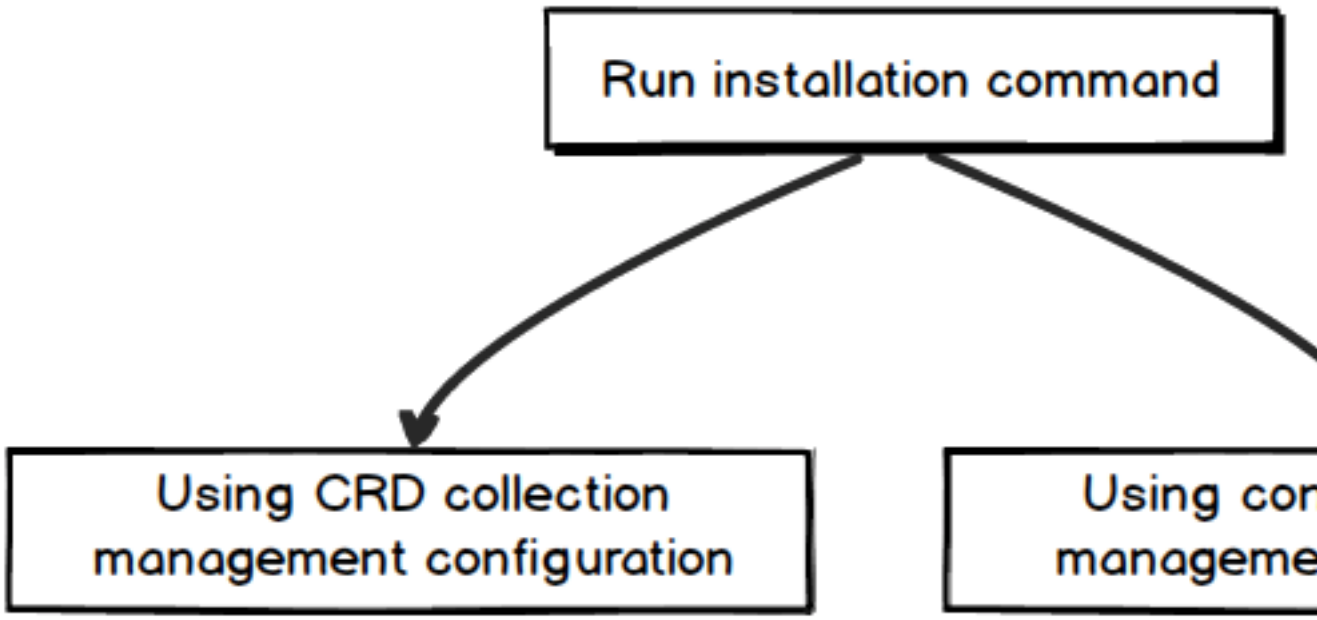
```
[ root @ iZbp17enxc  2us3624wex  h2Z   ilogtail ]#  docker   exec
 a287de895e  40  / etc / init . d / ilogtaild   stop
 kill   process   Name :  ilogtail   pid :  7
 kill   process   Name :  ilogtail   pid :  8
 stop   success
[ root @ iZbp17enxc  2us3624wex  h2Z   ilogtail ]#  docker   exec
 a287de895e  40  / etc / init . d / ilogtaild   start
 ilogtail   is   running
```

3.6.2 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

Figure 3-23: 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_116.
2. Replace `${ your_k8s_c luster_id }` in the following command with your Kubernetes cluster ID and run the command:

```
wget   http :// logtail - release - cn - hangzhou . oss - cn -
hangzhou . aliyuncs . com / kubernetes / alicloud - log - k8s -
install . sh  - O   alicloud - log - k8s - install . sh ;  chmod
```

```
    744  ./ alicloud - log - k8s - install . sh ;  sh  ./ alicloud -
  log - k8s - install . sh  ${ your_k8s_c  luster_id }
```

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 -${ your_k8s_c  luster_id }`. Under the project, machine group `k8s - group -${ your_k8s_c  luster_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 - ${ your_k8s_c  luster_id }` project. You cannot delete this Logstore because it is required for alibaba-log-controller troubleshooting.
- To store the collected logs in an existing project, run the `sh ./alicloud-log-k8s-install.sh${your_k8s_cluster_id}${your_project_name}` 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:

```
[ root @ iZbp ****** biaZ  ~]#  wget   http :// logtail - release -
 cn - hangzhou . oss - cn - hangzhou . aliyuncs . com / kubernetes
 / alicloud - log - k8s - install . sh  - O   alicloud - log - k8s
 - install . sh ;  chmod   744  ./ alicloud - log - k8s - install
 . sh ;  sh  ./ alicloud - log - k8s - install . sh   c12ba20
 ************** 86939f0b
....
....
....
 alibaba - cloud - log / Chart . yaml
 alibaba - cloud - log / templates /
 alibaba - cloud - log / templates / _helpers . tpl
 alibaba - cloud - log / templates / alicloud - log - crd . yaml
 alibaba - cloud - log / templates / logtail - daemonset . yaml
 alibaba - cloud - log / templates / NOTES . txt
 alibaba - cloud - log / values . yaml
 NAME :    alibaba - log - controller
 LAST   DEPLOYED :  Wed   May   16   18 : 43 : 06   2018
 NAMESPACE :  default
 STATUS :  DEPLOYED
 RESOURCES :
==>  v1beta1 / ClusterRol  eBinding
 NAME                      AGE
 alibaba - log - controller    0s
==>  v1beta1 / DaemonSet
 NAME       DESIRED    CURRENT    READY    UP - TO - DATE    AVAILABLE
    NODE   SELECTOR    AGE
```

```
 logtail    2          2          0        2             0           <
 none >          0s
==>  v1beta1 / Deployment
 NAME                      DESIRED    CURRENT    UP - TO - DATE
 AVAILABLE    AGE
 alibaba - log - controller    1          1          1             0
          0s
==>  v1 / Pod ( related )
 NAME                                     READY    STATUS
    RESTARTS    AGE
 logtail - ff6rf                            0 / 1      ContainerC
 reating    0           0s
 logtail - q5s87                            0 / 1      ContainerC
 reating    0           0s
 alibaba - log - controller - 7cf6d7dbb5 - qvn6w    0 / 1
 ContainerC  reating    0           0s
==>  v1 / ServiceAcc  ount
 NAME                      SECRETS    AGE
 alibaba - log - controller    1          0s
==>  v1beta1 / CustomReso  urceDefini  tion
 NAME                                     AGE
 aliyunlogc  onfigs . log . alibabaclo  ud . com    0s
==>  v1beta1 / ClusterRol  e
 alibaba - log - controller    0s
[ SUCCESS ]  install   helm   package  :  alibaba - log - controller
 success .
```

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:

Name	Description
{your-project-suffix}	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 mast enter `xxxx` .
{regionId}	The ID of the region where your project is located. You can view the #unique_17, for example, the region ID of `China East 1 ( Hangzhou )` is `cn - hangzhou`
{aliuid}	User ID (AliUid). You need to replace the parameter with your Alibaba Cloud account ID AliUid.  Note: AliUid is a string of characters. For more information about how to view your AliUid, see step 1 in #unique_117.
{access-key-id}	Your account access key ID. We recommend that you use the sub-account AccessKey and grant AliyunLogFullAccess permission. For more information, see #unique_118.

Name	Description
{access-key-secret}	Your account access key secret. We recommend that you use the sub-account AccessKey and grant AliyunLogFullAccess permission. For more information, see #unique_118.

After installation, Log Service automatically creates a machine group in the project. The machine group name is `k8s - group -${ your_k8s_c luster_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 @ iZbp1dsxxx xxqfbiaZ ~]# 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 165xxxxxxx
 x050 LTAxxxxxxx xxxx AIxxxxxxxx xxxxxxxxxx xxxxxxxxxx xe
....
....
....
 NAME : alibaba - log - controller
 LAST DEPLOYED : Fri May 18 16 : 52 : 38 2018
 NAMESPACE : default
 STATUS : DEPLOYED
 RESOURCES :
==> v1beta1 / ClusterRol eBinding
 NAME AGE
 alibaba - log - controller 0s
==> v1beta1 / DaemonSet
 NAME DESIRED CURRENT READY UP - TO - DATE
 AVAILABLE NODE SELECTOR AGE
 logtail - ds 2 2 0 2 0
 < none > 0s
==> v1beta1 / Deployment
 NAME DESIRED CURRENT UP - TO - DATE
 AVAILABLE AGE
```

```
 alibaba - log - controller    1          1          1             0
          0s
==>  v1 / Pod ( related )
 NAME                                        READY    STATUS
    RESTARTS    AGE
 logtail - ds - 7xf2d                            0 / 1      ContainerC
 reating    0          0s
 logtail - ds - 9j4bx                            0 / 1      ContainerC
 reating    0          0s
 alibaba - log - controller - 796f8496b6 - 6jxb2    0 / 1
 ContainerC  reating    0          0s
==>  v1 / ServiceAcc  ount
 NAME                       SECRETS    AGE
 alibaba - log - controller    1          0s
==>  v1beta1 / CustomReso  urceDefini  tion
 NAME                                        AGE
 aliyunlogc  onfigs . log . alibabaclo  ud . com    0s
==>  v1beta1 / ClusterRol  e
 alibaba - log - controller    0s
[ INFO ]  your   k8s   is   using   project  :  k8s - log - custom
 - xxx ,  region  :  cn - hangzhou ,  aliuid  : ***************,
 accessKeyI  d  :  LTA **********
[ SUCCESS ]  install   helm   package  :  alibaba - log - controller
 success .
```

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:

-	CRD Mode	Console mode
Operational complexity	Low	Medium
Function	Supports advanced configuration with the exception of Console mode	Medium
Complexity	Medium	Low
Network connection	Connect to the Kubernetes cluster	Connect to the Internet
Integration with deployment components	Supported	Not supported

-	CRD Mode	Console mode
Authentication method	Kubernetes authorization	Cloud account authentication

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_ho st` 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_ho st / home / logs / app_log /`.
- #unique_42

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 #unique_119.

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 `c12ba2028c xxxxxxxxxx 6939f0b` , then the installation command is

   ```
   wget http :// logtail - release - cn - hangzhou . oss - cn -
   hangzhou . aliyuncs . com / kubernetes / alicloud - log - k8s -
   install . sh - O alicloud - log - k8s - install . sh ; chmod
     744 ./ alicloud - log - k8s - install . sh ; sh ./ alicloud -
   log - k8s - install . sh c12ba2028c xxxxxxxxxx 6939f0b k8s -
   log - demo
   ```

2. After successful installation, in the Log Service console apply the historical collection configuration to the new machine group `k8s - group -${ your_k8s_c luster_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_c luster_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_c luster_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_ho st / var / lib / docker / overlay2 /
3fd0043af1 74cb0273c3 c7869500fb e2bdb95d13 b1e110172e
f57fe840c8 2155 / merged : must be superuser to unmount
umount : / logtail_ho st / var / lib / docker / overlay2 /
d5b10aa193 99992755de 1f85d25009 528daa749c 1bf8c16edf
f44beab6e6 9718 / merged : must be superuser to unmount
umount : / logtail_ho st / var / lib / docker / overlay2 /
5c3125dadd acedec29df 72ad0c52fa c800cd56c6 e880dc4e8a
640b1e16c2 2dbe / merged : must be superuser to unmount
......
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 r` 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 @ iZbp1dsu6v  77zfb40qfb  iaZ  ~]#  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 @ iZbp1dsu6v  77zfb40qfb  iaZ  ~]#  kubectl   exec   logtail
 - ds - gb92k  - n   kube - system   cat  / usr / local / ilogtail /
 app_info . json
{
   " UUID " : "",
   " hostname " : " logtail - ds - gb92k ",
   " instance_i  d " : " 0EBB2B0E - 0A3B - 11E8 - B0CE - 0A58AC1404
 02_172 . 20 . 4 . 2_15178109  40 ",
   " ip " : " 172 . 20 . 4 . 2 ",
   " logtail_ve  rsion " : " 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_tim  e " : " 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 @ iZbp1dsu6v  77zfb40qfb  iaZ  ~]#  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 / LogtailPlu  gin . cpp : 104 ]  logtail
 plugin   Resume : start
[ 2018 - 02 - 05   06 : 09 : 02 . 168807 ] [ INFO ] [ 9 ] [ build /
 release64 / sls / ilogtail / LogtailPlu  gin . cpp : 106 ]  logtail
 plugin   Resume : success
[ 2018 - 02 - 05   06 : 09 : 02 . 168822 ] [ INFO ] [ 9 ] [ build /
 release64 / sls / ilogtail / EventDispa  tcher . cpp : 369 ]  start
 add   existed   check   point   events ,  size : 0
[ 2018 - 02 - 05   06 : 09 : 02 . 168827 ] [ INFO ] [ 9 ] [ build /
 release64 / sls / ilogtail / EventDispa  tcher . 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 @ iZbp1dsu6v  77zfb40qfb  iaZ  ~]#  kubectl   exec   logtail -
 gb92k  - n   kube - system  / etc / init . d / ilogtaild   stop
 kill   process   Name :  ilogtail   pid :  7
 kill   process   Name :  ilogtail   pid :  9
 stop   success
[ root @ iZbp1dsu6v  77zfb40qfb  iaZ  ~]#  kubectl   exec   logtail -
 gb92k  - n   kube - system  / etc / init . d / ilogtaild   start
 ilogtail   is   running
```

3.6.3 Container text logs

Logtail supports collecting text logs generated in containers and uploading the collected logs together with the container metadata to Log Service.

Function features

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

- Set the log path within the container, no need to care about the mapping from this path to the host.
- Supports using labels to specify containers to be collected.
- Supports using labels to exclude specific containers.
- Supports environments to specify containers to be collected.
- Supports environments to exclude specific containers.
- Supports multiline logs (such as Java stack logs).
- Supports automatic tagging for container data.
- Supports automatic tagging for Kubernetes containers.

Limits

- Policy of stopping collection: When the container is stopped, Logtail stops collecting logs from the container after listening to the `die` event of the container (with a delay of 1–3 seconds). If a collection delay occurs during this time, it is possible to lose part of the logs before the stop.
- Docker storage drives: Currently, only overlay and overlay2 drives are supported. For other drive types, you must mount the log directory to your local PC.

- Logtail running methods: Logtail must be run as a container and follow the Logtail deployment method.
- Label: The label is the label information in the Docker inspect, not the label in the Kubernetes configuration.
- Environment: The environment is the environment information configured in the container startup.

Procedure

1. Deploy and configure the Logtail container.
2. Set the collection configuration in Log Service.

1. Logtail deployment and configuration

- Kubernetes

  For more information on Kubernetes log collection, see Collect Kubernetes logs.
- Other container management methods

  For more information on other container management methods such as Swarm and Mesos, see Collect standard Docker logs.

2. Collection configuration in Log Service

1. On the Logstore List page, click the Data Import Wizard icon to enter the configuration process.
2. Select a data source.

   Select Docker File under Third-Party Software and then click Next.
3. Configure the data source.

Configuration item	Required	Description
Docker file	Yes	Confirm if the target file being collected is a Docker file.

Configuration item	Required	Description
Label whitelist	Optional	LabelKey is required. If LabelValue is not empty, only containers whose label includes LabelKey = LabelValue are collected. If LabelValue is empty, all the containers whose label includes the LabelKey are collected.  Note: a. Key-value pairs have an OR relationship between each other, that is, a container is collected if its label includes any of the key-value pairs. b. Here the label is the label information in Docker inspect.
Label blacklist	Optional	LabelKey is required. If LabelValue is not empty, only containers whose label includes LabelKey = LabelValue are excluded. If LabelValue is empty, all the containers whose label includes the LabelKey are excluded.  Note: a. Key-value pairs have an OR relationship between each other, that is, a container is excluded if its label includes any of the key-value pairs. b. Here the label is the label information in Docker inspect.

Configuration item	Required	Description
Environment whitelist	Optional	EnvKey is required. If EnvValue is not empty, only containers whose environment includes EnvKey=EnvValue are collected. If EnvValueis empty, all the containers whose environment includes the EnvKey are collected.  Note: · Key-value pairs have an OR relationship between each other, that is, a container is collected if its environment includes any of the key-value pairs. · Here the environment is the environment information configured in the container startup.
Environment blacklist	Optional	EnvKeyis required. If EnvValue is not empty, only containers whose environment includes EnvKey=EnvValue are excluded. If EnvValue is empty, all the containers whose environment includes EnvKey are excluded.  Note: a. Key-value pairs have an OR relationship between each other, that is, a container is collected if its environment includes any of the key-value pairs. b. Here the environment is the environment information configured in the container startup.
Other configurations	-	For other collection configurations and parameter descriptions, see Collect text logs.

4. Description

   · In this topic, labels refer to label information contained in Docker inspect.

   · The namespace and container name in Kubernetes are mapped to labels `io . kubernetes . pod . namespace` and `io . kubernetes . container . name` in a docker. For example, the Pod you have created belongs to the backend-prod namespace, and the container name is worker-server. In this case,

you can configure two whitelist labels `io . kubernetes . pod . namespace : backend - prod` and `io . kubernetes . container . name : worker - server` to specify that only logs in the worker-server container can be collected.

- We recommend that you use only the `io . kubernetes . pod . namespace` and `io . kubernetes . container . name` labels in Kubernetes. For other scenarios, you can use an environment whitelist or blacklist.

5. Apply to the machine group.

   On the Apply to Machine Group page, select the Logtail machine group to be collected and click Apply to Machine Group to apply the configuration to the selected machine group. If you have not created a machine group, click Create Machine Group to create one.

6. Complete the process of accessing container text logs.

   To configure the Search, Analysis, and Visualization function and the Shipper & ETL function, complete the settings as instructed on the page.

Configuration example

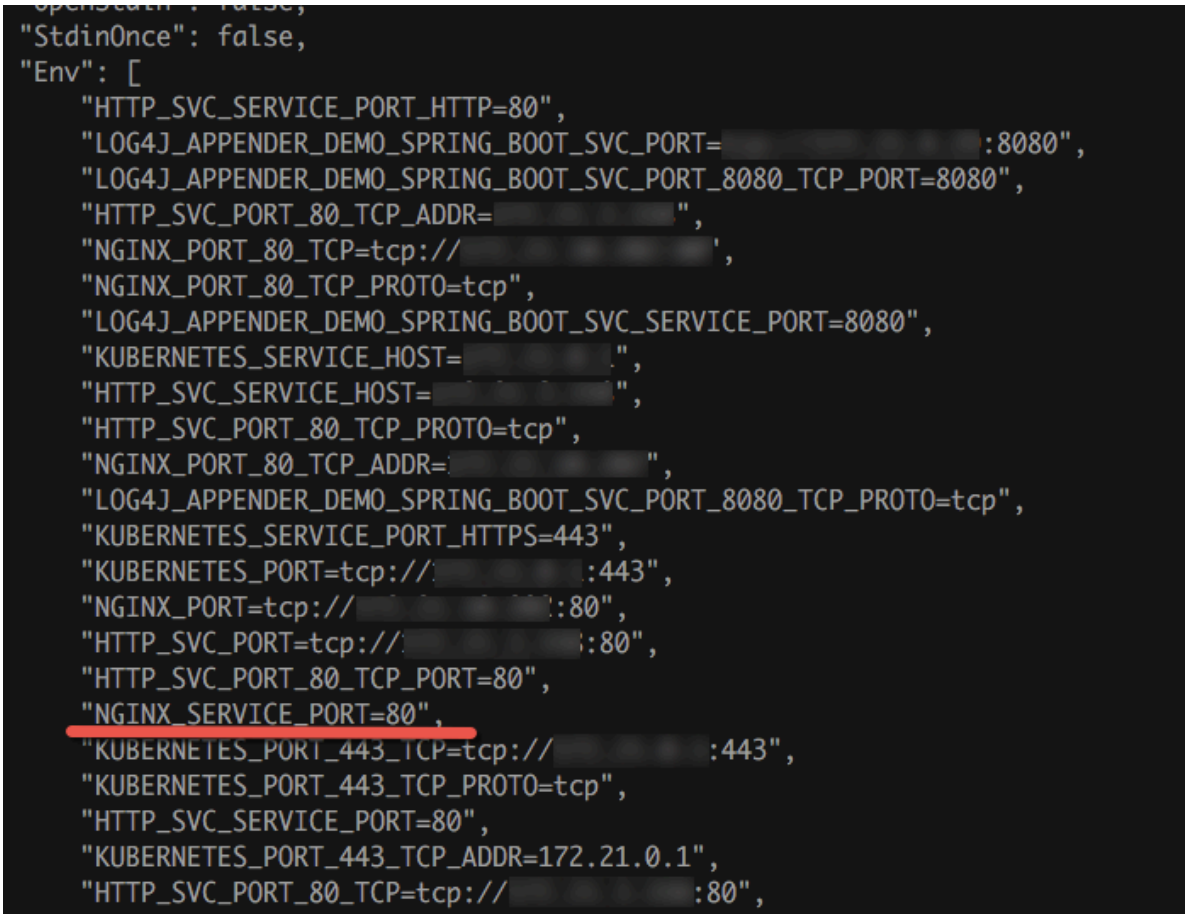
- Environment configuration

  Collect the logs of containers whose environment is `NGINX_PORT _80_TCP_PORT = 80` , and is not `POD_NAMESP ACE = kube - system` . The log file path is `/ var / log / nginx / access . log` , and the logs are parsed in the simple mode.

   **Note:**

The environment is the environment information configured in the container startup.

Figure 3-24: Example of Environment Configuration



The data source configuration in this example is as follows. For other collection configurations and parameter descriptions, see Collect text logs.

- Label configuration

Collect the logs of containers whose label is `io . kubernetes . container . name = nginx` , and is not type=pre . The log file path is `/ var / log / nginx / access . log` , and the logs are parsed in the simple mode.

Note:

The label is the label information in the Docker inspect, not the label in the Kubernetes configuration.

Figure 3-25: Example label Mode

The data source configuration in this example is as follows. For other collection configurations and parameter descriptions, see Collect text logs.

Default field

Normal Docker

The following fields are uploaded by each log by default.

Field	Description
`_image_nam e_`	Image name.
`_container _name_`	Container name
`_container _ip_`	Container IP address

Kubernetes

If the cluster is a Kubernetes cluster, the following fields are uploaded by each log by default.

Field	Description
`_image_nam e_`	Image name
`_container _name_`	Container name
`_pod_name_`	Pod name
`_namespace _`	Namespace where a Pod resides
`_pod_uid_`	The unique identifier of a Pod

Field	Description
_container_ip_	Pod IP address

3.6.4 Containers-standard output

Logtail supports using the standard output stream of the container as the input source, and uploading the standard output stream together with the container metadata to Log Service.

Features

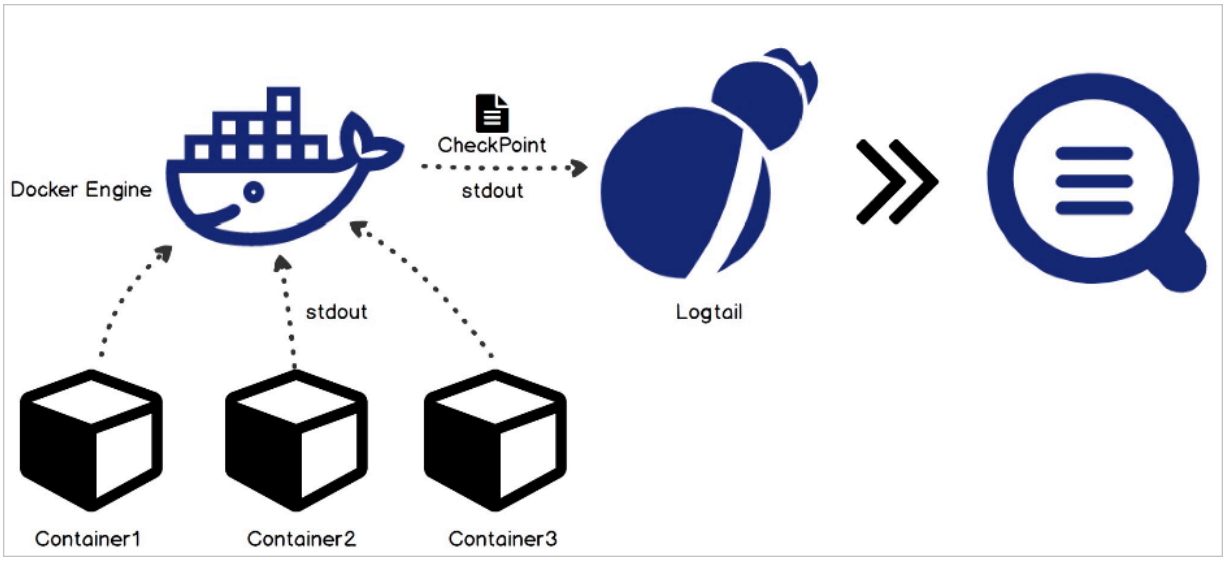
- Supports collection stdout and stderr
- Supports label specified collection containers
- Supports using labels to exclude specific containers
- Supports environments to to specify containers to be collected.
- Supports environments to exclude specific containers.
- Supports multi-line logs (like java stack logs)
- Supports automatic tagging of container data
- Supports automatic tagging for Kubernetes containers

Implementation principle

As shown in the following figure, Logtail communicates with the Domain Socket of Docker to query all of the containers running on Docker and then locate the containers to be collected according to label information. Then, Logtail uses the Docker log command to retrieve the specified container log.

When Logtail collects the standard output of the container, it periodically saves the collected point information in the checkpoint file. If Logtail is restarted after being stopped, the log will be collected from the last saved point.

Figure 3-26: Implementation principle



Limits

- Currently, this feature only supports Linux and depends on Logtail 0.16.0 and later versions. For version check and upgrade, see #unique_31.
- By default, Logtail uses `/ var / run / docker . sock` to access Docker Engine. Make sure that Domain Socket exists and has access permissions.
- Multiline log limit. To ensure that a logmade up of multiple lines is not split up due to output delay, multi-line logs will be cached for a short time by default. The default cache time is three seconds, but can be changed by using the `BeginLineT imeoutMs` parameter. However, this value cannot be less than 1000. Otherwise, the operation may be prone to error.
- Policy of stopping collection. When the container is stopped, Logtail stops collecting the standard output from the container after listening to the container to `die` event. If a collection delay occurs during this time, it is possible to lose parts of the output before the stop.
- Context limit. Each collection is deployed to the same context by default. If you need a different context for each type of container, they must be set individually.
- Data processing. The default field of collected data is `content`, which supports common processing configurations.

- Label. The label is the label information in the Docker inspect, not the label in the Kubernetes configuration.
- Environment The environment is the environment information configured in the container startup.

Configuration process

1. Deploy and configure the Logtail container.
2. Set the collection configuration in Log Service.

1. Logtail deployment and configuration

- Kubernetes
- Other container management methods

2. Collection configuration in Log Service

1. On the Logstore List page, click the Data Import Wizard icon to enter the configuration process.
2. Select the data source.

   Select Docker Stdout under Third-Party Software and then click Next.
3. Configure the data source.

   On the Configure Data Source page, complete your collection configuration. See the following example.

```
{
 " inputs ": [
     {
         " type ": " service_do  cker_stdou  t ",
         " detail ": {
             " Stdout ":  true ,
             " Stderr ":  true ,
             " IncludeLab  el ": {
                 " io . kubernetes . container . name ": " nginx "
             },
             " ExcludeLab  el ": {
                 " io . kubernetes . container . name ": " nginx -
 ingress - controller "
             },
             " IncludeEnv ": {
                 " NGINX_SERV  ICE_PORT ": " 80 "
             },
             " ExcludeEnv ": {
                 " POD_NAMESP  ACE ": " kube - system "
             }
         }
     }
 ]
```

```
}
```

4. Apply the configuration to the machine group.

   Enter the apply to machine group page. select the Logtail machine group to be collected and click Apply to Machine Group to apply the configuration to the selected machine group. If you have not created a machine group, click Create Machine Group to create one.

Description

The input source type is `service_docker_stdout`

Configuration items	Type	Required or not	Description
IncludeLabel	The mapping type, where key and value are both strings.	Required.	Empty by default. When this is empty, all Container data will be collected. When key is not empty and value is empty, all containers with a label containing this key will be collected. Note: 1. Key-value pairs have an OR relationship between each other, that is, a container is collected if its label includes any of the key-value pairs. 2. Here the label is the label information in Docker inspect.

Configuration items	Type	Required or not	Description
ExcludeLabel	The mapping type, where key and value are both strings.	Optional	Empty by default. When empty, no Containers will be excluded. When key is not empty and value is empty, all containers with a label that contains this key will be excluded. Note: 1. All key-value pairs have an OR relationship. As long as the label for a container includes one of the key-value pairs, it will be excluded. 2. Here the label is the label information in Docker inspect.
IncludeEnv	The mapping type, where key and value are both strings.	Optional	Empty by default. If empty, all containers are collected. If the key is not empty but the value is empty, all the containers whose environment includes this key are collected. Note: 1. Key-value pairs have an OR relationship between each other, that is, a container is collected if its environment includes any of the key-value pairs. 2. The environment is the environment information configured in the container startup.

Configuration items	Type	Required or not	Description
ExcludeEnv	The mapping type, where key and value are both strings.	Optional	Empty by default. If empty, no containers are excluded. If the key is not empty but the value is empty, all the containers whose environment includes this key are excluded.  Note: 1. Key-value pairs have an OR relationship between each other, that is, a container is excluded if its environment includes any of the key-value pairs. 2. The environment is the environment information configured in the container startup.
Stdout	bool	Optional	True by default. When false, stdout data will not be collected.
Stderr	bool	Optional	True by default. When false, stderr data will not be collected.
BeginLineRegex	String	Optional	Empty by default. When not empty it is the first match to the regular expression in the line . If a line matches the regular expression, then that line will be treated as a new log. Otherwise the line of data will be connected to the previous log.
BeginLineTimeoutMs	int	Optional	Timeout time for matchin a lin, measured in miliseconds, 3000 by default. Every 3 seconds, if a new log has not appeared, the last log will be output.

Configuration items	Type	Required or not	Description
BeginLineCheckLength	int	Optional	The length (in bytes) of the beginning of a row used to match with the regular expression. The default value is 10*1024. If the regular expression can match with the row within the first N bytes, configure this parameter to increase the matching efficiency.
MaxLogSize	int	Optional	The maximum length (in bytes) of a log. The default value is 512*1024. If the log exceeds this setting, it will not continue to be searched, rather it will be directly uploaded.

Default fields

Normal Docker

The following fields are uploaded by each log by default.

Field name	Description:
`_time_`	The data time. For example, `2018-02-02T02:18:41.979147844Z`.
`_source_`	The input source type, either stdout or stderr.
`_image_name_`	The image name.
`_container_name_`	The container name.

Kubernetes

If the cluster is a Kubernetes cluster, the following fields are uploaded by each log by default.

Field name	Description:
`_time_`	The data time. For example, `2018-02-02T02:18:41.979147844Z`.

Field name	Description:
`_source_`	The input source type, either stdout or stderr.
`_image_nam e_`	The image name.
`_container _name_`	The container name.
`_pod_name_`	The pod name.
`_namespace _`	The namespace where the pod resides.
`_pod_uid_`	The unique identifier for the pod.

Configuration example

General configuration

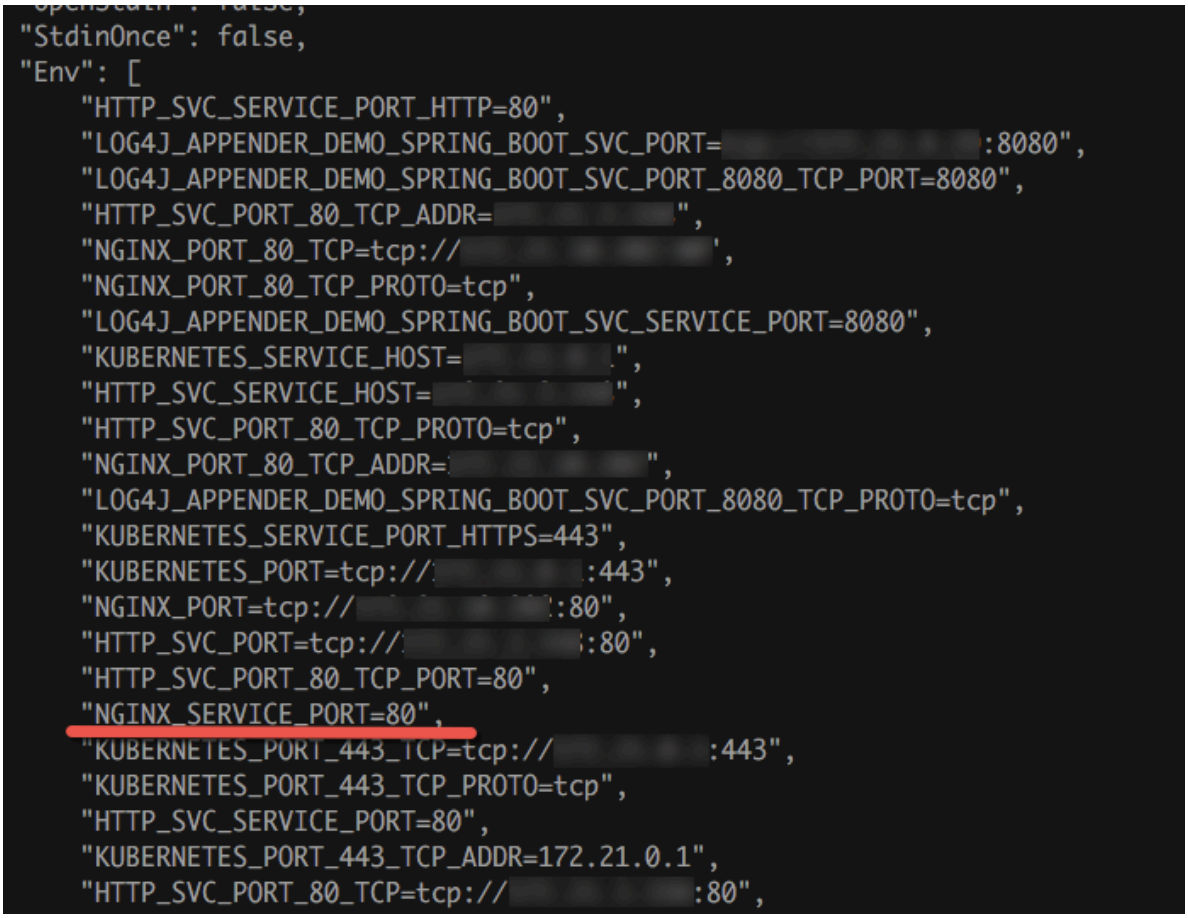
- Environment configuration configuration

  Collect the stdout logs and stderr logs of containers whose environment is `NGINX_PORT _80_TCP_PO RT = 80` , and is not `POD_NAMESP ACE = kube - system` :

   Note:

The environment is the environment information configured in the container startup.

Figure 3-27: Example of Environment Configuration



Collection configuration

```
{
    " inputs ": [
        {
            " type ": " service_do  cker_stdou  t ",
            " detail ": {
                " Stdout ":  true ,
                " Stderr ":  true ,
                " IncludeEnv ": {
                    " NGINX_PORT  _80_TCP_PO  RT ": " 80 "
                },
                " ExcludeEnv ": {
                    " POD_NAMESP  ACE ": " kube - system "
                }
            }
        }
    ]
```

```
}
```

- Label configuration

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

  > **Note:**
  > The label here is Docker not the label in the Kubernetes configuration.

  Figure 3-28: Label configuration example


```
{
    " inputs ": [
        {
            " type ": " service_do  cker_stdou  t ",
            " detail ": {
                " Stdout ":  true ,
                " Stderr ":  true ,
                " IncludeLab  el ": {
                    " io . kubernetes . container . name ": " nginx
 "
                },
                " ExcludeLab  el ": {
                    " type ": " pre "
                }
            }
        }
    ]
}
```

Collection configuration of multiline logs

Multi-line log collection is particularly important for the collection of Java exception stack output. Here we introduce a standard Java standard output log collection configuration.

- Log sample

```
2018 - 02 - 03   14 : 18 : 41 . 968   INFO  [ spring - cloud -
monitor ] [ nio - 8080 - exec - 4 ]  c . g . s . web . controller .
DemoContro  ller  :  service   start
2018 - 02 - 03   14 : 18 : 41 . 969   ERROR  [ spring - cloud -
monitor ] [ nio - 8080 - exec - 4 ]  c . g . s . web . controller .
DemoContro  ller  :  java . lang . NullPointe  rException
at   org . apache . catalina . core . Applicatio  nFilterCha  in .
internalDo  Filter ( Applicatio  nFilterCha  in . java : 193 )
at   org . apache . catalina . core . Applicatio  nFilterCha  in .
doFilter ( Applicatio  nFilterCha  in . java : 166 )
at   org . apache . catalina . core . StandardWr  apperValve .
invoke ( StandardWr  apperValve . java : 199 )
at   org . apache . catalina . core . StandardCo  ntextValve .
invoke ( StandardCo  ntextValve . java : 96 )
...
2018 - 02 - 03   14 : 18 : 41 . 968   INFO  [ spring - cloud -
monitor ] [ nio - 8080 - exec - 4 ]  c . g . s . web . controller .
DemoContro  ller  :  service   start   done
```

- Collection configuration

  Collect input logs of containers whose label is `app = monitor` and the beginning of a row is of the date type (to increase matching efficiency, only the first 10 bytes of the row is used to check for a match with the regular expression).

```
{
" inputs ": [
  {
    " detail ": {
      " BeginLineC  heckLength ":  10 ,
      " BeginLineR  egex ": "\\ d +-\\ d +-\\ d +. *",
      " IncludeLab  el ": {
        " app ": " monitor "
      }
    },
    " type ": " service_do  cker_stdou  t "
  }
]
}
```

Process collected data

Logtail supports common data processing methods for collected Docker standard output. We recommend that you use a regular expression to parse logs into time, module, thread, class, and info based on the multiline log format in the previous section.

- **Collection configuration:**

  Collect input logs of containers whose label is `app = monitor` **and the beginning of a row is of the date type (to increase matching efficiency, only the first 10 bytes of the row is used to check for a match with the regular expression).**

```
{
" inputs ": [
  {
    " detail ": {
      " BeginLineC  heckLength ":  10 ,
      " BeginLineR  egex ": "\\ d +-\\ d +-\\ d +. *",
      " IncludeLab  el ": {
        " app ": " monitor "
      }
    },
    " type ": " service_do  cker_stdou  t "
  }
],
" Processors  ":[
    {
        " type ": " processor_  regex ",
        " detail ": {
            " SourceKey ": " content ",
            " Regex ": "(\\ d +-\\ d +-\\ d + \\ d +:\\ d +:\\ d +\
\.\\ d +)\\ s +(\\ w +)\\ s +\\[([^]]+)]\\ s +\\[([^]]+)]\\ s +:\\ s
 +(. *)",
            " Keys ": [
                " time ",
                " module ",
                " thread ",
                " class ",
                " info "
            ],
            " NoKeyError ":  true ,
            " NoMatchErr  or ":  true ,
            " KeepSource ":  false
        }
    }
]
}
```

- **Sample output:**

  The output after processing the log `2018 - 02 - 03  14 : 18 : 41 . 968  INFO  [ spring - cloud - monitor ] [ nio - 8080 - exec - 4 ]  c . g . s . web . controller . DemoContro  ller  :  service  start  done` **is as follows:**

```
__tag__ : __hostname  __ : logtail - dfgef
_container  _name_ : monitor
_image_nam  e_ : registry . cn - hangzhou . aliyuncs . xxxxxxxxxx
xxxxx
_namespace  _ : default
_pod_name_ : monitor - 6f54bd5d74 - rtzc7
_pod_uid_ : 7f012b72 - 04c7 - 11e8 - 84aa - 00163f00c3  69
_source_ : stdout
```

```
_time_ : 2018 - 02 - 02T14 : 18 : 41 . 979147844Z
Time :  2018 - 02 - 02   02 : 18 : 41 .  968
level : INFO
module : spring - cloud - monitor
Thread :  fig
Class :  c . g . s .  web .  Controller .  demcontrol  ler
message : service   start   done
```

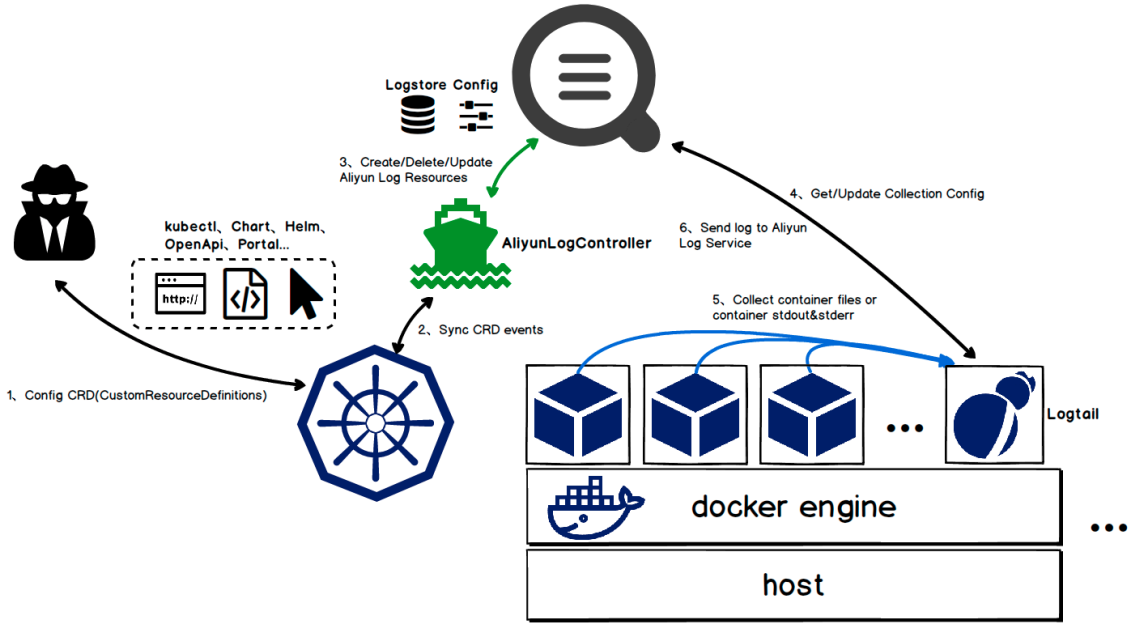
3.6.5 Configure Kubernetes log collection on CRD

Log collection is configured on the console by default. Log Service also provides CRD configuration for log collection for Kubernetes microservice development. This allows you to use kubectl to manage configurations.

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

Implementation principles

Figure 3-29: Implementation principles



Run the installation command to install the `alibaba - log - controller` Helm package. The Helm package mainly run the following operations:

1. Create aliyunlogconfigs CRD (Custom Resource Definition).
2. Deploy alibaba-log-controller.
3. Deploy Logtail DaemonSet.

The internal workflow of configuration is as follows:

1. Use `kubectl` or other tools to apply the aliyunlogconfigs CRD configuration.
2. alibaba-log-controller detects configuration update.
3. alibaba-log-controller automatically submits requests for Logstore creation, configuration creation, and configuration application to machine groups based on the CRD content and server status.
4. Logtail running in DaemonSet mode periodically sends requests for server configuration, obtains the new or updated configuration, and performs the rapid loading.
5. Logtail collects standard outputs or files from each container (pod) based on the configuration information.
6. Logtail sends processed and aggregated data to the Log Service.

Configuration method

Note:

If you have used the Logtail deployed in DaemonSet mode, you cannot manage configurations in CRD mode. For more information, see Migration process for the DaemonSet deployment mode in this document.

You must define the CRD of AliyunLogConfig to create configurations, and delete the corresponding CRD resource to delete the configuration. The CRD is configured as follows:

```
apiVersion : log . alibabaclo ud . com / v1alpha1 ## Default
value , no need for change
kind : AliyunLogC onfig ## Default value , no need for
change
metadata :
  name : simple - stdout - example ## Resource name , which
must be unique in the cluster
spec :
  logstore : k8s - stdout ## Logstore name , automatica lly
created if no name exists
  shardCount : 2 ## [ Optional ] Number of Logstore shards
. The default value is 2 . The value range is 1
to 10 .
  lifeCycle : 90 ## [ Optional ] Storage period of the
Logstore . The default value is 90 . The value range
  is 1 to 7300 . The value 7300 indicates permanent
storage .
  logtailCon fig : ## Detailed configurat ion
    inputType : plugin ## Input type of collection .
Generally , the value is file or plugin .
    configName : simple - stdout - example ## Collection
configurat ion name . The value must the same as the
  resource name ( metadata . name ).
    inputDetai l : ## Detailed configurat ion informatio n ,
see the example
```

```
      ...
```

After the configuration is completed and applied, alibaba-log-controller is created automatically.

View configuration

You can check the configuration on the Kubernetes CRD or console.

For how to view configuration on the console, see #unique_96.

> **Note:**
> If you use the CRD method to manage configuration, the configuration changes you have made on the console will be overwritten when you update configuration on the CRD.

- Run `kubectl get aliyunlogconfigs` to view all the configurations.

```
[root@iZbp1dsbiaZ ~]# kubectl get aliyunlogconfigs
NAME AGE
regex-file-example 10s
regex-stdout-example 4h
simple-file-example 5s
```

- Run `kubectl get aliyunlogconfigs ${config_name} -o yaml` to view the detailed configuration and status.

  The `status` field in the configuration shows the configuration execution result. If the configuration is successfully applied, the value of `statusCode` is 200 in the `status` field. If the value of `statusCode` is not 200, applying the configuration failed.

```
[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: |

clusterName: ""
creationTimestamp: 2018-05-17T08:44:46Z
generation: 0
name: simple-file-example
namespace: default
resourceVersion: "21790443"
selfLink: /apis/log.alibabacloud.com/v1alpha1/
namespaces/default/aliyunlogconfigs/simple-file-
example
uid: 8d3a09c4-59ae-11e8-851d-00163f008685
spec:
lifeCycle: null
```

```
logstore :  k8s - file
logtailCon  fig :
  configName :  simple - file - example
  inputDetai  l :
    dockerFile :  true
    dockerIncl  udeEnv :
      ALIYUN_LOG  TAIL_USER_  DEFINED_ID : ""
    filePatter  n :  simple . LOG
    logPath : / usr / local / ilogtail
    logType :  common_reg  _log
  inputType :  file
machineGro  ups :  null
project : ""
shardCount :  null
status :
status :  OK
statusCode :  200
```

Configuration example

Container standard output

In the container standard output, set `inputType` to `plugin` and fill the detailed information in the `plugin` field under `inputDetai l` . For more information on the configuration fields, see #unique_10.

- Simple collection mode

  Collect standard outputs (stdout and stdeer) of all containers except for those who has environment variable configuration `COLLECT_ST DOUT_FLAG = false` .

```
apiVersion :  log . alibabaclo  ud . com / v1alpha1
kind :  AliyunLogC  onfig
metadata :
 #  your  config  name ,  must  be  unique  in  your  k8s
cluster
  name :  simple - stdout - example
spec :
 #  logstore  name  to  upload  log
  logstore :  k8s - stdout
 #  logtail  config  detail
  logtailCon  fig :
   #  docker  stdout ' s  input  type  is  ' plugin '
    inputType :  plugin
   #  logtail  config  name ,  should  be  same  with  [
metadata . name ]
    configName :  simple - stdout - example
    inputDetai  l :
      plugin :
        inputs :
         -
           #  input  type
            type :  service_do  cker_stdou  t
            detail :
             #  collect  stdout  and  stderr
              Stdout :  true
              Stderr :  true
```

```
              #  collect   all   container ' s   stdout   except
containers   with  " COLLECT_ST  DOUT_FLAG : false "  in   docker
env   config
               ExcludeEnv :
                 COLLECT_ST  DOUT_FLAG : " false "
```

- Custom collection mode

  Collect the access log of Grafana and parse the access log into structured data.

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

  Figure 3-30: Custom collection mode

  The access log of Grafana is in the following format:

```
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_add  r = 172
. 16 . 64 . 154   time_ms = 0   size = 29   referer =
```

  Parse the access log using a regular expression. The detailed configuration is as follows:

```
apiVersion :  log . alibabaclo  ud . com / v1alpha1
kind :  AliyunLogC  onfig
metadata :
 #  your   config   name ,  must   be   unique   in   your   k8s
cluster
  name :  regex - stdout - example
spec :
 #  logstore   name   to   upload   log
  logstore :  k8s - stdout - regex
 #  logtail   config   detail
  logtailCon  fig :
   #  docker   stdouts   input   type   is   plugin
    inputType :  plugin
   #  logtail   config   name ,  should   be   same   with  [
metadata . name ]
    configName :  regex - stdout - example
```

```
    inputDetai  l :
      plugin :
        inputs :
         -
           #  input   type
            type :  service_do  cker_stdou  t
            detail :
             #  Collect   stdout   outputs   only   and   do   not
  collect   stdeer   outputs .
              Stdout :  true
              Stderr :  false
             #  Collect   only   stdout   outputs   whose   key
is  " GF_INSTALL  _PLUGINS "  in   the   environmen  t   variable
configurat  ion   from   the   container .
              IncludeEnv :
                GF_INSTALL  _PLUGINS : ''
        processors :
         -
           #  Use   a   regular   expression
            type :  processor_  regex
            detail :
             #  The   data   collected   by   the   docker   has
key  " content "  by   default .
              SourceKey :  content
             #  Regular   expression   for   extraction
              Regex : ' t =(\ d +-\ d +-\ w +:\ d +:\ d +\+\ d +)
lvl =(\ w +)  msg ="([^"]+)"  logger =(\ w +)  userId =(\ w +)
orgId =(\ w +)  uname =(\ S *)  method =(\ w +)  path =(\ S +)
status =(\ d +)  remote_add  r =(\ S +)  time_ms =(\ d +)  size =(\
d +)  referer =(\ S *). *'
             #  Extracted   keys
              Keys : [' time ', ' level ', ' message ', ' logger ',
' userId ', ' orgId ', ' uname ', ' method ', ' path ', ' status ',
' remote_add  r ', ' time_ms ', ' size ', ' referer ']
             #  Retain   the   original   fields
              KeepSource :  true
              NoKeyError :  true
              NoMatchErr  or :  true
```

After the configuration is applied, the data collected by Log Service is as follows:

Figure 3-31: Collected log data

Container file

- **Simple file**

  Collect log files from containers whose environment variable configuration contains key `ALIYUN_LOG_TAIL_USER_DEFINED_ID`. The log file path is `/data/logs/app_1` and the file name is `simple.LOG`.

```
apiVersion: log.alibabacloud.com/v1alpha1
kind: AliyunLogConfig
metadata:
  # your config name, must be unique in your k8s cluster
  name: simple-file-example
spec:
  # logstore name to upload log
  logstore: k8s-file
  # logtail config detail
  logtailConfig:
    # log file's input type is 'file'
    inputType: file
    # logtail config name, must same with [metadata.name]
    configName: simple-file-example
    inputDetail:
      # Set logType to "common_reg_log" for simple mode logs
      logType: common_reg_log
      # Log file folder
      logPath: /data/logs/app_1
      # File name, which supports wildcards, for example, log_*.log
      filePatter n: simple.LOG
      # Collect files from the container. dockerFile flag is set to true
      dockerFile: true
      # Only collect container with "ALIYUN_LOG_TAIL_USER_DEFINED_ID" in docker env config
      dockerIncludeEnv:
        ALIYUN_LOG_TAIL_USER_DEFINED_ID: ""
```

- **Complete regular expression files**

  The following is an example of a Java program log:

```
[2018-05-11T20:10:16,000][INFO][SessionTracker]
[SessionTrackerImpl.java:148] Expiring sessions
java.sql.SQLExcepti on: Incorrect string value: '\xF0\x9F\x8E\x8F",...' for column 'data' at row 1
at org.springfram ework.jdbc.support.AbstractFa
llbackSQLE xceptionTr anslator.translate(AbstractFa
llbackSQLE xceptionTr anslator.java:84)
at org.springfram ework.jdbc.support.AbstractFa
llbackSQLE xception
```

  A log entry may be divided into multiple lines because the log contains error stacking information. Therefore, you must set a regular expression for the

beginning of a line. To extract each field, use a regular expression. The detailed configuration is as follows:

```
apiVersion :  log . alibabaclo  ud . com / v1alpha1
kind :  AliyunLogC  onfig
metadata :
 #  your   config   name ,  must   be   unique   in   your   k8s
cluster
  name :  regex - file - example
spec :
 #  logstore   name   to   upload   log
  logstore :  k8s - file
  logtailCon  fig :
   #  log   file ' s   input   type   is  ' file '
    inputType :  file
   #  logtail   config   name ,  should   be   same   with  [
metadata . name ]
    configName :  regex - file - example
    inputDetai  l :
     #  Set   logType   to  " common_reg  _log "  for   logs   of
the   regular   expression   type .
      logType :  common_reg  _log
     #  Log   file   folder
      logPath : / app / logs
     #  File   name ,  which   supports   wildcards ,  for
example ,  log_ *. log
      filePatter  n :  error . LOG
     #  Regular   expression   for   first   line
      logBeginRe  gex : '\[\ d +-\ d +-\ w +:\ d +:\ d +,\ d +]\ s
\[\ w +]\ s . *'
     #  Parse   the   regular   expression
      regex : '\[([^]]+)]\ s \[(\ w +)]\ s \[(\ w +)]\ s \[([^:]+):
(\ d +)]\ s (. *)'
     #  List   of   extracted   keys
      key  : [" time ", " level ", " method ", " file ", " line ",
" message "]
     #  Logs   in   regular   expression . ` time `  in   the
logs   are   extracted   for   time   parsing   by   default .  If
  time   is   not   required ,  ignore   the   field .
      timeFormat : '% Y -% m -% dT % H :% M :% S '
     #  Collect   files   from   the   container .  dockerFile
flag   is   set   to   true
      dockerFile :  true
     #  Only   collect   container   with  " ALIYUN_LOG
TAIL_USER_  DEFINED_ID "  in   docker   env   config
      dockerIncl  udeEnv :
```

```
          ALIYUN_LOG  TAIL_USER_  DEFINED_ID : ""
```

After the configuration is applied, the data collected by Log Service is as follows:

Figure 3-32: Collected log data

- Delimiter pattern file

  Logtail supports log parsing in delimiter mode, an example is as follows:

```
apiVersion :  log . alibabaclo  ud . com / v1alpha1
kind :  AliyunLogC  onfig
metadata :
 #  your   config   name ,  must   be   unique   in   your   k8s
cluster
  name :  delimiter - file - example
spec :
 #  logstore   name   to   upload   log
  logstore :  k8s - file
  logtailCon  fig :
   #  log   file ' s   input   type   is  ' file '
    inputType :  file
    configName :  delimiter - file - example
   #  logtail   config   name ,  should   be   same   with  [
metadata . name ]
    inputDetai  l :
     #  Set   logType   to   delimiter_  log   for   logs   of
the   delimiter   type
      logType :  delimiter_  log
     #  Log   file   folder
      logPath : / usr / local / ilogtail
     #  File   name ,  which   supports   wildcards ,  for
example ,  log_ *. log
      filePatter  n :  delimiter_  log . LOG
     #  Use   a   multi - character   delimiter
      separator : '|&|'
     #  List   of   extracted   keys
      key  : [' time ', ' level ', ' method ', ' file ', ' line ',
' message ']
     #  Keys   for   parsing   time .  Ignore   the   field   if
time   parsing   is   not   required
      timeKey : ' time '
     #  Time   parsing   method .  Ignore   the   field   if   time
  parsing   is   not   required
      timeFormat : '% Y -% m -% dT % H :% M :% S '
     #  Collect   files   from   the   container .  dockerFile
flag   is   set   to   true
      dockerFile :  true
```

```
      #  Only   collect   container   with  " ALIYUN_LOG
TAIL_USER_  DEFINED_ID "  in   docker   env   config
      dockerIncl  udeEnv :
        ALIYUN_LOG  TAIL_USER_  DEFINED_ID : ''
```

- **JSON mode file**

  If each data line in a file is a JSON object, you can use the JSON method for parsing, an example is as follows:

```
apiVersion :  log . alibabaclo  ud . com / v1alpha1
kind :  AliyunLogC  onfig
metadata :
 #  your   config   name ,  must   be   unique   in   you   k8s
cluster
  name :  json - file - example
spec :
 #  logstore   name   to   upload   log
  logstore :  k8s - file
  logtailCon  fig :
   #  log   file ' s   input   type   is  ' file '
    inputType :  file
   #  logtail   config   name ,  should   be   same   with  [
metadata . name ]
    configName :  json - file - example
    inputDetai  l :
     #  Set   logType   to   json_log   for   logs   of   the
delimiter   type
      logType :  json_log
     #  Log   file   folder
      logPath : / usr / local / ilogtail
     #  File   name ,  which   supports   wildcards ,  for
example ,  log_ *. log
      filePatter  n :  json_log . LOG
     #  Keys   for   parsing   time .  Ignore   the   field   if
time   parsing   is   not   required
      timeKey : ' time '
     #  Time   parsing   method .  Ignore   the   field   if   time
  parsing   is   not   required
      timeFormat : '% Y -% m -% dT % H :% M :% S '
     #  Collect   files   from   the   container .  dockerFile
flag   is   set   to   true
      dockerFile :  true
     #  Only   collect   container   with  " ALIYUN_LOG
TAIL_USER_  DEFINED_ID "  in   docker   env   config
      dockerIncl  udeEnv :
        ALIYUN_LOG  TAIL_USER_  DEFINED_ID : ""
```

3.6.6 Kubernetes-Sidecar log collection mode

Logtail can collect logs in Sidecar mode from Kubernetes and create a Sidecar container for each service container requiring log collection, thereby facilitating multi-tenant isolation and improving collection performance.

Currently, the default log component installed in Kubernetes clusters is DaemonSet , which simplifies O&M operations, occupies a few resources, supports collection of container stdout and container files, and can be flexibly configured.

However, in DaemonSet mode, Logtail needs to collect logs from all container on a node. This leads to a bottleneck in performance and does not allow for total isolation among service logs. To resolve the preceding issue, Logtail now provides the Sidecar mode, which enables Logtail to create a Sidecar container for each service container requiring log collection. This mode greatly strengthens multi-tenant isolation and improves the collection performance. We recommend that you use the Sidecar mode for large-scale Kubernetes clusters and for clusters that function as a PaaS platform to serve multiple services.

Features

- The Sidecar mode can be applied to Container Service for Kubernetes, on-premises ECS Kubernetes, and on-premises Kubernetes in IDCs.
- In Sidecar mode, Logtail can collect Pod metadata, including the Pod name, Pod IP address, Pod namespace, and name and IP address of the node to which the Pod belongs.
- In Sidecar mode, Logtail can automatically create Log Service resources through CustomResourceDefinition (CRD), including projects, Logstores, indexes, Logtail Configs, and machine groups.
- The Sidecar mode supports dynamic scaling. You can adjust the number of replicas at any time, and the changes take effect immediately.

Concepts

In Sidecar mode, log collection requires that Logtail share the log directory with the service container. Briefly, the service container writes logs into the log directory, and Logtail monitors changes on log files in the log directory and collects logs. For more information, see the following official documents:

1. Introduction to the Sidecar log collection mode
2. Example of the Sidecar mode

Prerequisites

1. You have activated Log Service.

   If you have not activated Log Service, activate it.
2. You have installed #unique_44 for CRD-based settings.

Limits

1. Logtail must share the log directory with the service container.

2. Sidecar mode does not support collection of container stdout.

Sidecar configuration

The Sidecar configuration involves:

1. Setting basic operation parameters
2. Setting the mount path

The following is an example:

```
apiVersion :  batch / v1
kind :  Job
metadata :
  name :  nginx - log - sidecar - demo
  namespace :  default
spec :
  template :
    metadata :
      name :  nginx - log - sidecar - demo
    spec :
      restartPol  icy :  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 "]
        volumeMoun  ts :
       -  name :  nginx - log
          mountPath : / var / log / nginx
     #####  logtail   sidecar   container
     -  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
        livenessPr  obe :
          exec :
            command :
           - / etc / init . d / ilogtaild
           -  status
          initialDel  aySeconds :  30
          periodSeco  nds :  30
        resources :
          limits :
            memory :  512Mi
          requests :
            cpu :  10m
            memory :  30Mi
        env :
         #####  base   config
         #  user   id
         -  name : " ALIYUN_LOG  TAIL_USER_  ID "
            value : "${ your_aliyu  n_user_id }"
         #  user   defined   id
         -  name : " ALIYUN_LOG  TAIL_USER_  DEFINED_ID "
            value : "${ your_machi  ne_group_u  ser_define  d_id }"
```

```
          #  config   file   path   in   logtail ' s   container
          -  name : " ALIYUN_LOG  TAIL_CONFI  G "
             value : "/ etc / ilogtail / conf /${ your_regio  n_config
 }/ ilogtail_c  onfig . json "
          #####  env   tags   config
          -  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
        volumeMoun  ts :
        -  name :  nginx - log
           mountPath : / var / log / nginx
      #####  share   this   volume
       volumes :
      -  name :  nginx - log
         emptyDir : {}
```

Configuration 1: Set basic operation parameters.

The following shows major parameters and their settings:

```
#####  base   config
          #  user   id
          -  name : " ALIYUN_LOG  TAIL_USER_  ID "
             value : "${ your_aliyu  n_user_id }"
          #  user   defined   id
          -  name : " ALIYUN_LOG  TAIL_USER_  DEFINED_ID "
             value : "${ your_machi  ne_group_u  ser_define  d_id }"
          #  config   file   path   in   logtail ' s   container
          -  name : " ALIYUN_LOG  TAIL_CONFI  G "
```

```
            value : "/ etc / ilogtail / conf /${ your_regio  n_config
 }/ ilogtail_c  onfig . json "
```

Parameter	Description
`${ your_regio n_config }`	This parameter is determined by the region and network type of the project. Set the parameter to an appropriate value according to the network type. Valid values: · For the Internet: `region - internet`. For example, the value for the China (Hangzhou) region is `cn - hangzhou - internet`. · For Alibaba Cloud intranet: `region`. For example, the value for the China (Hangzhou) region is `cn - hangzhou`. In this parameter, region is a #unique_26/unique_26_Connect_42_table_eyz_pmv_vdb. Set it to the region to which the project belongs.
`${ your_aliyu n_user_id }`	This parameter specifies the user ID, which must be replaced with your Alibaba Cloud account ID in string format. For more information about how to query your ID, see section 2.1 in user ID configuration. Note: This parameter value must be your Alibaba Cloud account ID. RAM user IDs do not take effect.
`${ your_machi ne_group_u ser_define d_id }`	This parameter specifies the custom ID of a machine group in your cluster. The ID must be unique within the region where Log Service is deployed. For more information, see #unique_39.

Configuration 2: Set the mount path.

1. Logtail and the service container must be mounted to the same directory.
2. The `emptyDir` mount method is recommended.

The mount path example is shown in the preceding configuration example.

Log collection settings

Log collection can be set through CRD or the Log Service console. CRD-based settings support automatic creation of projects, Logstores, indexes, machine groups, and Logtail Configs, and can be easily integrated with Kubernetes. Therefore, CRD-based settings are recommended. Console-based settings are easier for users who debug or use Kubernetes log collection for the first time.

CRD-based settings

For more information, see #unique_119. Compared with the DaemonSet collection mode, CRD-based settings are subjected to the following limits:

1. You must specify the name of the project requiring log collection. Otherwise, logs are collected and sent to the project where the log component is installed by default.
2. You must specify the machine group for the settings to take effect. Otherwise, the settings are applied to the machine group to which the DaemonSet belongs by default.
3. The Sidecar mode supports only file collection, during which, `dockerFile` must be set to false.

For more information, see the corresponding example.

Console-based settings

1. Configure the machine group.

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

   a. Activate Log Service and create a project and a Logstore as needed. For more information, see #unique_112.
   b. On the machine group list page, click Create Machine Group.
   c. Set the identification to the custom ID `ALIYUN_LOG_TAIL_USER_DEFINED_ID`.

2. Set the collection mode.

   Set collection details for the target file. Currently, various modes are supported, such as the simple mode, Nginx access mode, delimiter mode, JSON mode, and regular mode. For more information, see Collect text logs.

   The settings of this example is shown in the following figure.

    Note:

Docker File must be disabled.


* 配置名称:
nginx-log-sidecar
* 日志路径:
/var/log/nginx
/**/
access
指定文件夹下所有符合文件名称的文件都会被监控到(包含所有层次的目
符模式匹配。Linux文件路径只支持/开头，例：/apsara/nuwa/.../app.Lo
如：C:\Program Files\Intel\...*.Log
是否为Docker文件:
如果是Docker容器内部文件，可以直接配置内部路径与容器Tag，Logta
进行过滤采集指定容器的日志，具体说明参考文档链接
模式:
分隔符模式
如何设置Delimiter类型配置
日志样例:
2018-09-26T03:16:53.033307075Z 10.200.98.220 - - "POST /PutData
Category=YunOsAccountOpLog&AccessKeyId=Uxxxx45A&Date=Fri%
A53%3A30%20GMT&Topic=raw&Signature=pD12XYLmGxKQ%2Bm
18204 200 37 "-" "aliyun-sdk-java" 1
请贴入需要解析的日志样例(支持多条) 常见样例>>
* 分隔符:
空格
引用符:
双引号
双引号（"）作为Quote时，内部包含分隔符的字段需要被一对Quote包
包含空格、制表符等字符，请修改格式。

Examples

Scenario:

1. **The Kubernetes cluster is an on-premises cluster in an IDC, and the region where Log Service is deployed is China (Hangzhou). Logs are collected from the Internet.**
2. **In the following examples, the mount object is `nginx - log` , and the mount type is `emptyDir` . They are mounted to the `/ var / log / nginx` directory in the nginx-log-demo and logtail containers, respectively.**
3. **The access log is `/ var / log / nginx / access . log` , and the destination Logstore is `nginx - access` .**
4. **The error log is `/ var / log / nginx / error . log` , and the destination Logstore is `nginx - error` .**

- **Sidecar settings:**

```
apiVersion :  batch / v1
kind :  Job
metadata :
  name :  nginx - log - sidecar - demo
  namespace :  default
spec :
  template :
    metadata :
      name :  nginx - log - sidecar - demo
    spec :
      restartPol  icy :  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 "]
        volumeMoun  ts :
       -  name :  nginx - log
          mountPath : / var / log / nginx
     #####  logtail   sidecar   container
     -  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
        livenessPr  obe :
          exec :
            command :
           - / etc / init . d / ilogtaild
           -  status
          initialDel  aySeconds :  30
          periodSeco  nds :  30
        env :
         #####  base   config
```

```
          #  user   id
          -  name : " ALIYUN_LOG  TAIL_USER_  ID "
             value : " xxxxxxxxxx "
          #  user   defined   id
          -  name : " ALIYUN_LOG  TAIL_USER_  DEFINED_ID "
             value : " nginx - log - sidecar "
          #  config   file   path   in   logtail ' s   container
          -  name : " ALIYUN_LOG  TAIL_CONFI  G "
             value : "/ etc / ilogtail / conf / cn - hangzhou -
 internet / ilogtail_c  onfig . json "
          #####  env   tags   config
          -  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
        volumeMoun  ts :
        -  name :  nginx - log
           mountPath : / var / log / nginx
      #####  share   this   volume
       volumes :
      -  name :  nginx - log
         emptyDir : {}
```

- **CRD settings:**

```
#  config   for   access   log
 apiVersion :  log . alibabaclo  ud . com / v1alpha1
 kind :  AliyunLogC  onfig
 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_LOG  TAIL_USER_  DEFINED_ID
 ]
   machineGro  ups :
  -  nginx - log - sidecar
  #  logtail   config   detail
   logtailCon  fig :
```

```
    #  log   file ' s   input   type   is  ' file '
     inputType :  file
    #  logtail   config   name ,  should   be   same   with  [
 metadata . name ]
     configName :  nginx - log - access - example
     inputDetai  l :
      #  Simple   logs   with   logType   set   to   common_reg
 _log
       logType :  common_reg  _log
      #  Log   folder
       logPath : / var / log / nginx
      #  File   name   with   wildcards   supported ,  for   example
 ,  log_ *. log
       filePatter  n :  access . log
      #  Sidecar   mode   with   dockerFile   set   to   false
       dockerFile :  false
      #  Line   start   regular   expression ,  which   is   set
 to  .*  is   the   log   contains   only   a   line
       logBeginRe  gex : '. *'
      #  Regular   expression   for   parsing
       regex : '(\ S +)\ s (\ S +)\ s \ S +\ s \ S +\ s "(\ S +)\ s
 (\ S +)\ s +([^"]+)"\ s +(\ S +)\ s (\ S +)\ s (\ d +)\ s (\ d +)\
 s (\ S +)\ s "([^"]+)"\ s . *'
      #  List   of   the   extracted   keys
       key  : [" time ", " ip ", " method ", " url ", " protocol ",
 " latency ", " payload ", " status ", " response - size ", ser -
 agent "]
#  config   for   error   log
```

```
#  config   for   error   log
 apiVersion :  log . alibabaclo  ud . com / v1alpha1
 kind :  AliyunLogC  onfig
 metadata :
  #  your   config   name ,  must   be   unique   in   you   k8s
 cluster
   name :  nginx - log - error - example
 spec :
  #  project   name   to   upload   log
   project :  k8s - nginx - sidecar - demo
  #  logstore   name   to   upload   log
   logstore :  nginx - error
  #  machine   group   list   to   apply   config ,  should   be
 same   with   your   sidecar ' [ ALIYUN_LOG  TAIL_USER_  DEFINED_ID
 ]
   machineGro  ups :
  -  nginx - log - sidecar
  #  logtail   config   detail
   logtailCon  fig :
    #  log   file ' s   input   type   is  ' file '
     inputType :  file
    #  logtail   config   name ,  should   be   same   with  [
 metadata . name ]
     configName :  nginx - log - error - example
     inputDetai  l :
      #  Simple   logs   with   logType   set   to   common_reg
 _log
       logType :  common_reg  _log
      #  Log   folder
       logPath : / var / log / nginx
      #  File   name   with   wildcards   supported ,  for   example
 ,  log_ *. log
       filePatter  n :  error . log
      #  Sidecar   mode   with   dockerFile   set   to   false
```

```
        dockerFile :  false
```

- View log collection results

  After the preceding settings are applied to the Kubernetes cluster, the Logtail container automatically creates the corresponding project, Logstore, machine group, and Logtail Config, and automatically sends the collected logs to Log Service. You can log on to the Log Service console to view details.

3.7 Logtail limits

Table 3-4: File collection

Item	Capability and limit
File encoding	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.
Log file size	Unlimited.
Log file rotation	Supported. Both `. log *` and `. log` are supported for file names.
Log collection behavior when log parsing is blocked	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.
Symbolic link	Monitored directories can be symbolic links.

Item	Capability and limit
Single log size	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.
Regular expression	Regular expressions can be Perl-compatible regular expressions.
Multiple collection configurations for the same file	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.
File opening behavior	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).
First log collection behavior	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.
Non-standard text log	If a log contains the '\0' lines, the log is truncated at the position of the first '\0' line.

Table 3-5: Checkpoint management

Item	Capability and limit
Checkpoint timeout period	If a file is not modified for more than 30 days, the checkpoint is deleted.
Checkpoint storage policy	Checkpoints are regularly saved (every 15 minutes) and are automatically saved when you exit the application.
Checkpoint storage path	By default, checkpoints are stored in the `/ tmp / logtail_ch eckpoint` directory. For more information about how to modify the directory, see #unique_128.

Table 3-6: Configuration

Item	Capability and limit
Configuration update	Your updated configuration takes effect with a latency of about 30 seconds.
Dynamic configuration loading	Supported. The update of a Logtail configuration does not affect other Logtail configurations.
Number of configurations	Theoretically unlimited. We recommend that you create a maximum of 100 collection configurations on a server.
Multi-tenant data segregation	Collection configurations for different tenants are isolated.

Table 3-7: Resources and performance

Item	Capability and limit
Log processing throughput	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 #unique_128 .

Item	Capability and limit
Maximum performance	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.
Number of monitored directories	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).
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 #unique_128. 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 #unique_128.
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

Item	Capability and limit
Network error handling	If the network is disconnected, Logtail retries and automatically adjusts the retry interval.

Item	Capability and limit
Resource quota out-of-limit processing	If the data transmission rate exceeds the quota of the Logstore, Logtail blocks log collection and automatically retries.
Maximum retry period before timeout	If data transmission fails for more than 6 successive hours, Logtail discards the data.
Status self-check	Logtail automatically restarts if an exception occurs, for example, an application unexpectedly exits or resource usage exceeds the quota.

Table 3-9: Others

Item	Capability and limit
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.

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.

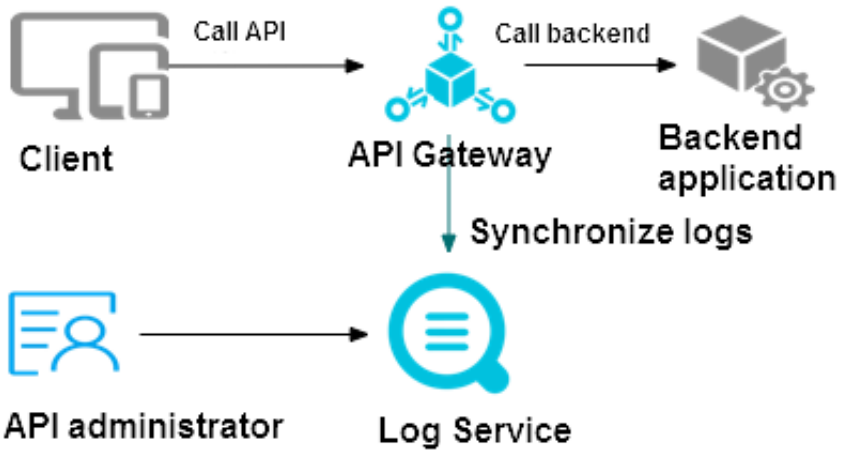
Type	Cloud service	Activation method	Remarks
Elastic computing	ECS	Install Logtail.	Logtail introduction
	Container Service/ Container Service for Kubernetes	Activate the service in the Container Service or Container Service for Kubernetes console.	Text logs and output
Storage	OSS	Activate the service in the OSS console.	#unique_131
Network	SLB	Activate the service in the SLB console.	Access logs of Layer-7 SLB
	VPC	Activate the service in the VPC console.	#unique_133
	API Gateway	Activate the service in the API Gateway console.	Access logs of API Gateway
Security	ActionTrail	Activate the service in the ActionTrail console.	#unique_135
	Anti-DDoS Pro /BGP-line Anti-DDoS Pro	Activate the service in the Anti-DDoS Pro console.	Anti-DDoS Pro overview and BGP-line Anti-DDoS Pro overview

Type	Cloud service	Activation method	Remarks
	Anti-Bot Service	Activate the service in the Anti-Bot Service console.	Anti-Bot Service logs
Application	Log Service	Activate the service in the Log Service console.	#unique_138

4.2 API Gateway Access Log

Alibaba Cloud API Gateway provides API hosting service to facilitate micro-service aggregation, frontend/backend isolation, and system integration. An access log is a log generated by Web services. Each API request corresponds to an access record, containing caller IP, requested URL, response latency, returned status code, number of bytes for each request and response, and other information. With the preceding information, you can understand the operation status of your Web services.

Figure 4-1: API gateway



With Log Service, you can collect access logs of the API Gateway by using Data Import Wizard.

Data migration

1. Online log query: You can perform a rapid accurate or fuzzy search using any keyword in the log. This feature can be used to locate a problem or count queries.
2. Detailed call logs: You can search for details of API call logs.

3. Customized analysis chart: You can customize any log item into a statistical chart according to the statistical requirement to meet your business needs.
4. Preset analysis report: In the API Gateway, some global statistical charts are predefined, including request volume, success rate, failure rate, latency, the number of applications that call APIs, failure statistics, Top grouping, Top API, and Top latency.

Field Description

Log Fields	Description
apiGroupUid	The API group ID.
apiGroupName	API group name
apiUid	The API ID.
apiName	The API name.
apiStageUid	The API stage ID.
apiStageName	The API stage name.
httpMethod	The called HTTP method.
path	The requested path.
domain	The called domain name.
statusCode	The HTTP status code.
ErrorMessage	Error message
appId	The application ID of the caller.
appName	The application name of the caller.
clientIp	The client IP of the caller.
Exception	The specific error message returned by backend.
providerAliUid	The account ID of the API provider.
region	such as cn-hangzhou
requestHandleTime	The request time (GMT).
RequestId	The request ID, which is globally unique.
requestSize	The size of the returned data (in bytes).
Responsesize	The size of the returned data (in bytes).
Servicelatency	The backend latency (in milliseconds).

Procedure

1. Create a project and a Logstore.

   For how to create a project and a Logstore, see Preparation.

   Skip this step if a Logstore already exists.
2. Enter the data access wizard.

   After creating a Logstore, click the Data Import Wizard icon on theLogstore List page .
3. Select a data source.

   Click API Gateway in Cloud Services, and then click Next to go to the Configure Data Source step.
4. Configure the data source.

   In the Configure Data Source step, check whether you have completed the following configurations:

   a. Activate the API Gateway service.

      API Gateway provides a complete API hosting service, helping you open capabilities, services, and data to your partners in the form of API.

      If you have not activated the API Gateway service, activate it as instructed on the relevant page.

   b. Complete Resource Access Management (RAM) authorization.

      Authorize Log Service by using RAM before establishing a dispatch rule, so that Log Service can collect your API Gateway logs.

      Click Authorize in the upper-right corner for quick authorization.

   c. Establish a dispatch rule.

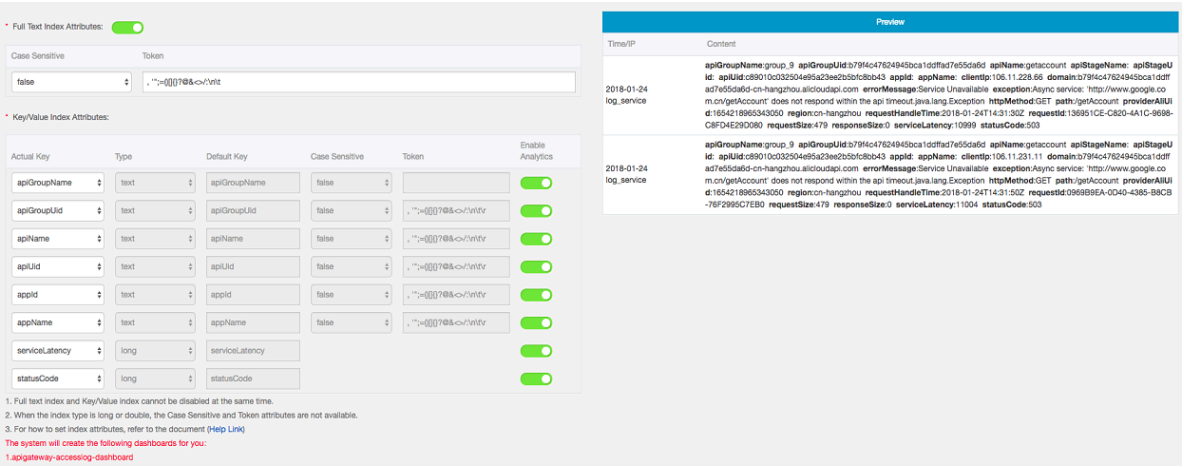
      If you do this for the first time, the system automatically imports API Gateway logs and establishes a dispatch rule. If you have configured API Gateway log collection before, a message indicating the log dispatch rule already exists is displayed. You can select to delete the existing dispatch rule.

   Click Next to enter the Search, Analysis, and Visualization page.

5. Configure Search, Analysis, and Visualization.

   Configure indexes as shown in the following figure. The configuration of the indexes is related to your log search and analysis efficiency.You will also use this configuration in Dashboard, so proceed with caution when modifying this configuration.

   Figure 4-2: Configure Indexes



   Click Next to complete the configuration. Log shipper can be configured separately when necessary.

You have finished the data import wizard initialization. You can select the configured Logstore api-gateway-access-log to query and analyze logs, or go to Dashboard to view reports.

4.3 OSS access logs

4.3.1 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.

4.3.2 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 you 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 4-3: Associate one or more buckets with your Logstore

   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:

   - 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:

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 4-4: 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.

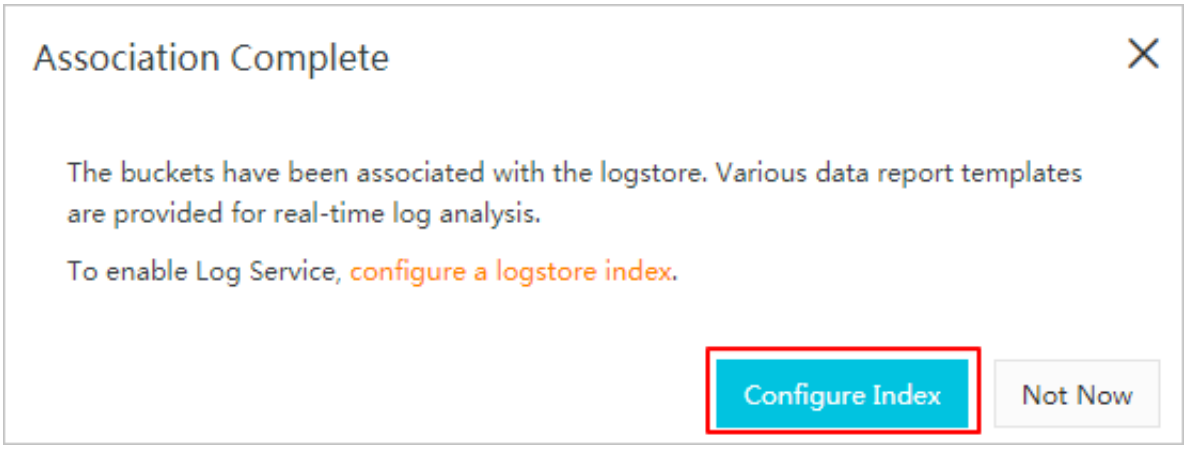
4.3.3 Query OSS access logs

Log Service can query and analyze collected Object Storage Service (OSS) access logs in real time, and clearly display analysis results by using multiple visual charts.

Configure query and analysis

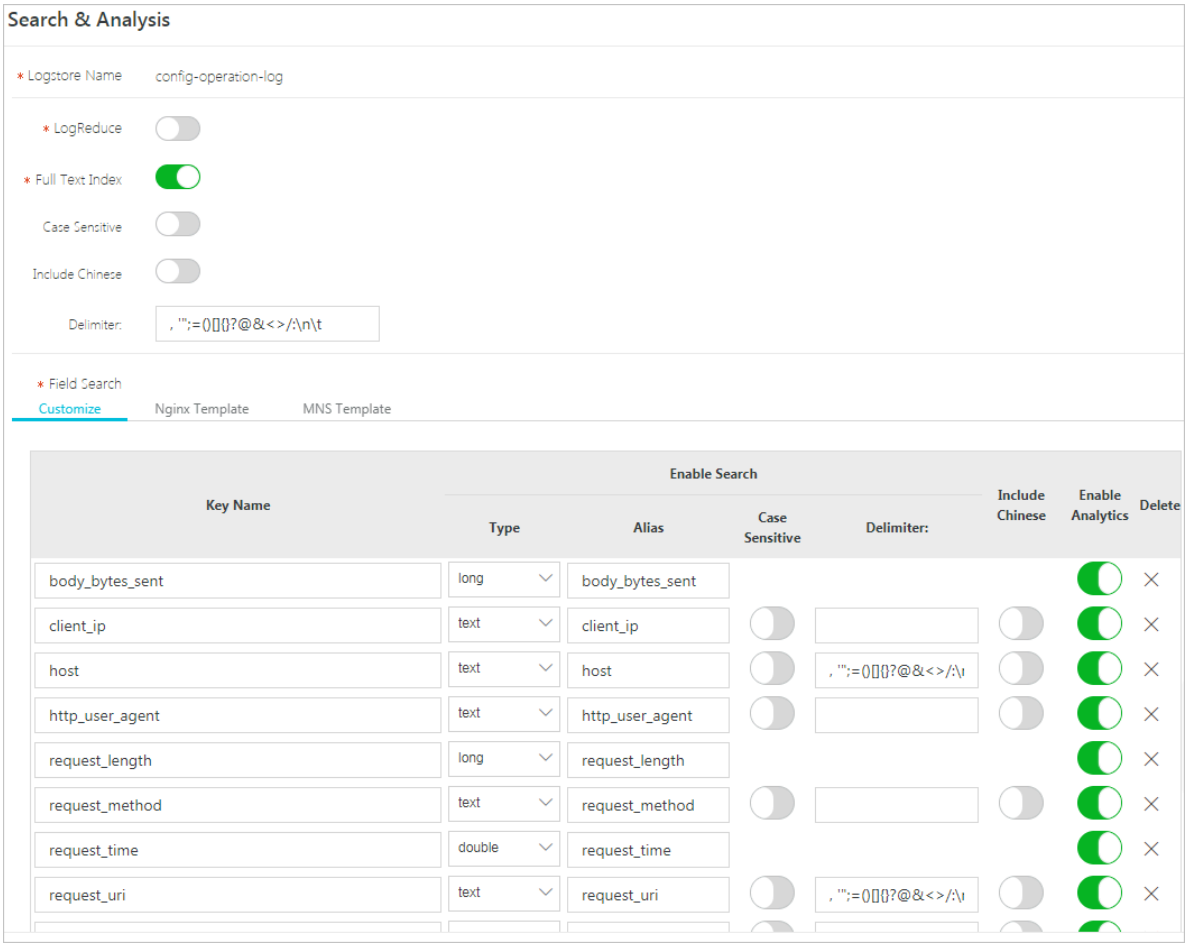
1. After you have created the association, click Configure Index in the displayed dialog box to go to the Log Service console.

Figure 4-5: Configure query and analysis



2. Log Service has indexes preconfigured for querying OSS logs. For more information about field description, see #unique_145. Confirm the configurations and click Next.

Figure 4-6: Configure the indexes

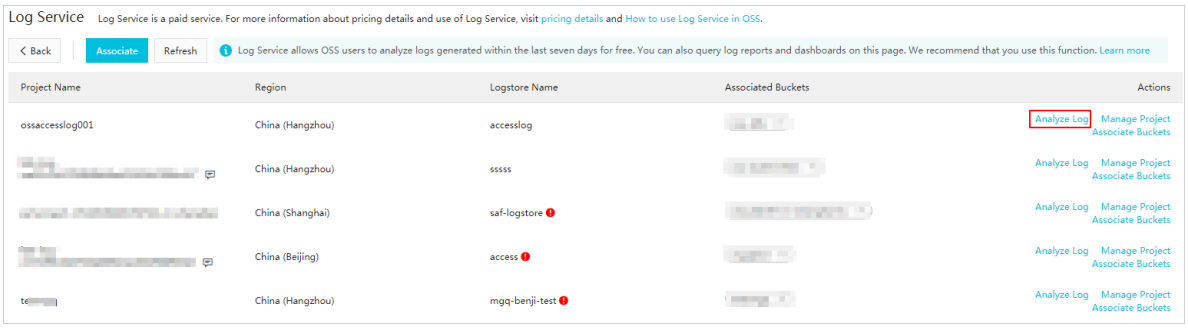


Note:

By default, Log Service creates four specific dashboards for the Logstore associated with one or more buckets. After you complete the configurations, 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 the dashboard name in the left-side navigation pane to view the dashboard.

Figure 4-7: Analyze Logs



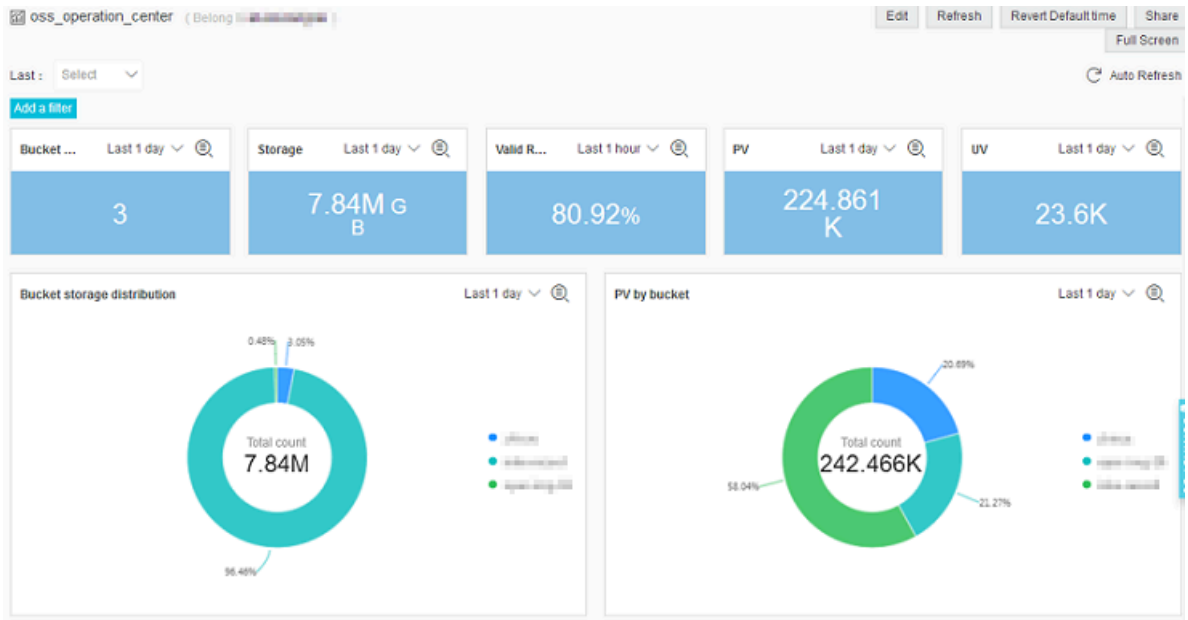
3. Configure the Log Shipper and extract, transform, and load (ETL) as needed, or directly click Confirm.

Default dashboards

Log Service provides the following default dashboards:

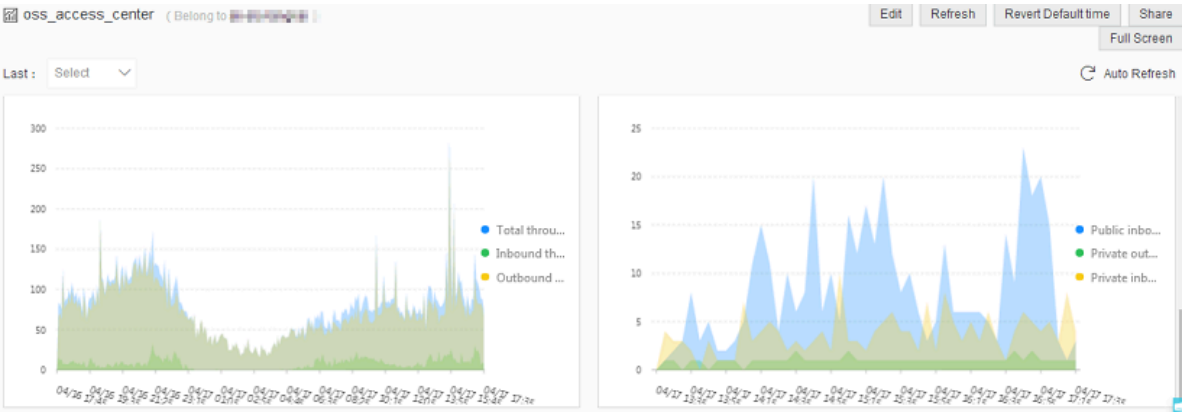
- oss_operation_center: displays overall operation status.

Figure 4-8: Analyze Logs



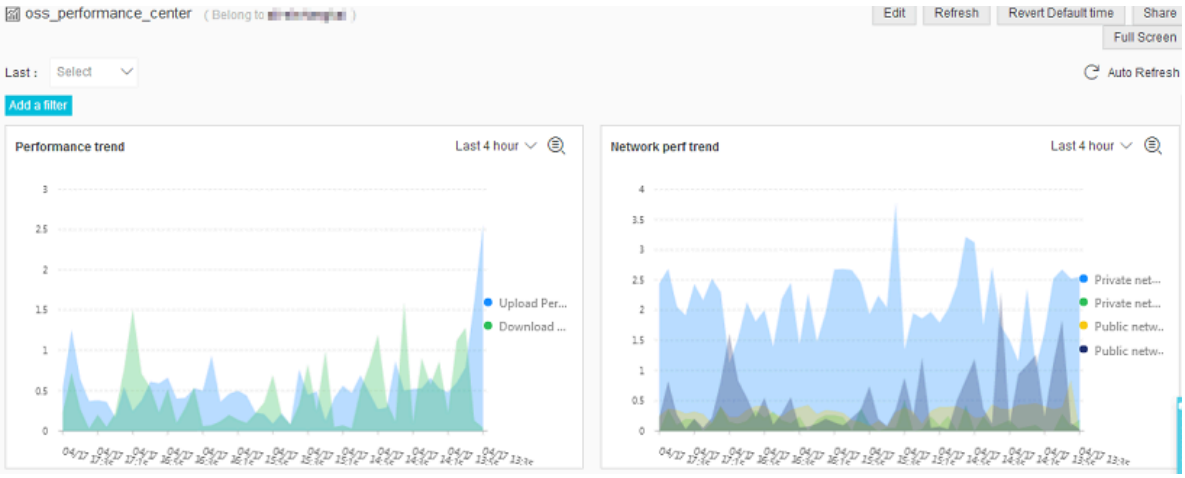
- oss_access_center: displays statistics of access logs.

  Figure 4-9: oss_access_center



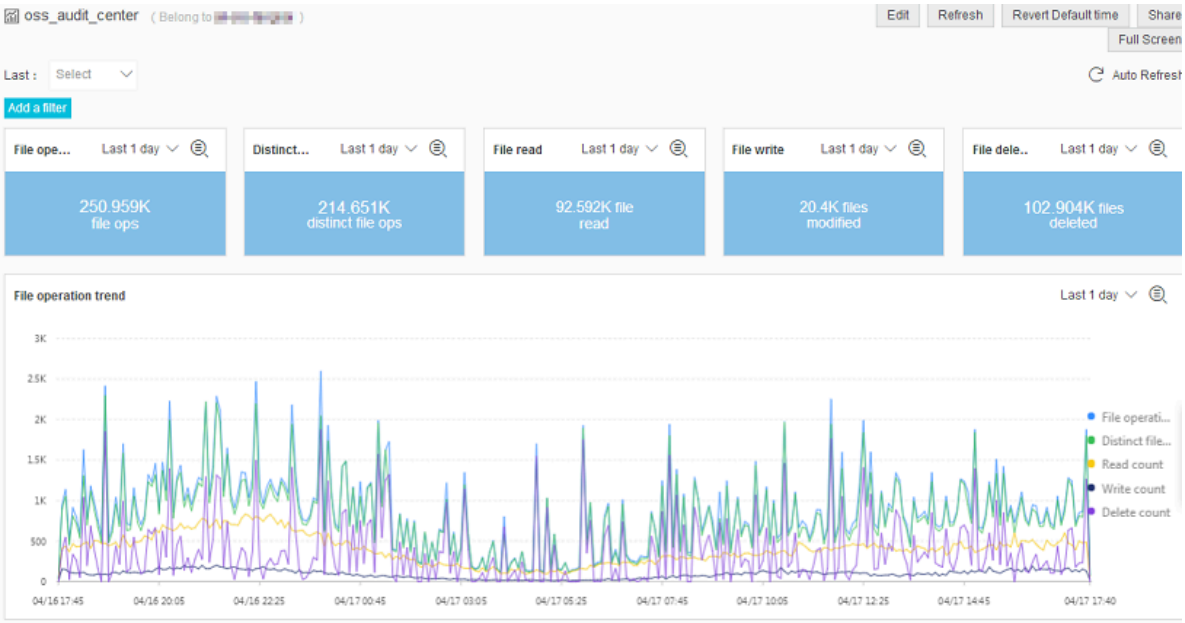
- oss_performance_center: displays statistics of performance.

  Figure 4-10: oss_performance_center



- oss_audit_center: displays statistics of file deletion and modification.

  Figure 4-11: oss_audit_center



4.3.4 Log fields

This topic describes all log fields for Object Storage Service (OSS) access logging.

Types of OSS logs

Table 4-1: Types of OSS logs

Log type	Description
Access log	This log records all access data of the corresponding buckets. Log Service collects the log data in real time.

Log type	Description
Batch deletion log	This log records the objects deleted in batch deletion operations. Log Service collects the log data in real time. Note: When you call the DeleteObjects API operation, the access log generates a request record. The information about the files 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 files, check the corresponding batch deletion log. You can set the `request_id` parameter to associate the batch deletion request with the files that you want to delete.
Hourly metering log	This log records specific 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 OSS access logging to record, collect, store, and analyze logs of OSS access operations, batch deletion operations, and hourly metering operations. OSS built-in logging is a built-in feature of OSS to record and store logs of access operations, and record information about access to OSS storage.

OSS access logs contain all information about OSS access operations. However, these logs have different log fields from OSS built-in logs. The following table describes the differences of their log fields.

OSS built-in log field	Log Service-OSS log field	Description
Remote IP	client_ip	The IP address that you used to make a request. The proxy or your firewall may block this field.
Time	time	The time when OSS received a request.
Request-URI	request-uri	The URI that you requested, including the query-string parameter.

OSS built-in log field	Log Service-OSS log field	Description
HTTP Status	http_status	The HTTP status code that OSS returned.
SentBytes	response_body_length	The traffic consumed when you downloaded logs from OSS.
RequestTime (ms)	response_time	The time consumed to complete a request, in milliseconds.
Referer	referer	The HTTP referer in a request.
User-Agent	User-Agent	The user-agent header in an HTTP request.
HostName	host	The domain name that you requested.
Request ID	request_id	The unique ID used to identify a request.
LoggingFlag	logging_flag	Indicates whether the access logging feature has been enabled.
Requester Aliyun ID	requester_id	Your Alibaba Cloud ID. This field is displayed as a hyphen (-) for anonymous access.
Operation	operation	The type of a request.
Bucket	bucket	The name of the bucket that you requested.
Key	object	The key that you requested. Note: The object field of Log Service is URL encoded.
ObjectSize	object_size	The size of an object.
Server Cost Time (ms)	server_cost_time	The time consumed by the OSS instance to process a request, in milliseconds.

OSS built-in log field	Log Service-OSS log field	Description
Error Code	error_code	The error code that OSS returned.
Request Length	request_length	The length of your request , in bytes.
UserID	owner_id	The ID of a bucket owner.
Delta DataSize	delta_data_size	The variation of the size of a bucket. This field is displayed as a hyphen (-) if the bucket size does not change.
Sync Request	sync_request	Indicates whether you made a back-to-origin request from Content Delivery Network (CDN). This field is displayed as a hyphen (-) if this is not a back-to-origin request.

Access log

Table 4-2: Access log

Field name	Description	Example
access_id	The AccessKey ID of your Alibaba Cloud account.	mEEJX**********
__topic__	The name of the topic in a log. This field is displayed as `oss_access _log` .	-
time	The time when you accessed OSS. This is also the time when OSS received a request. Use the value of `__time__` if a timestamp is required.	27/Feb/2018:13:58:45
owner_id	The Alibaba Cloud ID of the bucket owner.	12345678
User-Agent	The user-agent header in an HTTP request.	curl/7.15.5

Field name	Description	Example
logging_flag	Indicates whether logging has been enabled to periodically export logs to OSS buckets.	true
bucket	The name of a bucket.	bucket123
content_length_in	The value of Content-Length in a request header, in bytes.	12345
content_length_out	The value of Content-Length in a response header, in bytes.	12345
object	The URL encoded object of your request. You can use `select url_decode ( object )` to decode the object when querying logs.	data%2Fcur_file.txt
object_size	The size of a request object, in bytes.	1234
operation	The type of an access operation. For more information about access types and descriptions, see Access types.	GetObject
bucket_location	The cluster where a bucket is located. This field is displayed in the format of `oss -< region >- id` .	oss-cn-beijing-f
request_uri	The URL encoded URI of a request, including the query-string parameter. You can use `select url_decode ( request_ur i )` to decode the URI when querying logs.	/1518085703067732%2Fcur_file.txt HTTP/1.1

Field name	Description	Example
error_code	The error code that OSS returned. For more information about error codes and descriptions, see #unique_149.	NoSuchKey
request_length	The size of an HTTP request, including the header, in bytes.	376
client_ip	The IP address that you used to make a request.	1.2.3.4
response_body_length	The size of the body in an HTTP response, excluding the header.	123
http_method	The method of an HTTP request.	GET
referer	The HTTP referer in a request.	http://www.abc.com
requester_id	Your Alibaba Cloud ID. This field is displayed as a hyphen (-) for anonymous access.	12345678
request_id	The request ID that is used in OSS technical support to troubleshoot issues.	5A7C39674857FB9FFFFFF
response_time	The response time of a request, in milliseconds.	123
server_cost_time	The processing time of an OSS instance, in milliseconds. This is the time consumed by the OSS instance to process a request.	123
http_type	The type of an HTTP request. This field is displayed as HTTP or HTTPS.	http

Field name	Description	Example
sign_type	The type of a signature. For more information about signature types and descriptions, see Signature types.	NormalSign
http_status	The status code of an HTTP connection returned in a request to OSS.	200
sync_request	The type of a synchronization request. For more information about synchronization request types and descriptions, see Synchronization request types.	cn
bucket_storage_type	The type of bucket storage. For more information about bucket storage types and descriptions, see Bucket storage types.	standard
host	The domain name that you requested.	bucket123.oss-cn-beijing.aliyuncs.com
vpc_addr	The virtual IP address (VIP) that corresponds to the domain name that you requested. This field is in the integer format, and used in OSS technical support to troubleshoot issues.	1234567890
vpc_id	The Virtual Private Cloud (VPC) ID that you used when accessing OSS in a VPC. This field is used in OSS technical support to troubleshoot issues.	1234

Field name	Description	Example
delta_data_size	The variation of the size of an object. This field is displayed as 0 if the object size does not change. This field is displayed as a hyphen (-) for requests other than uploads.	280

Batch deletion log

When you call the DeleteObjects API operation, the access log generates a request record. The information about the files 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 files, check the corresponding batch deletion log. The fields and descriptions about deleting multiple logs are shown in the following table. You can set the `request_id` parameter to associate the batch deletion request with the files that you want to delete.

Table 4-3: Batch deletion log

Field	Description	Example
__topic__	The name of the topic in a log. This field is displayed as `oss_batch_delete_log`.	-
client_ip	The IP address that you used to make a request.	1.2.3.4
user_agent	The user-agent header in an HTTP request.	curl/7.15.5
bucket	The name of a bucket.	bucket123
error_code	The error code that OSS returned. For more information about error codes and descriptions, see #unique_149.	NoSuchKey

Field	Description	Example
request_length	The size of an HTTP request, including the header, in bytes.	376
response_body_length	The size of the body in an HTTP response, excluding the header.	123
object	The URL encoded object of your request. You can use `select url_decode ( object )` to decode the object when querying logs.	data%2Fcur_file.txt
object_size	The size of a request object , in bytes.	1234
operation	The type of an access operation. For more information about access types and descriptions, see Access types.	GetObject
bucket_location	The cluster where a bucket is located. This field is displayed in the format of `oss -< region >- id` .	oss-cn-beijing-f
http_method	The method of an HTTP request.	POST
referer	The HTTP referer in a request.	http://www.abc.com
request_id	The request ID that is used in OSS technical support to troubleshoot issues.	5A7C39674857FB9FFFFFF
http_status	The status code of an HTTP connection returned in a request to OSS.	200

Field	Description	Example
sync_request	The type of a synchronization request. For more information about synchronization request types and descriptions, see Synchronization request types.	cdn
request_uri	The URL encoded URI of a request, including the query-string parameter. You can use `select url_decode ( request_ur i )` to decode the URI when querying logs.	/1518085703067732%2Fcur_file.txt HTTP/1.1
host	The domain name that you requested.	bucket123.oss-cn-beijing.aliyuncs.com
logging_flag	Indicates whether logging has been enabled to periodically export logs to OSS buckets.	true
server_cost_time	The processing time of an OSS instance, in milliseconds. This is the time consumed by the OSS instance to process a request.	123
owner_id	The Alibaba Cloud ID of the bucket owner.	12345678
requester_id	Your Alibaba Cloud ID. This field is displayed as a hyphen (-) for anonymous access.	12345678

Field	Description	Example
delta_data_size	The variation of the size of an object. This field is displayed as 0 if the object size does not change. This field is displayed as a hyphen (-) for requests other than uploads.	280

Hourly metering log

This log records specific 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

Field	Description	Example
__topic__	The name of the topic in a log. This field is displayed as `oss_meteri ng_log`.	-
owner_id	The Alibaba Cloud ID of the bucket owner.	12345678
bucket	The name of a bucket.	bucket123
cdn_in	The CDN inbound traffic, in bytes.	123
cdn_out	The CDN outbound traffic, in bytes.	123
get_request	The number of GET requests.	123
intranet_in	The intranet inbound traffic, in bytes.	123
intranet_out	The intranet outbound traffic, in bytes.	123
network_in	The Internet inbound traffic, in bytes.	123
network_out	The Internet outbound traffic, in bytes.	123
put_request	The number of PUT requests.	123

Field	Description	Example
storage_type	The type of bucket storage. For more information about bucket storage types and descriptions, see Bucket storage types.	standard
storage	The storage capacity of a bucket, in bytes.	123
metering_datasize	The size of metering data in non-standard storage.	123
process_img_size	The size of a processed image, in bytes.	123
process_img	The name of a processed image.	123
sync_in	The synchronous inbound traffic, in bytes.	123
sync_out	The synchronous outbound traffic, in bytes.	123
start_time	The timestamp when a metering operation started .	1518084000
end_time	The timestamp when a metering operation ended . The metering data was hourly collected.	1518087600
region	The region where a bucket is located.	cn-beijing
bucket_location	The cluster where a bucket is located. This field is displayed in the format of `oss -< region >- id` .	oss-cn-beijing-f

Operation types

Table 4-5: Operation types

Operation name	Description
AbortMultiPartUpload	Stops a multipart upload.

Operation name	Description
AppendObject	Appends an object.
CommitTransition	CommitTransition.
CompleteUploadPart	Completes a multipart upload.
CopyObject	Copies an object.
DeleteBucket	Deletes a bucket.
DeleteLiveChannel	Deletes a LiveChannel.
DeleteObject	Deletes an object.
DeleteObjects	Deletes multiple objects.
ExpireObject	Makes an object expire.
GetBucket	Queries objects.
GetBucketAcl	Obtains permissions of a bucket.
GetBucketCors	Queries the cross-origin resource sharing (CORS) rules of a bucket.
GetBucketEventNotification	Queries the notification configurations of a bucket.
GetBucketInfo	Queries the information about a bucket.
GetBucketLifecycle	Queries the lifecycle configurations of a bucket.
GetBucketLocation	Queries the region where a bucket is located.
GetBucketLog	Queries the access log configurations of a bucket.
GetBucketReferer	Queries the hotlink protection configurations of a bucket.
GetBucketReplication	Queries the cross-region replication configurations.
GetBucketReplicationProgress	Queries the progress of a cross-region replication.
GetBucketStat	Queries the information about a bucket.
GetBucketWebSite	Queries the static website hosting status of a bucket.
GetLiveChannelStat	Queries the status of a LiveChannel.
GetObject	Reads an object.

Operation name	Description
GetObjectAcl	Obtains the Access Control List (ACL) of an object.
GetObjectInfo	Queries the information about an object.
GetObjectMeta	Queries meta data of an object.
GetObjectSymlink	Queries the details of the symlink file.
GetPartData	Queries the data in all parts of an object.
GetPartInfo	Queries the information about all parts of an object.
GetProcessConfiguration	Queries the image processing configurations of a bucket.
GetService	Queries buckets.
HeadBucket	Queries the information about a bucket.
HeadObject	Queries the information about an object.
InitiateMultipartUpload	Initializes the file for multipart upload.
ListMultiPartUploads	Queries multipart upload events.
ListParts	Queries statuses of all parts of an object.
Options	Options.
PostObject	Uploads an object by using a form.
PostProcessTask	Commits data processing operations, such as taking snapshots.
PostVodPlaylist	Creates a video-on-demand (VOD) playlist of a LiveChannel.
ProcessImage	Processes an image.
PutBucket	Creates a bucket.
PutBucketCors	Specifies the CORS rule for a bucket.
PutBucketLifecycle	Specifies the lifecycle configuration of a bucket.
PutBucketLog	Specifies the access log for a bucket.
PutBucketWebSite	Specifies static website hosting mode for a bucket.
PutLiveChannel	Creates a LiveChannel.
PutLiveChannelStatus	Specifies the status of a LiveChannel.

Operation name	Description
PutObject	Uploads an object.
PutObjectAcl	Modifies the ACL of an object.
PutObjectSymlink	Creates a symlink file.
RedirectBucket	Redirects the request to a bucket endpoint.
RestoreObject	Restores an object.
UploadPart	Resumes uploading a file from a checkpoint.
UploadPartCopy	Copies a part.
get_image_exif	Queries the exchangeable image file format (Exif) data of an image.
get_image_info	Queries the length and width of an image .
get_image_infoexif	Queries the length, width, and Exif data of an image.
get_style	Queries the style of a bucket.
list_style	Queries all styles of a bucket.
put_style	Creates a style of a bucket.

For more information about each operation, see #unique_153.

Synchronization request types

Table 4-6: Synchronization request types

Synchronization request type	Description
-	Indicates a common request.
cdn	Indicates a back-to-origin request from CDN.

Signature types

Table 4-7: Signature types

Signature type	Description
NotSign	Indicates that a request was not signed.

Signature type	Description
NormalSign	Indicates that a request was signed with a normal signature.
UriSign	Indicates that a request was signed with a URL signature.
AdminSign	Indicates that a request was signed by an administrator.

For more information about signatures, see #unique_154.

Bucket storage types

Table 4-8: Bucket storage types

Storage type	Description
standard	Standard storage
archive	Archive storage
infrequent_access	Infrequent access storage

For more information about each storage type, see #unique_155.

4.4 Access logs of Layer-7 Server Load Balancer

Alibaba Cloud Server Load Balancer can distribute traffic for multiple Elastic Compute Service (ECS) instances, and support Layer-4 Server Load Balancer based on TCP and Layer-7 Server Load Balancer based on HTTP/HTTPS. By using Server Load Balancer, the impact on the business is reduced when a single ECS instance has an exception so that the system availability is enhanced. Working with the dynamic expansion and contraction of Auto Scaling, backend servers can respond to the changes of business traffic quickly.

Each access request to Server Load Balancer records the access logs. The access logs collect the details of all the requests sent to Server Load Balancer, including request time, client IP address, latency, request path, and server response. As an Internet access point, Server Load Balancer hosts a large number of access requests. By using the access logs, you can analyze the user behavior on the client, the geographical distribution of the client users, and troubleshoot the issues.

Use Log Service to collect the Server Load Balancer access logs. You can monitor, probe, diagnose, and report the Layer-7 access logs of HTTP/HTTPS continuously and understand Server Load Balancer instances more comprehensively.

> Note:
> Only Layer-7 Server Load Balancer supports the access logs function. The access logs function is available in all regions. For more information, see #unique_157.

Function advantages

- Simple. Free developers and maintenance staff from tedious and time-consuming log processing so that they can concentrate on business development and technical research.
- Massive. Access logs are proportional to request PVs of Server Load Balancer instances. The data size is usually large. Therefore, the performance and cost issues must be considered when processing access logs. Log Service can analyze 100 million logs in a second and has obvious cost advantages compared with the open-source solutions.
- Real-time. Scenarios such as DevOps, monitoring, and alerting require real-time log data. Traditional data storage and analysis tools cannot meet this requirement. For example, it takes long time to ETL data to Hive at which a lot of work is spent on data integration. Powered by its powerful computing capability, Log Service can process and analyze access logs in seconds.
- Flexible. You can enable or disable the access log function at the level of Server Load Balancer instance. You can enable or disable the access log function at the level of Server Load Balancer instance. Additionally, you can set the storage period (1–365 days) and the Logstore capacity of logs is dynamically scalable to meet business growth requirements.

Configure Log Service to collect Layer-7 Server Load Balancer access logs

Prerequisites

1. You have activated Server Load Balancer and Log Service. The created #unique_158, Log Service project, and Logstore are in the same region.

   > Note:
   > Only Layer-7 Server Load Balancer supports the function of access logs. For the available regions, see Access logs.

2. If you are a RAM user, you must be authorized to use the SLB access logging. For more information, see #unique_159.

Procedure

1. Log on to the Log Service console.
2. After project and Logstore are created, follow the page prompts to enter the data import wizard. You can also click the Data Import Wizard icon on the Logstore List page to enter the configuration process.
3. Select a data source.

   Click Server Load Balancer in Cloud Services and then click Next.
4. RAM authorization.

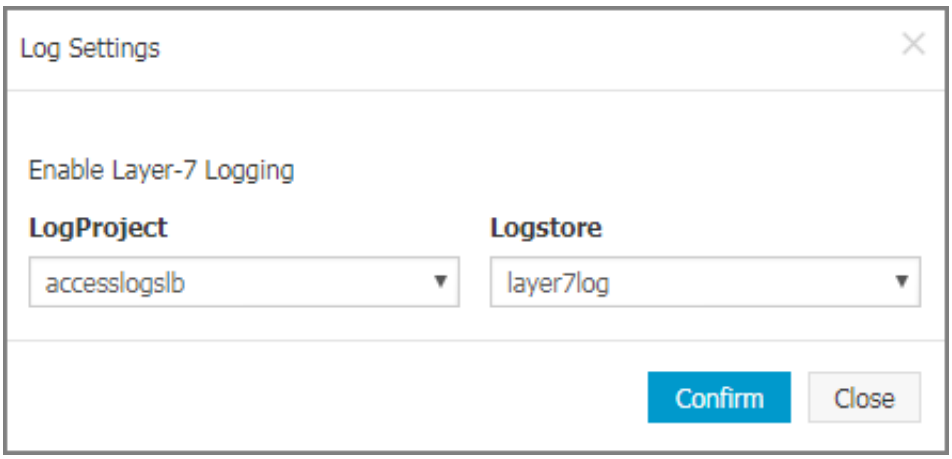
   Click Authorize as instructed on the page. Then, click Confirm Authorization Policy to authorize Server Load Balancer to access Log Service.
5. Set dispatch rule. Click Dispatch configuration to go to the Server Load Balancer console.

   a. ClickLogs > Access Login the left-side navigation pane.
   b. Click Configure at the right of the Server Load Balancer instance.

      Note:

Make sure the Log Service project and the SLB instance are in the same region.

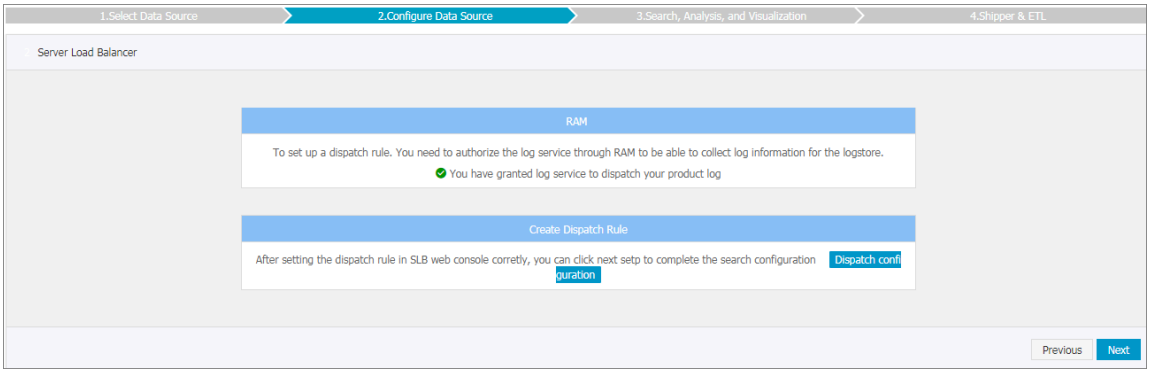
Figure 4-12: Log Settings



c. Select the project and Logstore of Log Service. Then, click Confirm.

d. After the configuration is complete, close the dialog box. Return to the data import wizard and click Next.

Figure 4-13: Configure Data Source



6. Search, analysis, and visualization.

   Log Service presets the query indexes required by Server Load Balancer. For the field descriptions, see Field description in this document. Click Next.

   Note:
   The dashboard {LOGSTORE}-slb_layer7_access_center and {LOGSTORE}-slb_layer7_operation_center are created by default. After the configuration is complete, you can view it on the Dashboard page.

7. Click Confirm to complete the data access.

Subsequent operations

- Query logs in real time

  You can perform a rapid accurate or fuzzy query by using any keyword in the log. This feature can be used for problem location or statistical query.

- Preset analysis reports

  Server Load Balancer predefines some global statistics graphs, including Top client access, distribution of request status codes, Top URI access, traffic trend of request messages, and statistics of RealServer response time.

- Customize analysis charts

  You can perform an ad-hoc query for any log item according to the statistical requirement and save the results as a chart to meet your daily business requirements.

- Set log monitoring alarms

  You can perform customized analysis on Server Load Balancer request logs and save the results as a quick query. Set the quick query as an alarm. When the computing results of real-time logs exceed the defined threshold, the system sends an alarm notification.

Field descriptions

Field	Description
body_bytes_sent	The size of the HTTP body (in bytes) sent to the client.
client_ip	The request client IP.
host	The host is obtained from the request parameters first. If no value is obtained, obtain the host from the host header. If the value is still not obtained, use the backend server IP of the processing request as the host.
http_host	The host header contents in the request message.
http_referer	The HTTP referer header contents in the request message received by the proxy.

Field	Description
http_user_agent	The HTTP user-agent header contents in the request message received by the proxy.
http_x_forwarded_for	The x-forwarded-for contents in the request message received by the proxy.
http_x_real_ip	The real client IP.
read_request_time	The time (in milliseconds) for the proxy to read request.
request_length	The length of the request message, including startline, HTTP header, and HTTP body.
request_method	The method of the request message.
Request_time	The interval (in seconds) between the time when proxy receives the first request message and the time when proxy returns the response.
request_uri	The URI of the request message received by the proxy.
scheme	The request schema (http or https).
server_protocol	The HTTP protocol version received by the proxy. For example, HTTP/1.0 or HTTP/1.1.
slb_vport	The listening port of Server Load Balancer.
slbid	The Server Load Balancer instance ID.
ssl_cipher	The used cipher, such as ECDHE-RSA-AES128-GCM-SHA256/.
ssl_protocol	The protocol used to establish the SSL connection, such as TLSv1.2.
status	The status of proxy responding to the message.
tcpinfo_rtt	The tcp rtt time (in microseconds) on the client.
time	The log recorded time.

Field	Description
Upstream_addr	The IP address and port of the backend server.
upstream_response_time	The time (in seconds) during which Server Load Balancer establishes a connection on the backend, receives the data, and closes the connection.
upstream_status	The response status code of the backend server received by the proxy.
vip_addr	The vip address.
write_response_time	The time (in milliseconds) for the proxy to write responses.

4.5 DDoS log collection

4.5.1 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 appears. Log Service supports Pay-As-You-Go billing method, and provides FreeTier quota. For more information, see #unique_163.

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:

  ```
  __topic__ : ddos_acces s_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:

  ```
  __topic__ : ddos_acces s_log and cc_blocks > 0 | SELECT
    ip_to_coun try ( if ( real_clien t_ip ='-', remote_add r ,
  ```

```
real_clien t_ip )) as country , count ( 1 ) as " number of
  attacks " group by country
```

- For example, view the PV access by the following statement:

```
__topic__ : ddos_acces s_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:

```
__topic__ : ddos_acces s_log | select ip_to_prov ider ( if (
real_clien t_ip ='-', remote_add r , real_clien t_ip )) as
provider , round ( sum ( request_le ngth )/ 1024 . 0 / 1024 . 0
, 3 ) as mb_in group by provider having ip_to_prov
ider ( if ( real_clien t_ip ='-', remote_add r , real_clien
t_ip )) <> '' order by mb_in desc limit 10
```

4.5.2 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.
2. Enable Anti-DDoS Pro function, purchase Anti-DDoS Pro instances.
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

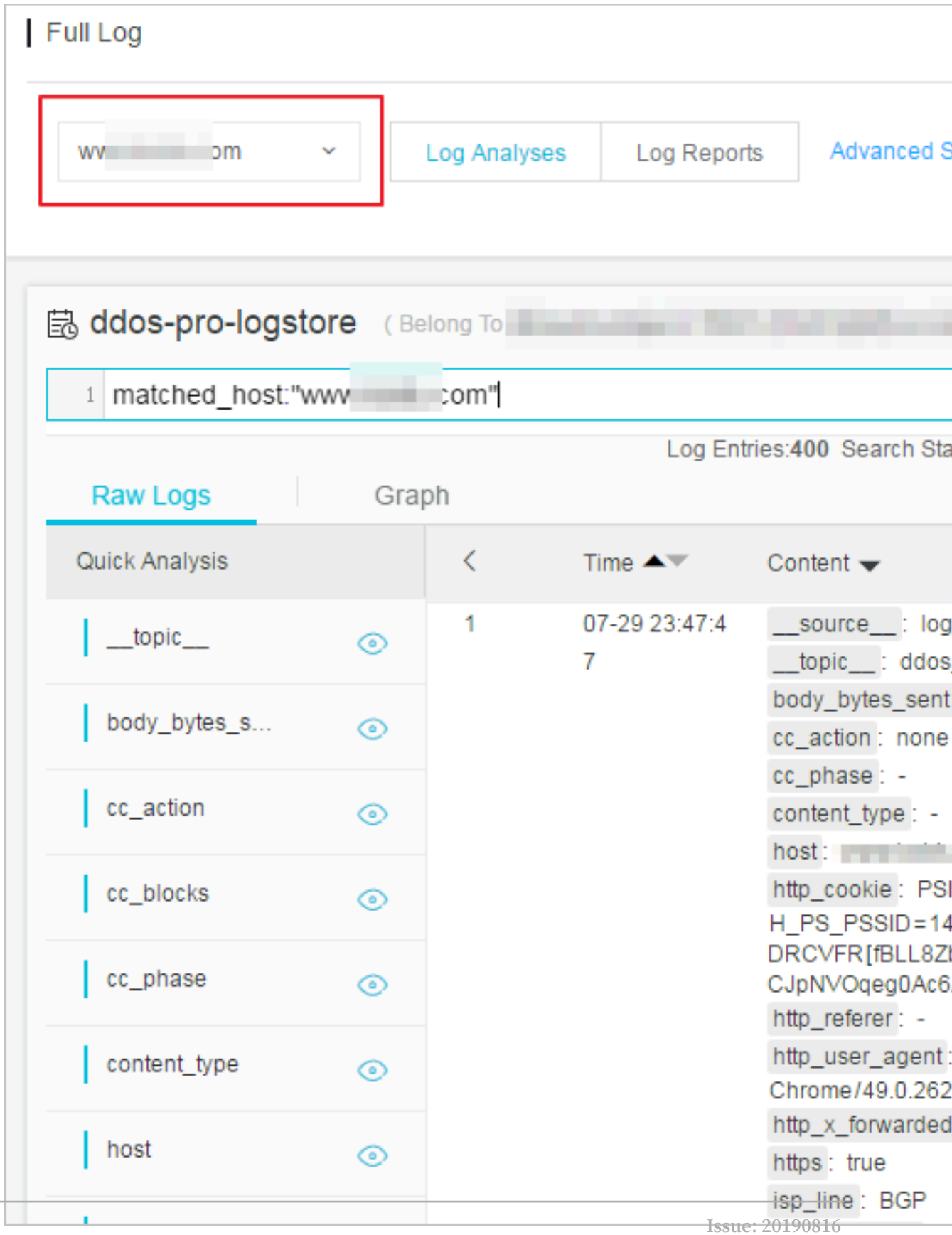
1. Log on to the Anti-DDoS Pro console and select Log > Full Log in the left-side navigation pane. Enter the Full Log page.

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 4-14: 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

Default configuration item	Configuration content
Project	By default, `ddos - pro - logstore` project is created.
Logstore	By default, Logstore is created. Logstore name is determined by the domain of the DDoS you purchased. · DDoS instances in mainland China: `ddos-pro-project-Alibaba Cloud Account ID-cn-hangzhou`. · 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_163.

Default configuration item	Configuration content
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 #unique_167.

You can query and analyze the collected logs in real time on the currentFull 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.

Field	Description	Example
__topic__	The topic of the log is fixed to ddos_access_log.	-
body_bytes_sent	Request to send the size of the Body. The unit is byte.	2
content_type	Content type.	application/x-www-form-urlencoded
host	Source website.	api.zhihu.com
http_cookie	Request cookie.	k1=v1;k2=v2
http_referer	Request referer. If none, the - is displayed.	http://xyz.com
http_user_agent	User agent request.	Dalvik/2.1.0 (Linux; U; Android 7.0; EDI-AL10 Build/HUAWEIEDISON-AL10)
http_x_forwarded_for	The upstream user IP that is redirected by the proxy.	-

Field	Description	Example
https	Whether the request is an HTTPS request, wherein: · true: the request is an HTTPS request. · false: the request is an HTTP request.	true
matched_host	The source website of the matching configuration may be a pan-domain name. If not matching, the - is displayed.	*.zhihu.com
real_client_ip	Access the customer real IP. If not available, the - is displayed.	1.2.3.4
isp_line	Line information, such as BGP, telecommunication , Unicom.	Telecommunication
remote_addr	Request client IP connection.	1.2.3.4
remote_port	Request client port connection.	23713
request_length	The length of the request. The unit is byte.	123
request_method	The HTTP request method .	GET
request_time_msec	Request time. The unit is microsecond.	44
request_uri	Request path.	/answers/377971214/ banner
server_name	The matching host name. If not matching, the `default` is displayed.	api.abc.com
status	HTTP status code.	200
time	Time.	2018-05-02T16:03:59+08: 00

Field	Description	Example
cc_action	CC protection policy, such as none, challenge, pass, close, captcha, wait, logon, n.	close
cc_blocks	Indicates whether CC protection is blocked, wherein: · 1: Blocked. · Other codes: Passed.	1
cc_phase	CC protection policy, including seccookie, server_ip_blacklist, static_whitelist, server_header_blacklist, server_cookie_blacklist, server_args_blacklist, qps_overmax.	server_ip_blacklist
ua_browser	Browser.	ie9
ua_browser_family	Browser series.	Internet explorer
ua_browser_type	Browser type.	web_browser
ua_browser_version	Browser version.	9.0
ua_device_type	Client device type.	computer
ua_os	Client operating system.	windows_7
ua_os_family	Client operating system series.	windows
upstream_addr	Return source address list, the format is `IP : Port`. Multiple addresses are separated by commas.	1.2.3.4:443
upstream_ip	The actual return source address IP.	1.2.3.4
upstream_response_time	The response time of the source. The unit is second.	0.044

Field	Description	Example
upstream_status	Return source request HTTP status.	200
user_id	Alibaba Cloud user ID.	12345678

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.

4.5.3 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.

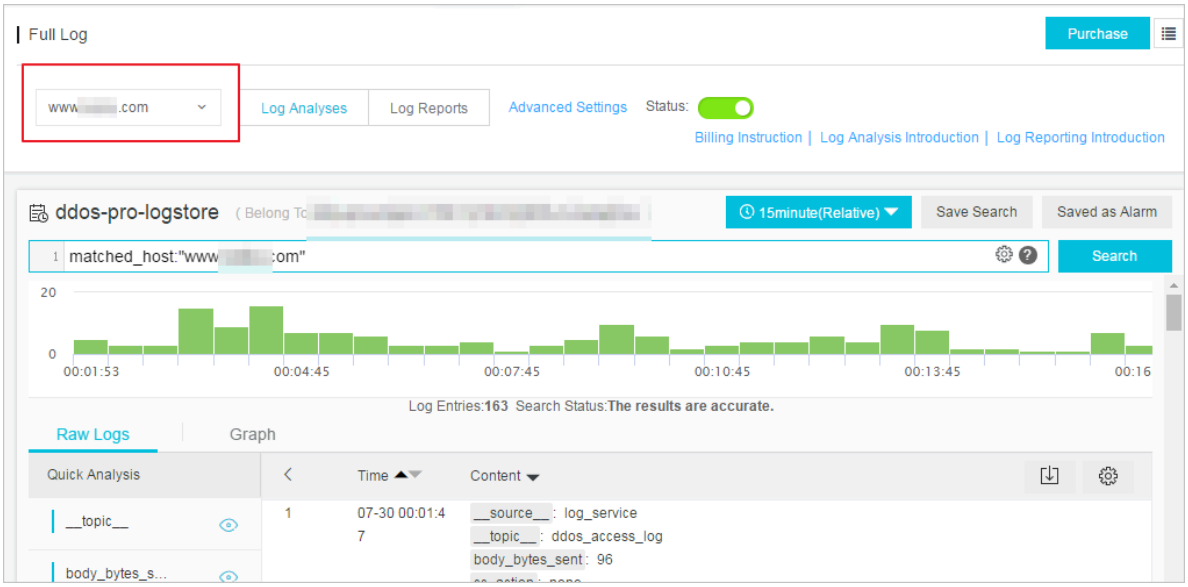
Procedure

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_ho st : www . aliyun . com`, to view the log data based on the selected website.

   Figure 4-15: Log analysis

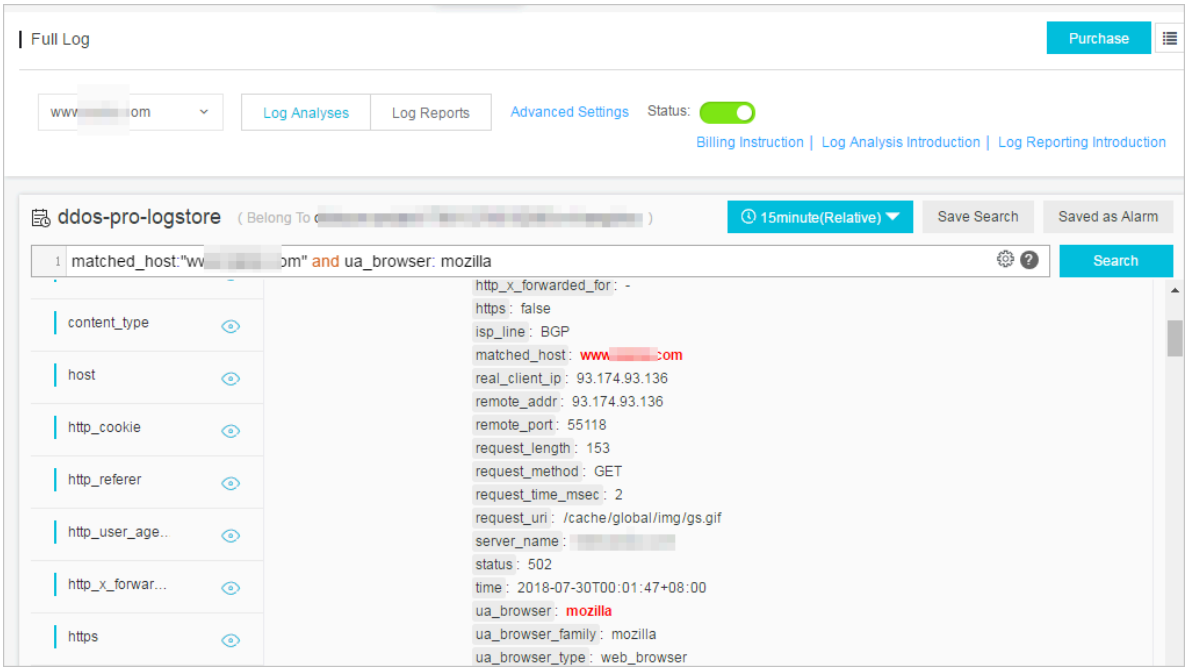


4. Enter query analysis statement, select the log time range and click Query.

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 4-16: 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.

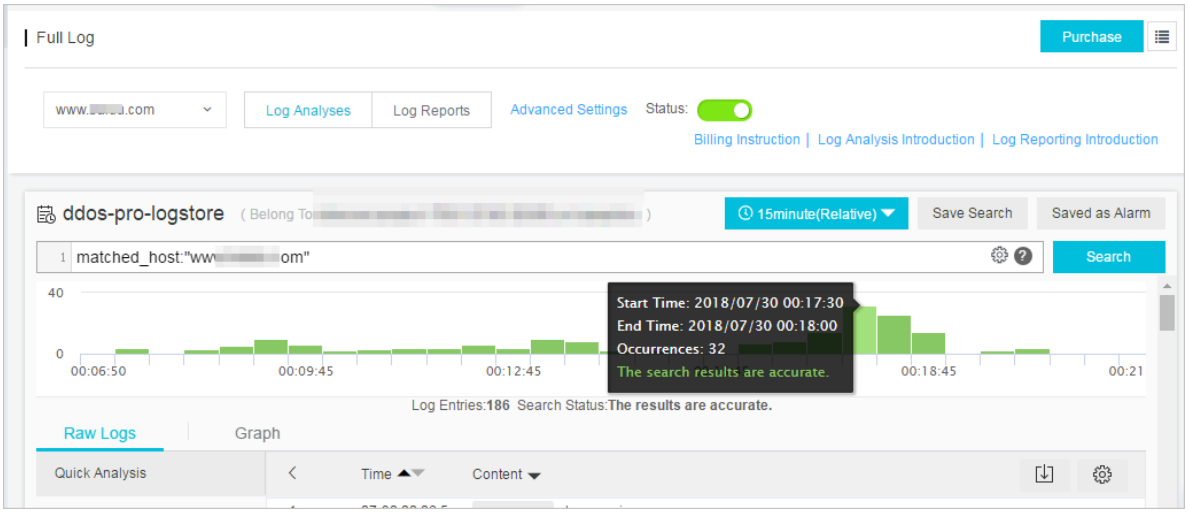
- View the log time distribution

  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 4-17: Log time distribution



- View the raw logs

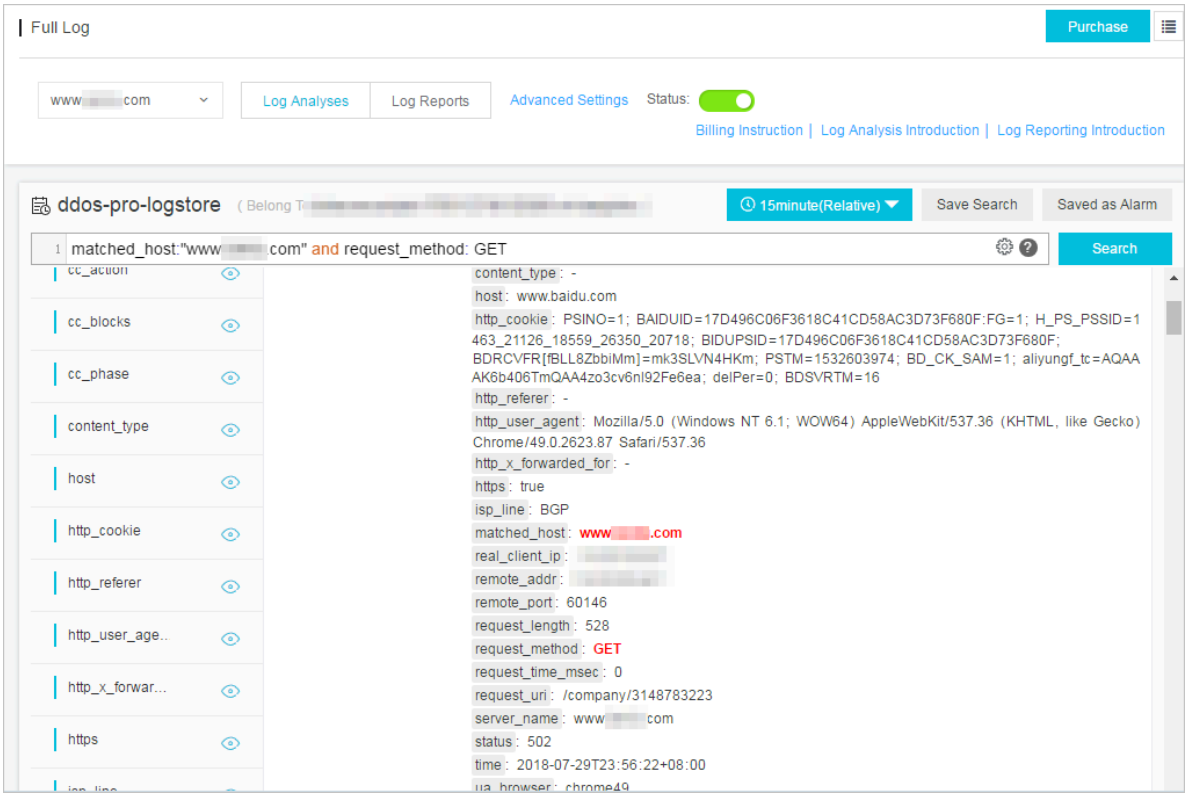
  In the Raw log, the details of each log are displayed in pagination, including time and content of these fields. You can sort the columns, download the current query results, or click the gear to select specific fields to be displayed.

  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_me thod : GET`, the following statement is automatically added to the search box:

```
Raw search statement and request_me thod : GET
```

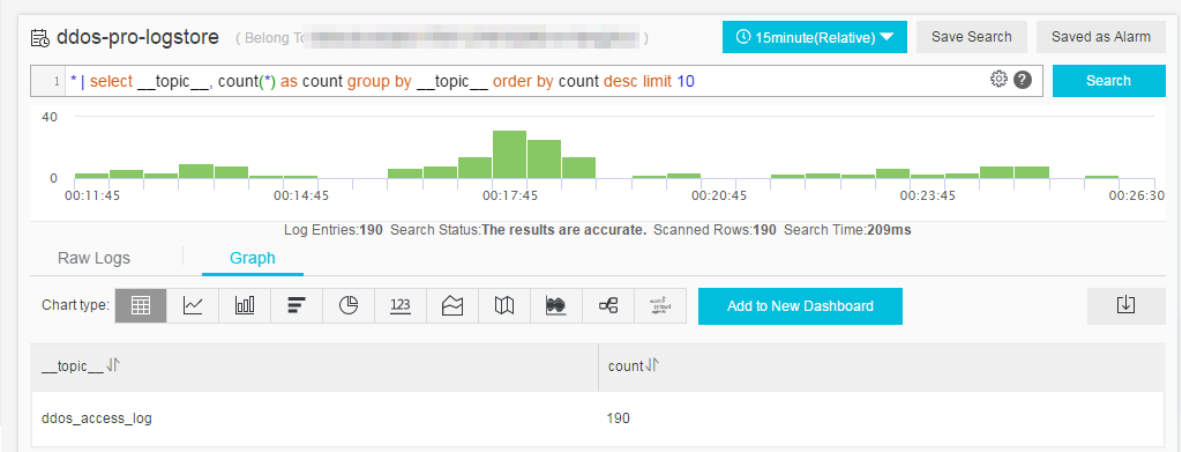
Figure 4-18: Raw logs



- 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.

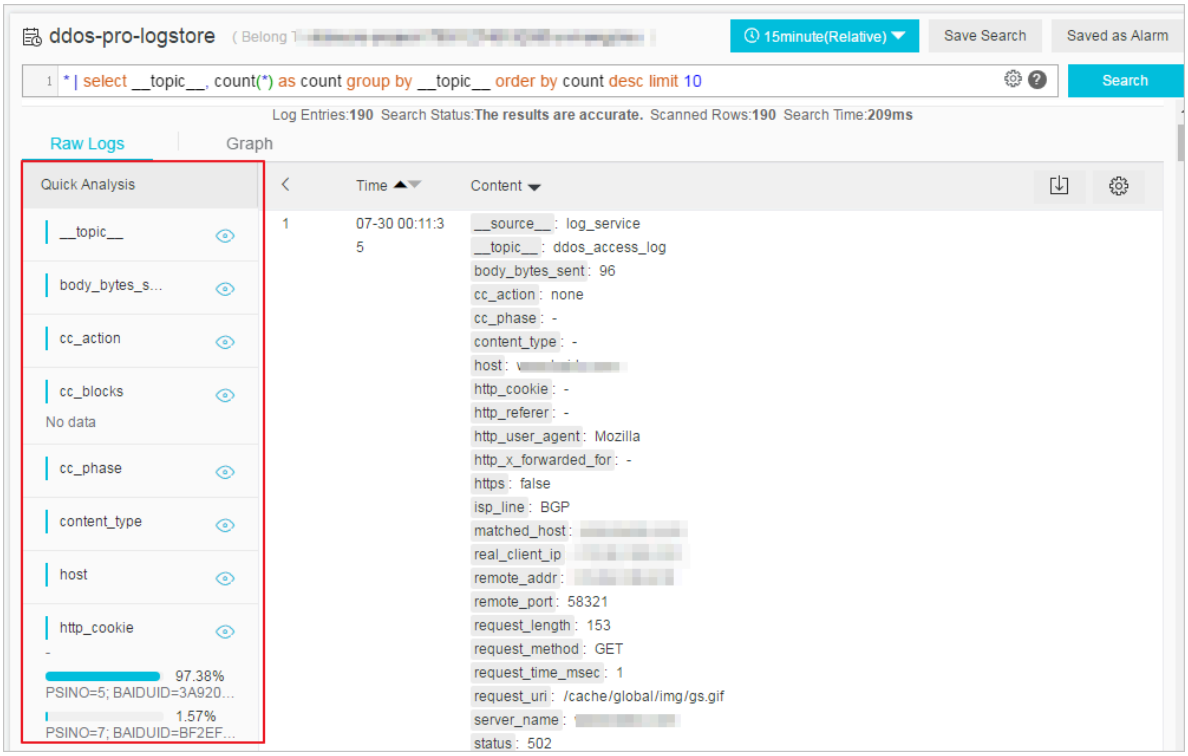
  Figure 4-19: Statistic chart



- 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 4-20: 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
```

Type	Description
Query (Search)	The query conditions can be generated by keywords, fuzzy, numerical values, interval range and combination conditions. If left empty or `*`, all data is displayed.
Analysis (Analytics)	Calculate and count the query results or the full amount of data.

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_ho  st :  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_clien  t_ip :  1 . 2 . 3 . 4  and  matched_ho  st :  www .
    aliyun . com  and  status :  404
    ```

    Note:

Fields used in the examples `matched_ho st` , `cc_blocks` , `real_clien t_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:

  ```
  request_ti me_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:

  ```
  request_ti me_msec in ( 5000 10000 ]
  ```

  The query can also be performed by the following statement:

  ```
  request_ti me_msec > 5000 and request_ti me_msec <=
  10000
  ```

- Check whether Japanese characters are used.

  Query for the presence of specific fields:

  ■ Query logs in the `ua_browser` field: `ua_browser : *`.

  ■ Query logs that do not belong to 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 data analysis and statistics. For more information about the syntax and functions supported by Log Service, see #unique_174.

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_175.

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:

```
matched_ho st : www . aliyun . com and cc_blocks : 1
| select time , real_clien t_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:

```
matched_ho st : www . aliyun . com and cc_blocks : 1
| select time ,
         round (( to_unixtim e ( now ()) - __time__ )/ 86400 ,
1 ) as " days_passe d ", real_clien t_ip , http_user_ agent
      order by time desc
      limit 10
```

> **Note:**
> Use `round (( to_unixtim e ( now ()) - __time__ )/ 86400 , 1 )`, first part `to_unixtim e`, 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:

```
matched_ho st : www . aliyun . com and cc_blocks : 1
| select date_trunc (' day ', __time__ ) as dt ,
        count ( 1 ) as PV
      group by dt
      order by dt
```

  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` , `miniute` , `hour` , `week` , `month` , `year` . For more information about function, see #unique_176.

- 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:

```
matched_ho st : www . aliyun . com and cc_blocks : 1
| select from_unixt ime ( __time__ - __time__ % 300 ) as
dt ,
        count ( 1 ) as PV
      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_unixt ime` . 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_forma t` . For more information, see #unique_176.

Client IP-based query analysis

DDoS log has a field `real_clien t_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_add r` field to directly connected to the client IP.

- Country attack distribution

  Distribution of source countries of CC attacks on a website:

```
matched_ho st : www . aliyun . com and cc_blocks : 1
| SELECT ip_to_coun try ( if ( real_clien t_ip ='-',
remote_add r , real_clien t_ip )) as country ,
        count ( 1 ) as " number of attacks "
        group by country
```

  Note:
  Use the function `if ( condition , option1 , option2 )` to select the field `real_clien t_ip` or `real_clien t_ip` (when `real_clien t_ip` is -). Pass the obtained IP to the function `ip_to_coun try` to get the country information corresponding to this IP.

- Access distribution

  To get more detailed province-based distribution, use the `ip_to_prov ince` function, for example:

```
matched_ho st : www . aliyun . com and cc_blocks : 1
| SELECT ip_to_prov ince ( if ( real_clien t_ip ='-',
remote_add r , real_clien t_ip )) as province ,
        count ( 1 ) as " number of attacks "
        group by province
```

  Note:
  Another IP function `ip_to_prov ince` 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 `ip_to_geo` function, for example:

```
matched_ho st : www . aliyun . com and cc_blocks : 1
| SELECT ip_to_geo ( if ( real_clien t_ip ='-', remote_add r
, real_clien t_ip )) as geo ,
        count ( 1 ) as " number of attacks "
        group by geo
```

```
limit   10000
```

Note:
Use another IP function `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 `ip_to_provider`, determining whether the IP is Internet or Intranet `ip_to_domain`, see #unique_177.

4.5.4 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.

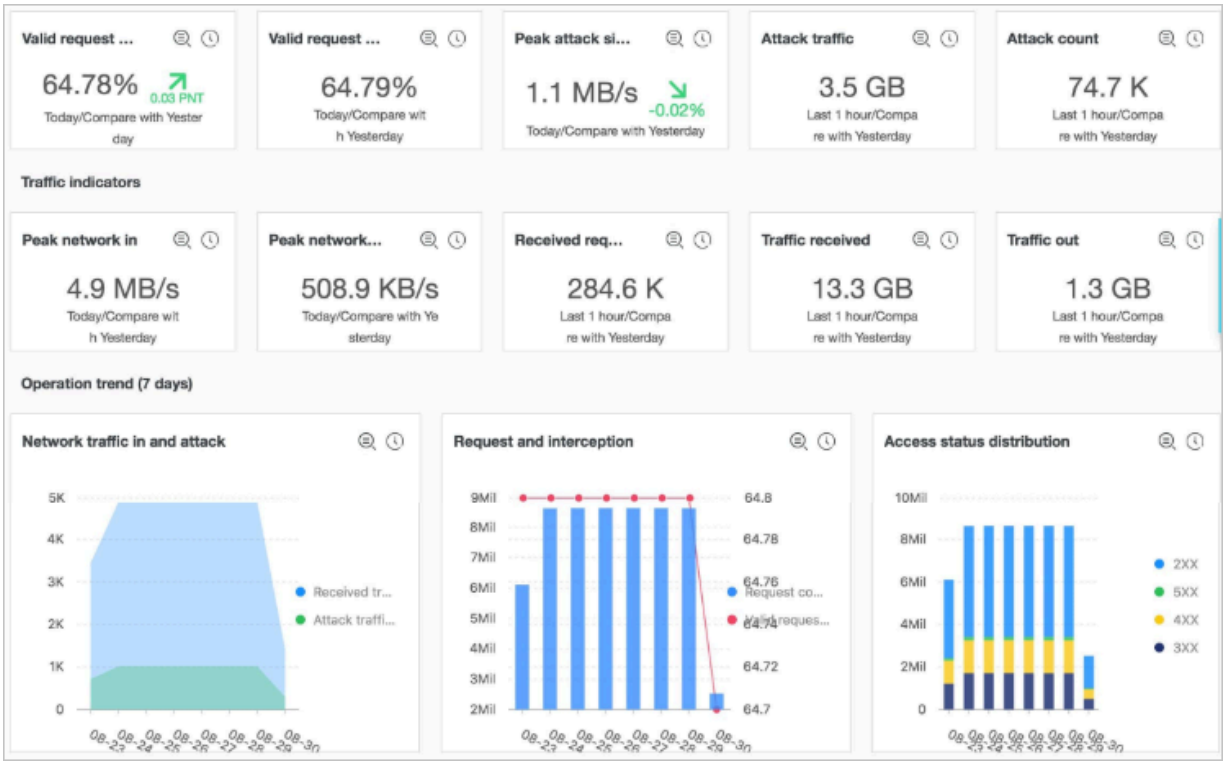
   Dashboard page of Log Service is embedded in the current page, and the filter condition is automatically added. For example, use `matched_host : www.aliyun.com` to view log reports based on selected website.

Figure 4-21: 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.

Dashboard	Dashboard name	Description
ddos-pro-logstore_ddos_operation_center	DDoS operation center	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.
ddos-pro-logstore_ddos_access_center	DDoS 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.

Figure 4-22: Default dashboard



Besides viewing the report, the following operations can be performed:

- Select time range
- Add or edit filter condition

- View charts

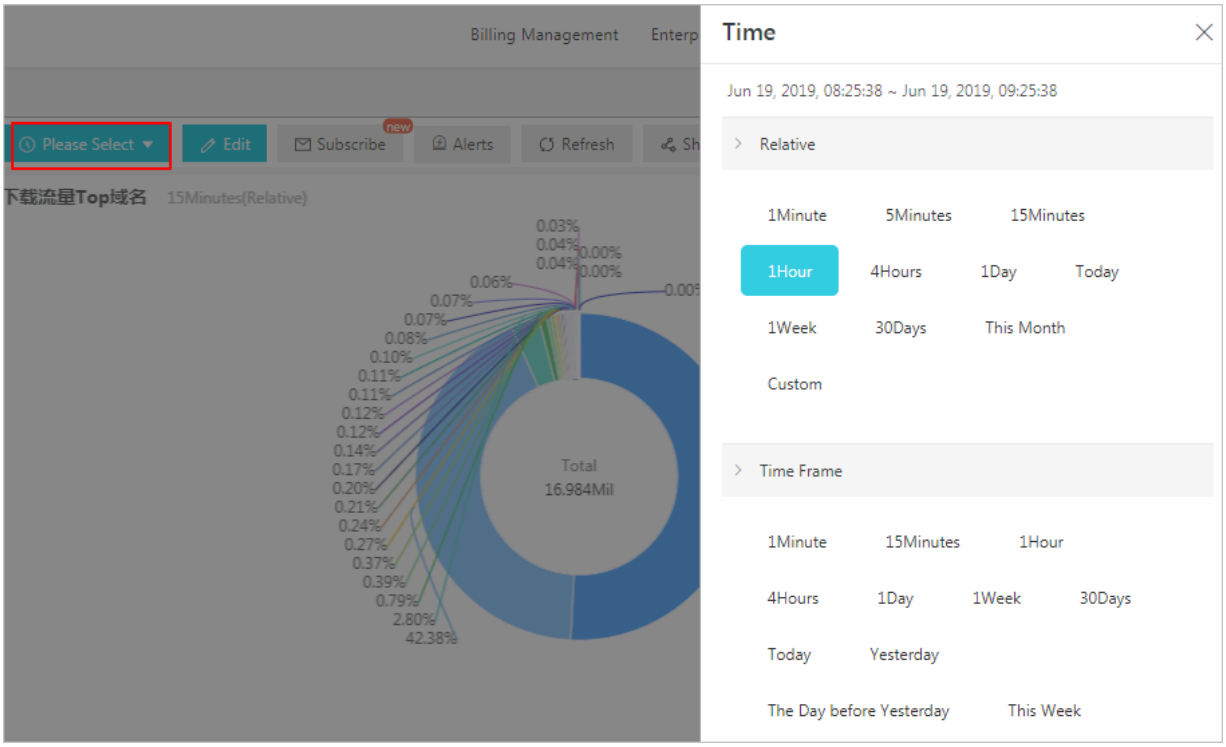
Time picker

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.

Figure 4-23: Set the time range



Filter conditions

Select the website and click Log Reports to enter the dashboard page. System automatically adds filter condition, such as `matched_ho st : 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 4-24: 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.

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_180.

Chart type	Description
Number	Displays important indicators, such as effective request rate, and attack peaks.
Line/area map	Displays trend graphs for certain important indicators within a specific time period, such as inbound bandwidth trends and attack interception rates.

Chart type	Description
Map	Displays the geographical distribution of visitors and attackers, such as CC attack country, access hotspot.
Pie chart	Displays the distribution of the information, such as the top 10 of the websites being attacked, client type distribution.
Table	Displays information such as the list of attackers, typically divided into multiple columns.
Maps	Displays the geographical distribution of the data.

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.

Chart	Type	Default time range	Description	Example
Valid request package rate	Single value	1 hour (relative)	A valid request, that is, the number of non-CC attacks or 400 error requests in the total number of all requests.	95%
Valid request flow rate	Single value	1 hour (relative)	Valid request percentage of the total flow of all requests.	95%
Received traffic	Single value	1 hour (relative)	The sum of valid request inflows. The unit is MB.	300 MB

Chart	Type	Default time range	Description	Example
Attack traffic	Single value	1 hour (relative)	The sum of inbound traffic of CC attacks. The unit is MB.	30 MB
Outbound traffic	Single value	1 hour (relative)	The sum of valid request outbound traffic. The unit is MB.	300 MB
Network in bandwidth peak.	Single value	1 hour (relative)	The highest peak of incoming traffic rate requested by the website. The unit is bytes/s.	100 Bytes/s
Network out bandwidth peak.	Single value	1 hour (relative)	The highest peak of outbound traffic rate requested by the website. The unit is bytes/s.	100 Bytes/s
Received data packets	Single value	1 hour (relative)	The number of incoming requests for valid requests (non-CC attacks), measured in units.	30, 000
Attack data packets	Single value	1 hour (relative)	The sum of the number of requests for the CC attack , measured in units.	100

Chart	Type	Default time range	Description	Example
Attack peak	Single value	1 hour (relative)	The highest peak of CC attack. The unit is number per minute.	100 per minute
Inbound bandwidth and attack trends	Two-line diagram	1 hour (entire point)	Trend chart of valid requests per minute and traffic bandwidth for attack requests. The unit is KB/s.	-
Request and interception trends	Two-line diagram	1 hour (entire point)	Trend chart of the total number of requests and intercepted CC attack requests per minute. The unit is number per minute.	-
Valid request rate trend	Two-line diagram	1 hour (entire point)	Trend chart of the number of valid requests per minute (non-CC attacks or 400 error requests) in the total number of all requests.	-

Chart	Type	Default time range	Description	Example
Access status distribution trend	Flow chart	1 hour (entire point)	Trend chart of various request processing statuses (400, 304, 20) per minute. The unit is number per minute.	-
CC attacks distribution	World map	1 hour (relative)	The sum of the number of CC attacks in the source country.	-
CC attack distribution	Map of China	1 hour (relative)	The sum of the number of CC attacks in the source province (China).	-
List of attacks	Table	1 hour (relative)	The attacker information of the first 100 attacks, including IP, city, network, number of attacks, and total traffic.	-
Attack access line distribution	Pie chart	1 hour (relative)	CC attack source access DDoS protection line distribution, such as telecommunications, Unicom, and BGP.	-

Chart	Type	Default time range	Description	Example
Top 10 attacked websites	Donut chart	1 hour (relative)	Top 10 attacked websites	-

- 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.

Chart	Type	Default time range	Description	Example
Page view	Single value	1 hour (relative)	The total number of requests.	100,000
Unique visitors	Single value	1 hour (relative)	Total number of independent access clients.	100,000
Inbound traffic	Single value	1 hour (relative)	The sum of inbound traffic of the website. The unit is MB.	300 MB
Network in bandwidth peak.	Single value	1 hour (relative)	The highest peak of inbound traffic rate requested by the website. The unit is bytes/s.	100 Bytes/s
Network out bandwidth peak.	Single value	1 hour (relative)	The highest peak of inbound traffic rate requested by the website. The unit is bytes/s.	100 Bytes/s

Chart	Type	Default time range	Description	Example
Traffic bandwidth trend	Two-line diagram	1 hour (entire point)	Trend chart of website inbound and outbound traffic per minute. The unit is KB/s.	-
Request and interception trends	Two-line diagram	1 hour (entire point)	Trend chart of the total number of requests and intercepted CC attack requests per minute . The unit is number per minute.	-
PV/UV access trends	Two-line diagram	1 hour (entire point)	Trend chart of PV and UV per minute. Measured in units.	-
Visitor distribution	World map	1 hour (relative)	The distributi on of visitors PV (page view) in the source country.	-
Visitor heat map	Amap	1 hour (relative)	Visitor geographic access heat map.	-
Inbound traffic distribution	World map	1 hour (relative)	Sum of inbound traffic distribution in the source country. The Unit is MB.	-

Chart	Type	Default time range	Description	Example
Inbound traffic distribution	Map of China	1 hour (relative)	Sum of inbound traffic distribution in the source province. The Unit is MB.	-
Access line distribution	Donut chart	1 hour (relative)	Source-based access DDoS protection line distribution, such as telecommunications, Unicom, and BGP.	-
Inbound traffic network provider distribution	Donut chart	1 hour (relative)	The distribution of inbound traffic that visitors access by network operators. For example, telecommunications, Unicom, mobile connections, education network. The Unit is MB.	-

Chart	Type	Default time range	Description	Example
Most visited clients	Table	1 hour (relative)	The top 100 most visited clients, including IP, city, network, request method distribution, incoming traffic, number of incorrect accesses, number of intercepted CC attacks.	-
Access domain name	Donut chart	1 hour (relative)	The top 20 most visited domain names.	-
Referer	Table	1 hour (relative)	The top 100 most redirected referer URLs, hosts, and frequency.	-
Client type distribution	Donut chart	1 hour (relative)	The top 20 most visited user agents, such as iPhone, iPad, Windows IE, Chrome.	-
Request content type distribution	Donut chart	1 hour (relative)	The top 20 most requested content types, such as HTML, Form, JSON, streaming data.	-

4.5.5 Billing method

DDoS log collection function is charged according to the charge items of the Log Service. If no log data is generated, no billing is made. Log Service is billed by resource usage and provides the FreeTier quota for DDoS Logstore.

DDoS log collection function provides functions such as log collection, storage, real-time query and analysis, and dashboards. The real-time query and analysis of log data relies on Log Service. Therefore, this feature is charged according to Log Service billing method. Log Service is billed by the resource usage and provides the FreeTier quota for DDoS Logstore. The specific fee depends on the amount of your log data. If you have Log Service enabled, but you have not turned on logging function for any website, no charge appears.

Deduction and outstanding payment

Log Service is billed by the resource usage, and the billing cycle is one day. For more information about deduction and outstanding payment, see #unique_182/unique_182_Connect_42_section_l45_r5n_vdb.

Billing item

Billing item	Description
Read and write traffic	· The read and write traffic is calculated by the traffic for transmitting compressed logs. DDoS logs are generally compressed by 5 to 10 times. · Read and write traffic also includes a loss of consumption interface that generates read traffic, generally, by using API/SDK and consumer group SDK. According to the compressed transmission traffic calculation, logs can be compressed in the API/SDK mode. Note: In the Log Service console, Preview function under Log Consumption also can generate micro-flow traffic consumption. · The data generated by the index-based query and analysis is free of read and write traffic charges. For example, the log query analysis, dashboards, and alarms in the console are not charged.
Storage space	The storage space is the sum of data size after compression and the indexed data size.

Billing item	Description
Indexing traffic	· The indexing traffic is calculated by actual index fields. Storage fee is collected in full during writing. DDoS logs enable full indexing by default. · The traffic of fields having both FullText and KeyValue indexes is calculated only once. · Indexes occupy the storage space and thus the storage space fee is collected.
Active shard rent	Only shards currently in readwrite status are counted. Rent of merged/split shards is not collected.  Note: By default, Log Service creates two shards, and enables the Auto Split Shard feature. Typically, each shard can proceed 430 GB of write data volume per day.
Read/write count	The write count of logs written into Log Service is a subject to the log generation speed. The background realization mechanism minimizes the read/write count.
Internet read traffic	The data traffic generated when Internet programs read log data collected by Log Service.

FreeTier quota

Log Service is not charged in the following cases:

- Log Service is activated, and DDoS logging function has not been enabled for any website.
- The amount of website logs that enable DDoS logging is within the free quota.
- Index-based query analysis, reports, and alarms are not charged.

Log Service provides the free quota for your DDoS Logstore. If the data volume is less than the free-quota limit, no charges appears.

Billing item	FreeTier quota
Read and write traffic	30 GB/day
Storage space	3 days
Indexing traffic	100 GB/day
Active shard rent	4 days/month
Read/write count	1 million times/day

Billing item	FreeTier quota
Internet read traffic	0
Read traffic consumption	0
Read count consumption	0

Note:
Log data storage time is set to 3 days by default, and when you modify for more than 3 days, extra charges can appear.

Billing method

When the log volume of the website that enables the log analysis function exceeds the free quota, Log Service charges the excess of the quota amount.

Billing item	Extra payment
Read/write traffic (USD/GB)	0.045
Storage space (USD/GB/day)	0.002875
Indexing traffic (USD/GB)	0.0875
Active shard rent (USD/day)	0.01
Read/write count (USD/million times)	0.03
Internet read traffic (USD/GB)	0.2

Billing example

- FreeTier quota: The average log is about 1600 bytes, about 60 million logs are generated per day, and the storage period is 3 days. The total log volume is about 96 GB per day, not exceeding the quota.
- Index: The log volume is 150 GB per day, and the 50 GB is charged (150 GB - 100 GB), which is 0.0875 x 50 = 17.5 USD per day.
- Write transmitting: The log volume is 300 GB per day, logs are compressed in six times. The actual compression size is about 50GB, and the 20GB is charged (50GB - 30GB), which is 0.045 x 20 = 0.9 USD per day.
- Storage space size:
  - 10 GB of data per day, 2 GB after compression, and 10 GB of indexing traffic. The storage period is 30 days, and the maximum storage capacity after 30 days is 30

× (10+2) = 360 GB, with a 3-day free quota, it is 27 × (10+2) = 324 GB, and the maximum charge for one day storage is 0.002875 × 324 = 0.9315 USD.

- 1 GB of data per day, 200 MB after compression, and 1 GB indexing traffic. The cumulative maximum storage capacity after 30 days is 30 × (1000 + 200) ≈ 36 GB, with a 3-day free quota, it is 27 × (1000 + 200) ≈ 32.4 GB, and the maximum charge for one day storage is 0.002875 × 32.4 = 0.09315 USD.

- Active shard rent: Currently, there are 10 shards, 7 read/write shards, and 3 read-only shards. DDoS Logstores are only charged per day. The rental fee for 3 (7 - 4) shards is 0.03 USD per day.
- Read/write count: The number of website logs is 10 billions per day, and the write count is about 500,000 (on average, 2,000 per time), free of charge.
- Internet traffic: 2 GB of Log Service data was delivered to non-Alibaba Cloud products, resulting in an external network read traffic of 0.4 USD.

Billing FAQ

- How can I modify the storage time of website logs?

  1. Log on to the Log Service console, click the Project name to enter the Logstore list. The default Project for DDoS log is `ddos-pro-project-`*`Alibaba Cloud Account ID`*.
  2. Click Modify in the Action column.
  3. On the Data Storage Time page, click Modify.

- How can I view the current log volume and estimate the cost?
  - To view the cost measurement data on day basis go to Alibaba Cloud Expense Management Center.
  - 1. Log on to the DDoS IP protection console and click Full Log on the left.
    2. Select the website which log volume you want to view, and click Log Analysis on the right.
    3. Enter the following query statement in the query box, the time range is `Yesterday ( entire point time )`:

       ```
       __topic__ : ddos_acces s_log | select count ( 1 ) as PV
       ```

    4. Click Query and select Statistics Chart with the chart type Table .

    You can get data volume of the previous day, and estimate the cost according to your current log storage time.
  - How can I configure Log Service to trigger an alarm when a large number of logs is generated?

    When a large number of DDoS logs is collected, the free quota of Log Service may be exceeded, and the certain charge can appear. If you want to receive an alarm notification when there is such a risk, you can configure Log Service to trigger an alarm when a large number of logs is generated.

    1. Log on to the DDoS IP protection console and click Full Log on the left.
    2. Select the website which log volume you want to view, and click Log Analysis on the right.
    3. Enter the following query statement in the query box, and click Query:

       ```
       *| select count ( 1 ) as PV
       ```

    4. Click Save as Quick Query in the upper-right corner of the query page to enter the information about the query, such as `ddos - metering - pv`. Then click OK.
    5. Click Save as Alarm and create an alarm configuration, see the following figure. Check the log volume of the past 1 hour every 5 minutes, and trigger an alarm if more than 5.6 million logs are generated.

        Note:

To ensure that the daily log volume is less than 100 GB free quota, the average hourly import volume is estimated to be: 100 GB ÷ 1600 bytes ÷ 24 hours ≈ 2.8 million. The example is two times of the hourly log volume, which is 5.6 millions, and can be adjusted according to the actual situation and needs.

4.6 Logs of BGP-line Anti-DDoS Pro

4.6.1 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 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

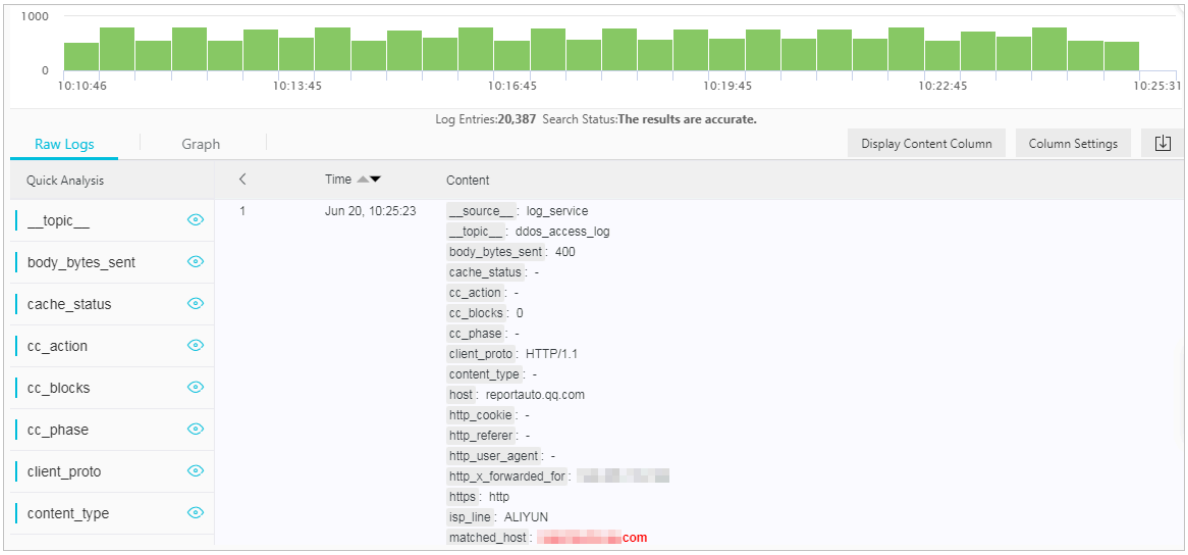
access exceptions. You can also query information such as read and write delays and exception distribution by ISP.

For example, use the following statement to query the access log entries of Anti-DDoS Pro:

```
__topic__ : DDoS_acces s_log
```

The query results are displayed, as shown in the following figure.

Figure 4-25: 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.

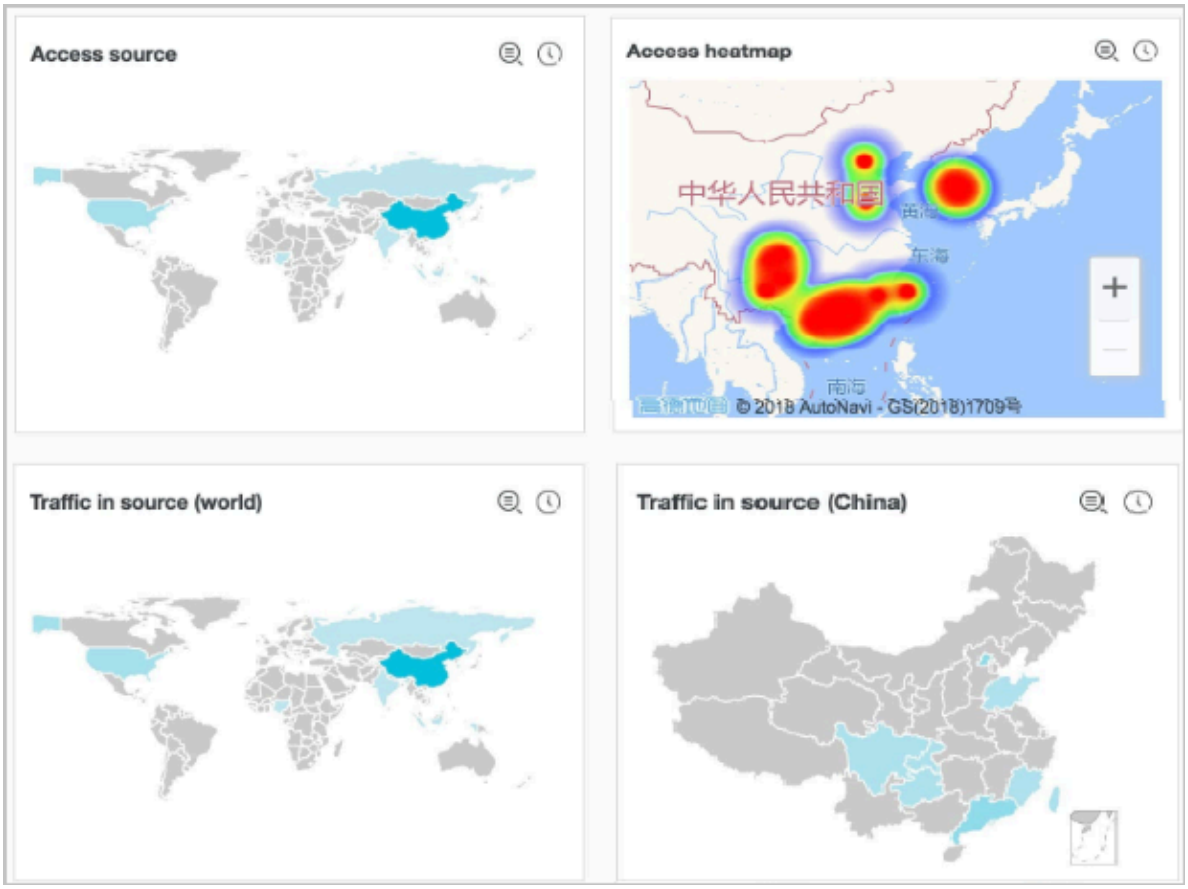
  For example, use the following statement to analyze the geographical distribution of HTTP flood attacks:

```
__topic__ : DDoS_acces s_log and cc_blocks > 0 | SELECT
  ip_to_coun try ( if ( real_clien t_ip ='-', remote_add r ,
```

```
real_clien t_ip )) as country , count ( 1 ) as " number "
group by country
```

The analysis results are displayed in a dashboard as follows:

Figure 4-26: HTTP flood attacks

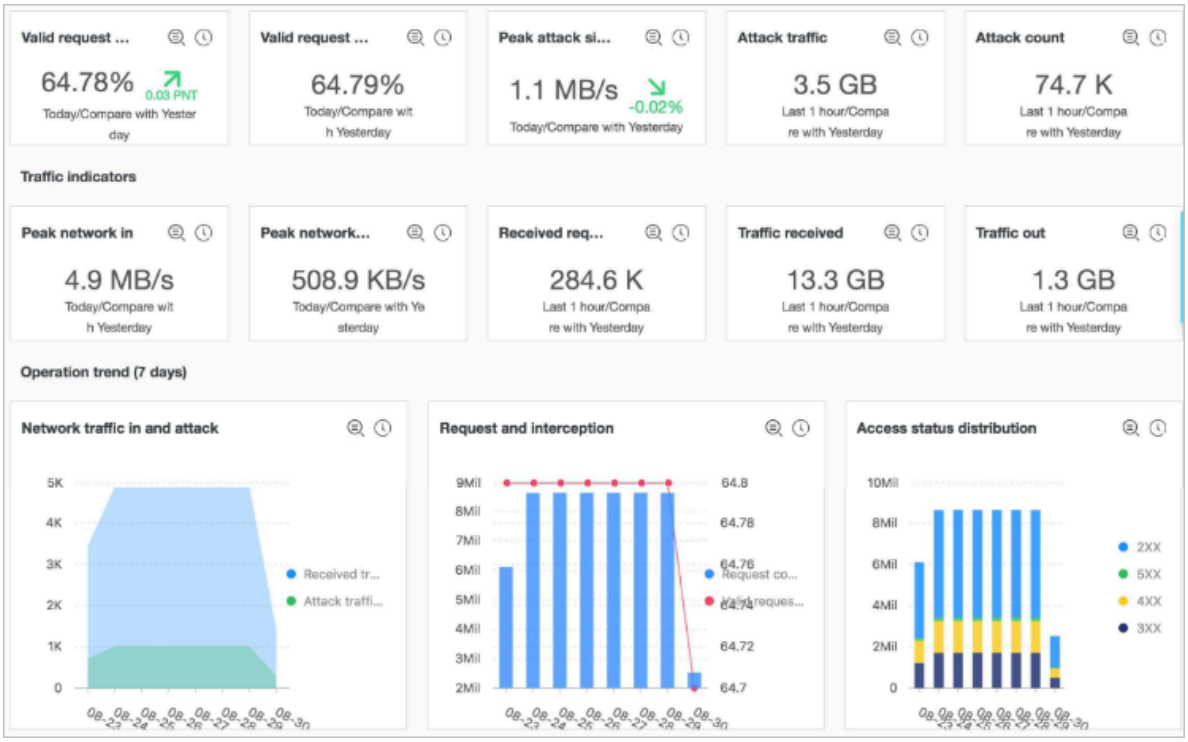


- For example, use the following statement to view PVs:

```
__topic__ : DDoS_acces s_log | select count ( 1 ) as PV
```

The analysis results are displayed in a dashboard, as shown in the following figure.

Figure 4-27: 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.

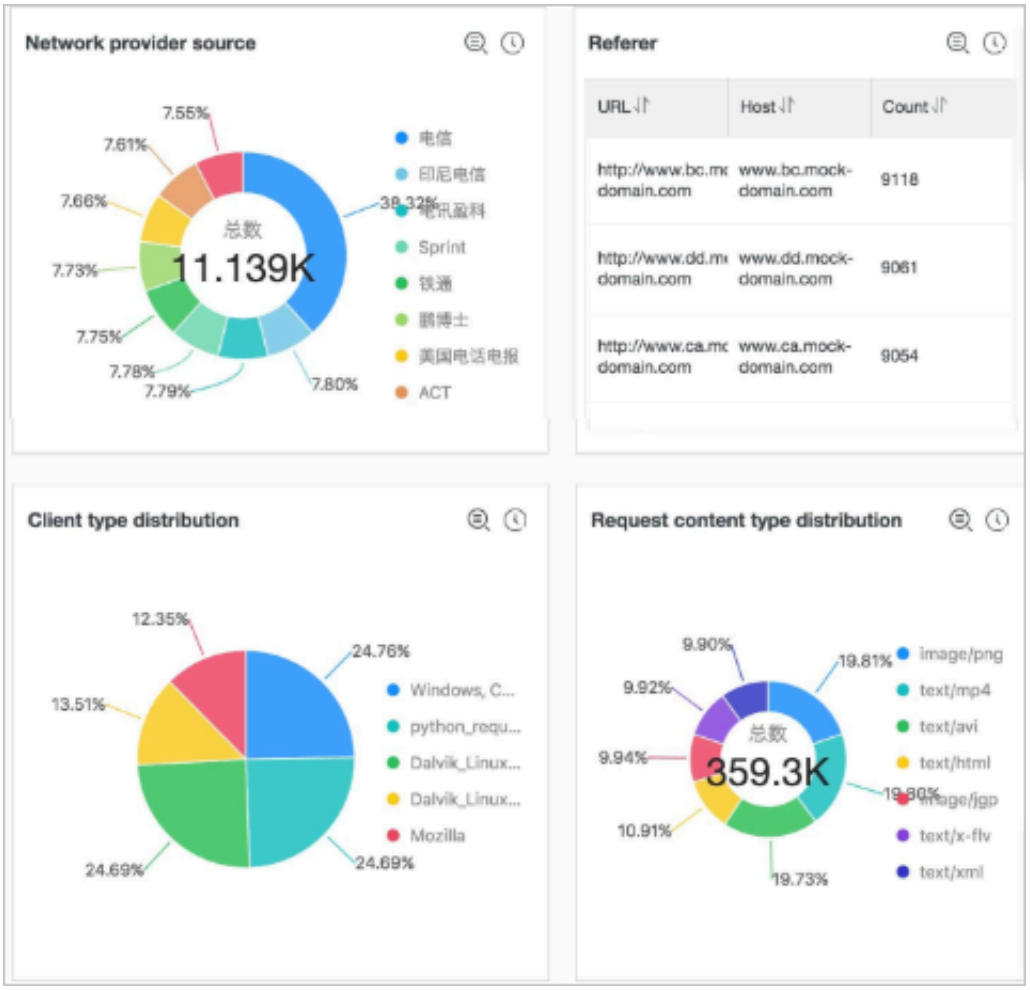
For example, use the following statement to view the visitor distribution by ISP:

```
__topic__ : ddos_acces s_log | select ip_to_prov ider ( if ( real_clien t_ip ='-', remote_add r , real_clien t_ip )) as provider , round ( sum ( request_le ngth )/ 1024 . 0 / 1024 . 0 , 3 ) as mb_in group by provider having ip_to_prov
```

```
ider ( if ( real_clien  t_ip ='-',  remote_add  r ,  real_clien
t_ip )) <> ''  order   by   mb_in   desc   limit   10
```

The analysis results are displayed in a dashboard, as shown in the following figure.

Figure 4-28: Visitor distribution



4.6.2 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

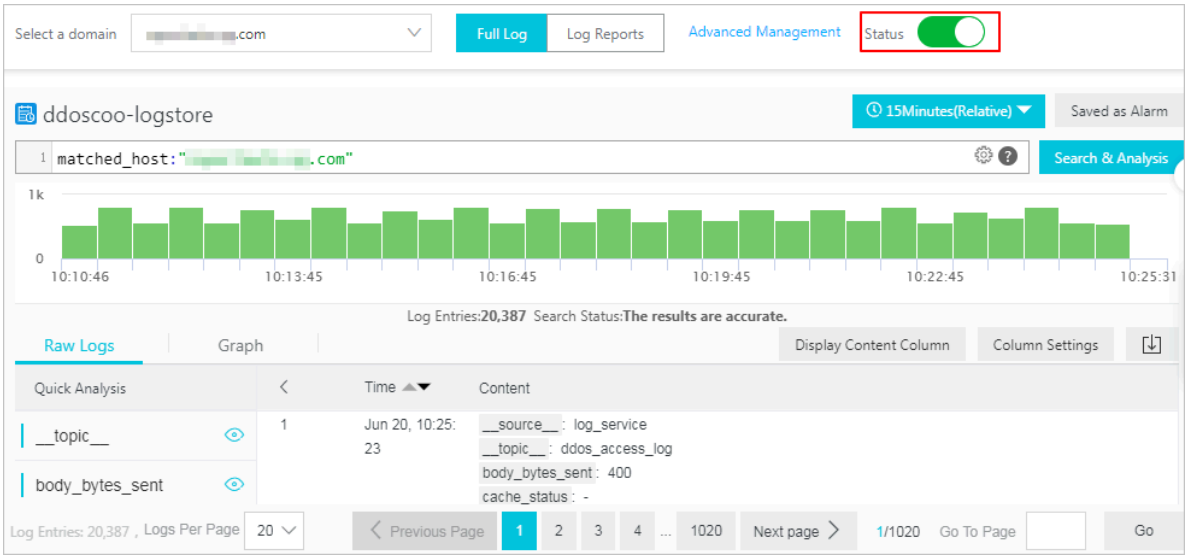
Log Service can collect website access logs and HTTP flood attack logs of Alibaba Cloud Anti-DDoS Pro in real time. Log Service also supports real-time retrieval and analysis of the collected log data and displays the query results in the form of dashboards. By analyzing the access and attack activities 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.

   Figure 4-29: Enable or disable the log collection function



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

Default configuration item	Description
Project	By default, a project is created. The name of the project is `ddoscoo-project-`*`Alibaba Cloud account ID`*`-cn-hangzhou`.

Default configuration item	Description
Logstore	By default, a Logstore is created. The name of the Logstore is `ddoscoo - logstore` . All log entries generated by the Anti-DDoS Pro log collection function are stored in this Logstore.
Region	By default, the project is created under the China (Hangzhou) region.
Shard	By default, two shards are created and the automatic shard splitting function is enabled.
Log TTL	A log entry can be stored for 30 days. After 30 days, the log entry is automatically deleted.
Log storage capacity	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.
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 #unique_186.

You can query and analyze collected log entries in real time on the Log Service page. For more information about log fields, see #unique_187. 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 #unique_186.
- 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.

4.6.3 Manage log storage space

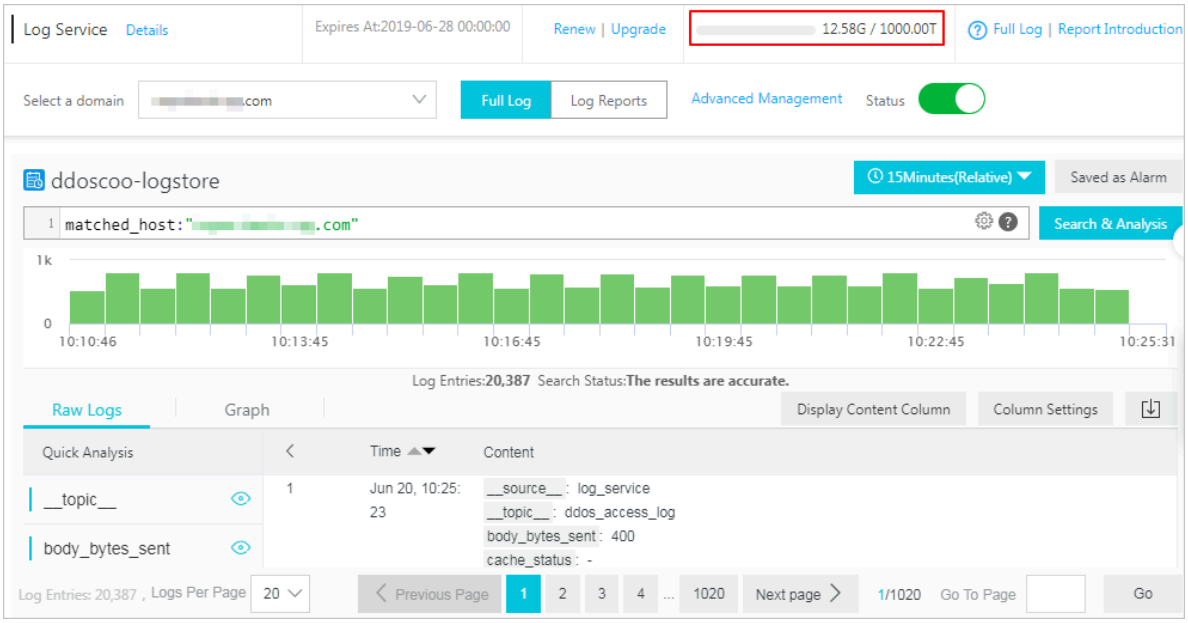
After you enable the Anti-DDoS Pro log collection function, log data is pushed to the specified Logstore in real time. You can view the usage of the log storage space in the Anti-DDoS Pro console.

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.



4.6.4 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.

Field	Description	Example
__topic__	The topic of the log entry. The value of this field is fixed to ddos_access_log.	N/A
body_bytes_sent	The size of the body in the access request, in bytes.	2
content_type	The content type.	application/x-www-form-urlencoded
host	The source website.	api.zhihu.com
http_cookie	The request cookie.	k1=v1;k2=v2
http_referer	The request referer. If no referer exists, a hyphen (-) is displayed.	http://xyz.com
http_user_agent	The User-Agent of the request.	Dalvik/2.1.0 (Linux; U; Android 7.0; EDI-AL10 Build/HUAWEIEDISON-AL10)
http_x_forwarded_for	The IP address of the upstream user redirected by proxy.	N/A
https	Indicates whether the request is an HTTPS request. · true: The request is an HTTPS request. · false: The request is an HTTP request.	true
matched_host	The matching source site, which may be a wildcard domain name. If no match is found, a hyphen (-) is displayed.	*.zhihu.com
real_client_ip	The real IP of the visitor. If no real IP is returned, a hyphen (-) is returned.	1.2.3.4
isp_line	Line information, such as BGP, China Telecom, and China Unicom.	China Telecom

Field	Description	Example
remote_addr	The IP address of the client that initiates the connection request.	1.2.3.4
remote_port	The port number of the client that initiates the connection request.	23713
request_length	The size of the request in bytes.	123
request_method	The HTTP method of the request.	GET
request_time_msec	The request time in ms.	44
request_uri	The request URI.	/answers/377971214/banner
server_name	The name of the matching host. If no match is found, the value is `default`.	api.abc.com
status	The HTTP status code.	200
time	The time when the log entry was generated.	2018-05-02T16:03:59+08:00
cc_action	The HTTP flood protection action. Valid values include none, challenge, pass, close, captcha, wait, and login.	close
cc_blocks	Indicates whether HTTP flood attacks are blocked. Valid values: · 1: block · Other values: pass	1
cc_phase	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.	server_ip_blacklist

Field	Description	Example
ua_browser	The browser.	ie9
ua_browser_family	The browser series.	internet explorer
ua_browser_type	The browser type.	web_browser
ua_browser_version	The browser version.	9.0
ua_device_type	The type of the client device.	computer
ua_os	The operating system of the client.	windows_7
ua_os_family	The operating system series of the client.	windows
upstream_addr	The list of origin addresses that are separated with commas (,). Each address is in the format of `IP : Port`.	1.2.3.4:443
upstream_ip	The real origin IP address.	1.2.3.4
upstream_response_time	The response time in seconds for the back-to-origin process.	0.044
upstream_status	The HTTP status of the back-to-origin request.	200
user_id	The user ID of the Alibaba Cloud account.	12345678
querystring	The request string.	token=bbcd&abc=123

4.6.5 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.

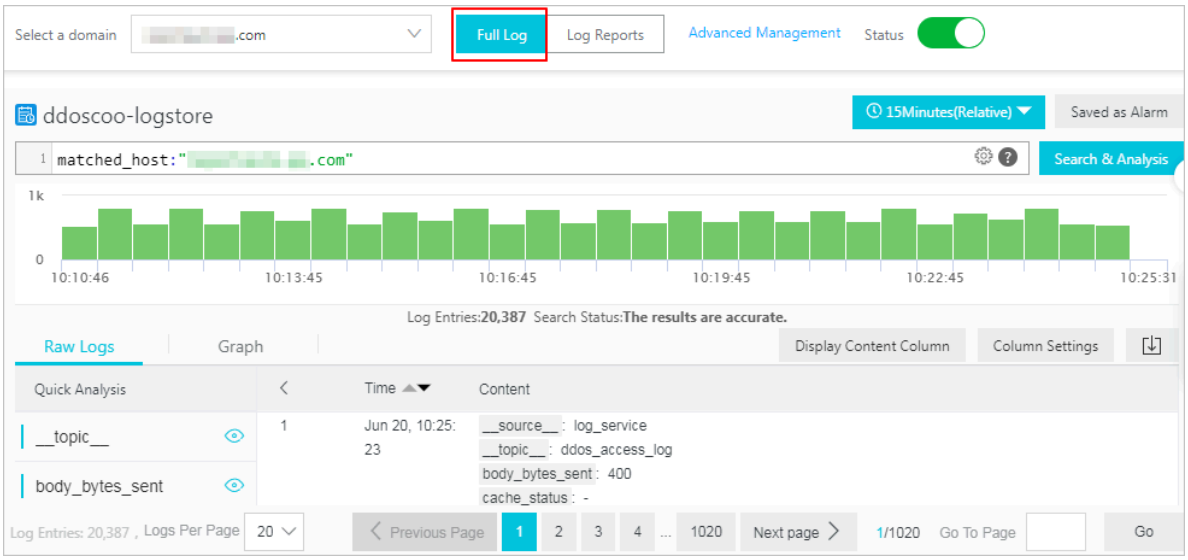
Procedure

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_ho st :" 0523 . yuanya . aliyun . com "`.

   Figure 4-30: Full log

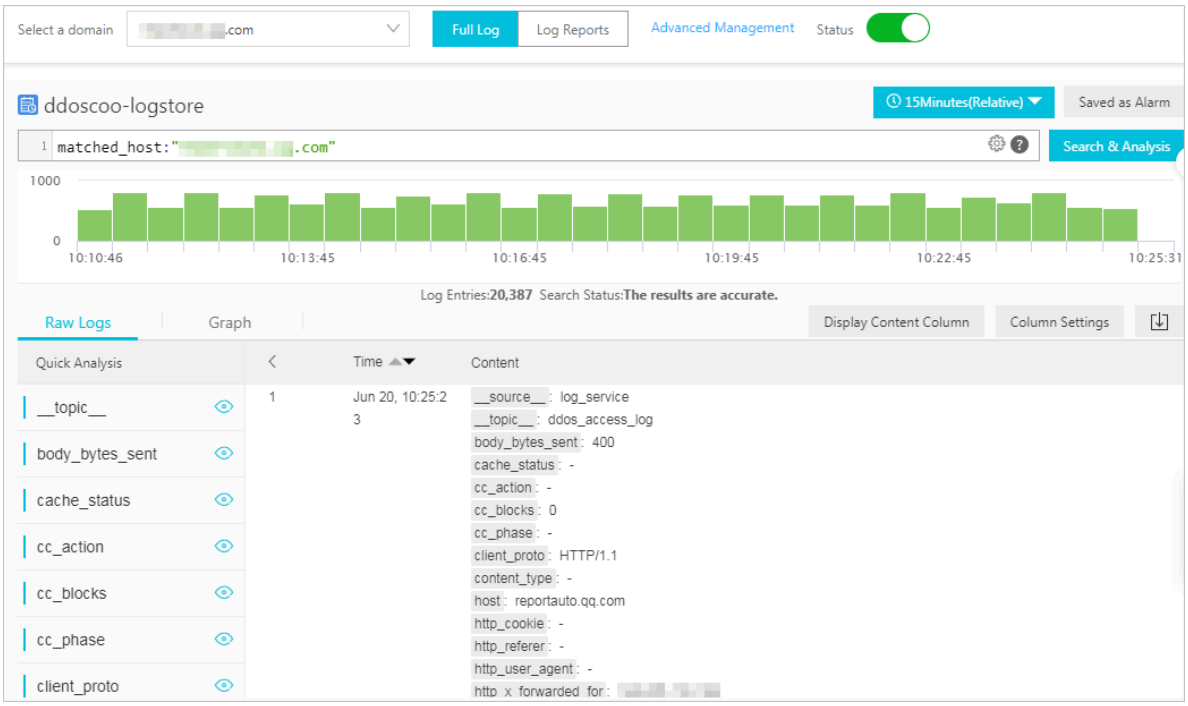


4. Enter a statement, select a time range, and then click Search & Analysis.

   Note:

Logs of Anti-DDoS Pro are stored for 180 days. By default, you can only query logs of the past 180 days.

Figure 4-31: 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.

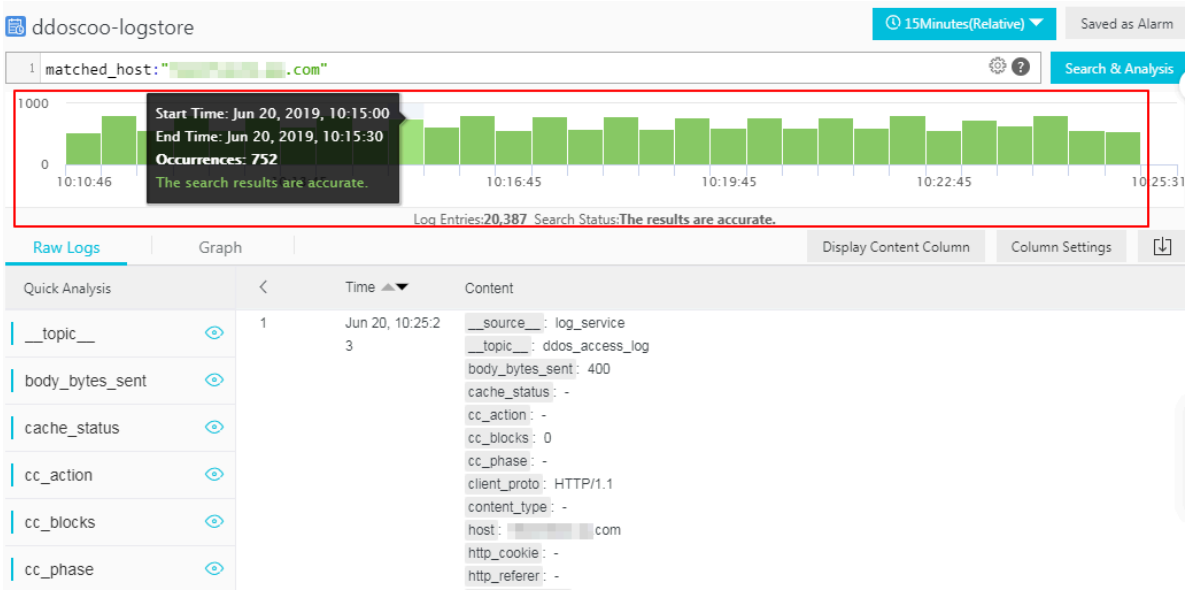
- View the distribution of log entries by time

  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 4-32: Distribution of log entries by time



- View raw logs

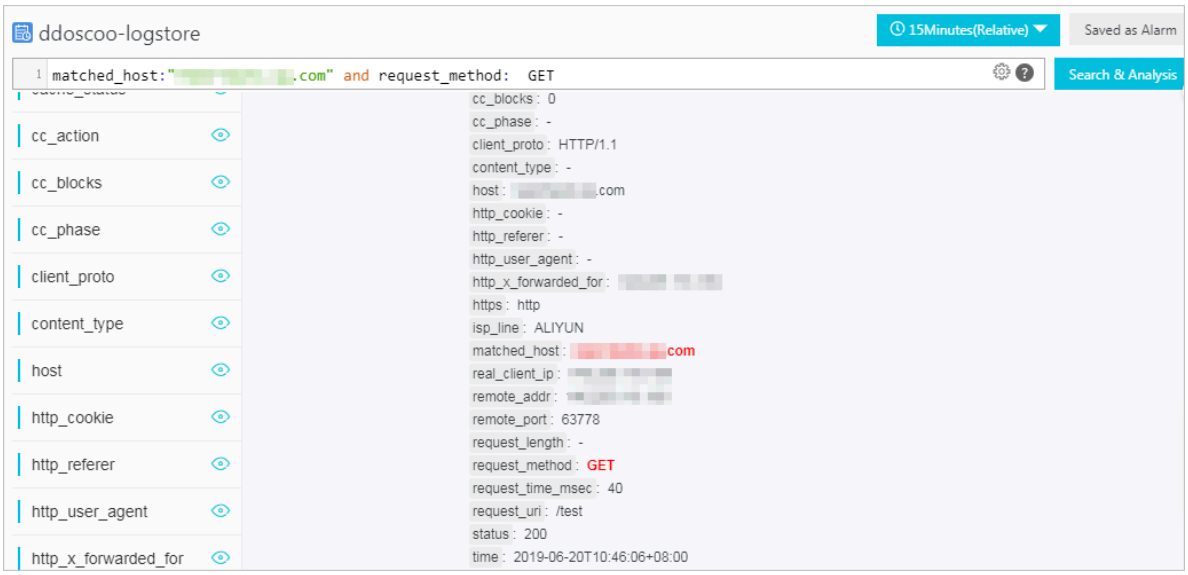
  On the Raw Logs tab page, each log entry is detailed on an individual page, which displays the time when the log is generated and the log content. The log content contains fields and their values. You can sort log entries, download logs, and click Column Settings to select column items to be displayed.

  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_me thod : GET` field, the following statement is automatically generated in the search box:

```
< The original search statement > and request_me thod :
GET
```

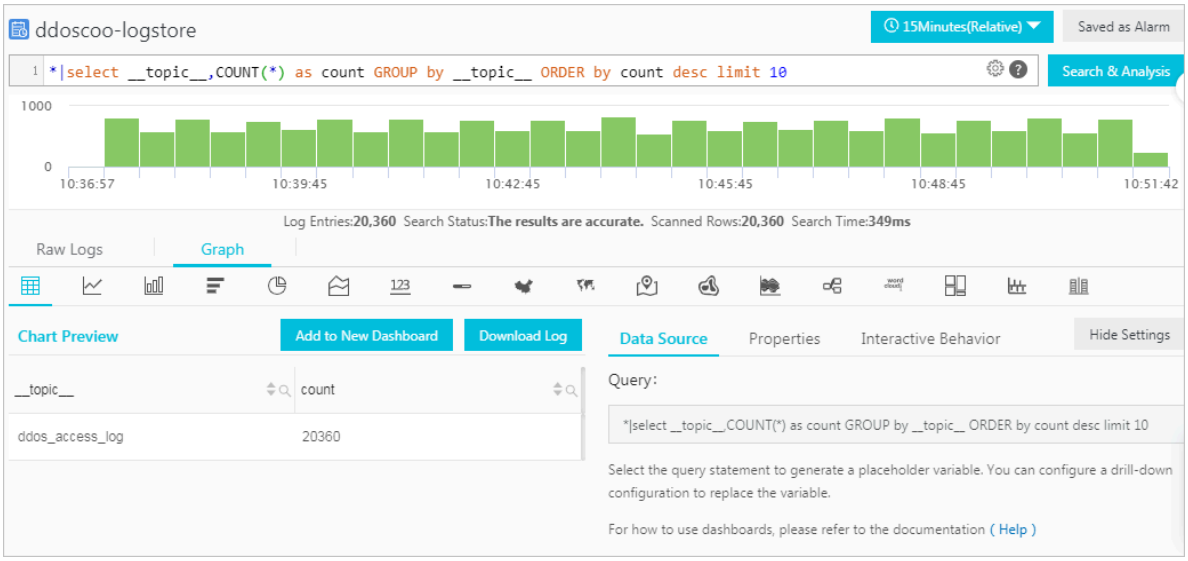
Figure 4-33: Raw logs



- 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.

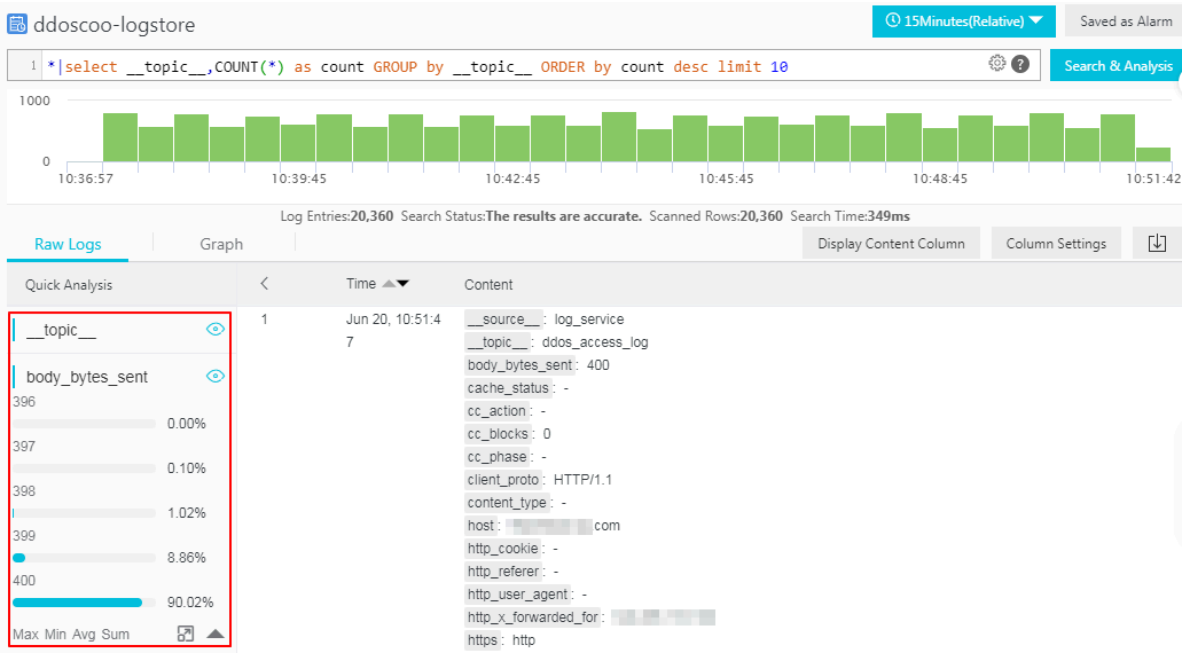
  Figure 4-34: Analysis charts



- Quick analysis

  The quick analysis feature provides you with an eay-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 4-35: 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 (|).

```
$ Search  | $ Analytics
```

Clause	Description
Query	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.
Analysis	You can use the analysis clause to process query results.

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 #unique_187.

  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_ho st : 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_clien t_ip : 10 . 2 . 3 . 4 and matched_ho st : www
    . aliyun . com and status : 404
    ```

     Note:

In the preceding examples, `matched_ho st` , `cc_blocks` , `real_clien t_ip` , and `status` are fields defined in access and attack logs of Anti-DDoS Pro. For more information about log fields, see #unique_187.

- Search by specifying numeric fields

  You can use the following statement to search for log entries where the response time exceeds 5 seconds:

  ```
  request_ti me_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_ti me_msec in ( 5000 10000 ]
  ```

  You can also use the following statement:

  ```
  request_ti me_msec > 5000 and request_ti me_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_174.

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_175.

Time-based log query and analysis

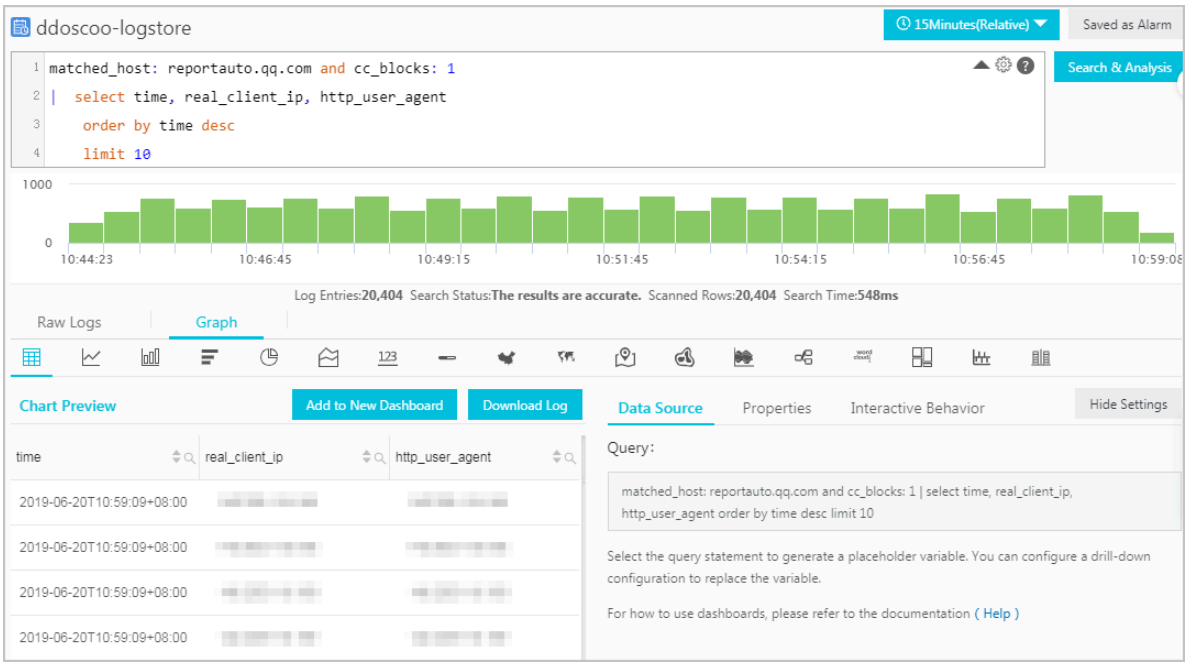
Each log entry of Anti-DDoS Pro has a `time` field in the `yyyy - MM - ddTHH : mm : ss +< time zone >` format. For example, in `2018 - 05 - 31T20 : 11 : 58 + 08 : 00`, the time zone is `UTC + 8`. Each log entry has a built-in field: `__time__`, which indicates the time when the log entry is generated, 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_ho  st :  www . aliyun . com  and  cc_blocks :  1
|  select  time ,  real_clien  t_ip ,  http_user_  agent
    order  by  time  desc
    limit  10
```

Figure 4-36: 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:

```
matched_ho  st :  www . aliyun . com    and   cc_blocks :  1
|   select   time ,
          round (( to_unixtim  e ( now ()) -  __time__ )/ 86400
 ,  1 )  as  " days_passe  d ",              real_clien  t_ip ,
 http_user_  agent
       order   by   time   desc
       limit   10
```

  Note:
  In the preceding example, `round (( to_unixtim e ( now ()) - __time__ )/ 86400 , 1 )` is used for calculation. The `to_unixtim e` 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 4-37: Query results

time	days_passed	real_client_ip	http_user_agent
2019-06-20T11:04:11+08:00	20.9	[redacted]	[redacted]
2019-06-20T11:04:11+08:00	20.9	[redacted]	[redacted]
2019-06-20T11:04:11+08:00	20.9	[redacted]	[redacted]
2019-06-20T11:04:11+08:00	20.9	[redacted]	[redacted]
2019-06-20T11:04:10+08:00	20.9	[redacted]	[redacted]
2019-06-20T11:04:10+08:00	20.9	[redacted]	[redacted]
2019-06-20T11:04:10+08:00	20.9	[redacted]	[redacted]
2019-06-20T11:04:10+08:00	20.9	[redacted]	[redacted]
2019-06-20T11:04:10+08:00	20.9	[redacted]	[redacted]
2019-06-20T11:04:10+08:00	20.9	[redacted]	[redacted]

- 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:

```
 matched_ho  st :  www . aliyun . com    and   cc_blocks :  1
|  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

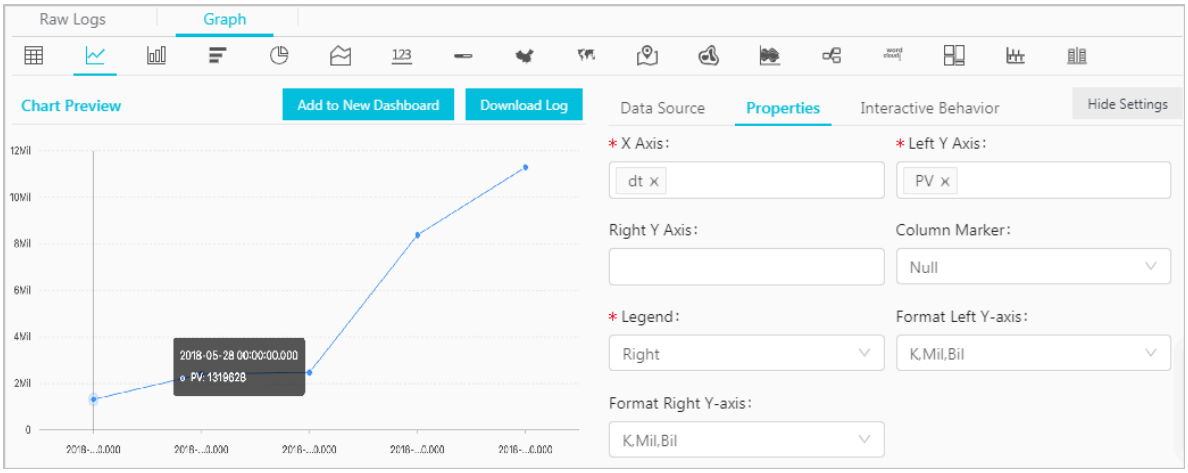
on other time units, such as `second` , `minute` , `hour` , `week` , `month` , and `year` . For more information about time-related functions, see #unique_176.

Figure 4-38: Analysis results

dt	PV
2018-05-28 00:00:00.000	1319628
2018-05-29 00:00:00.000	2402020
2018-05-30 00:00:00.000	2473332
2018-05-31 00:00:00.000	8381076
2018-06-01 00:00:00.000	11293642

The analysis results can be displayed in a line chart.

Figure 4-39: 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:

```
matched_host : www.aliyun.com and cc_blocks : 1
| select from_unixtime ( __time__ - __time__ % 300 ) as
dt ,
        count ( 1 ) as PV
```

```
group   by   dt
order   by   dt
limit   1000
```

Note:

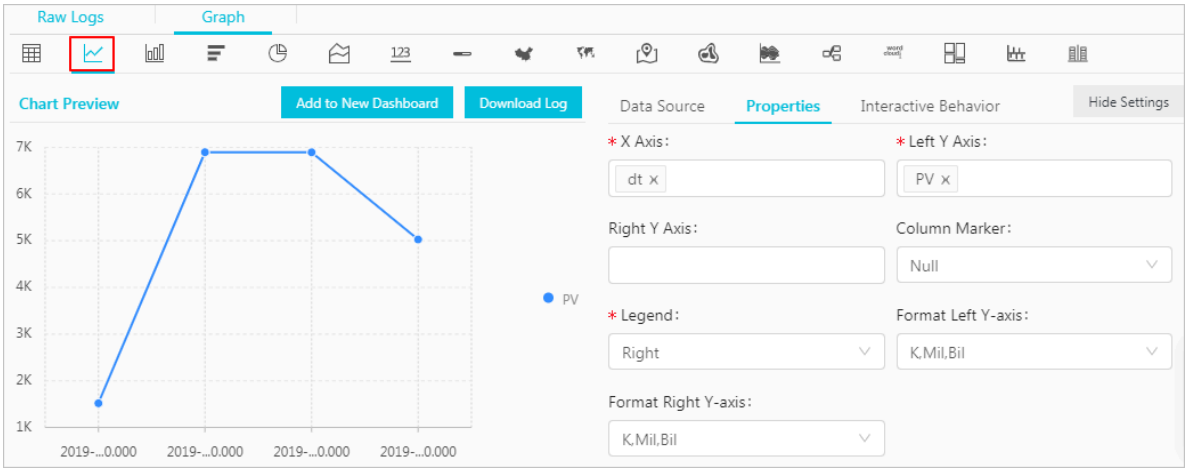
In the preceding example, the `__time__ - __time__ % 300` expression contains the built-in time field, and the expression result is formatted by using the `from_unixt ime` function. Each log entry is assigned to a group that indicates a time range of 5 minutes (300 seconds). The total number of log entries 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 4-40: Analysis results

dt	PV
2019-06-20 17:35:00.000	1430
2019-06-20 17:40:00.000	6893
2019-06-20 17:45:00.000	6902
2019-06-20 17:50:00.000	5230

The analysis results can be displayed in a line chart.

Figure 4-41: Line chart



For more information about time-related functions, see #unique_176. For example, the `date_parse` and `date_forma t` 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_clien t_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_add r` field.

- Distribution of attackers by country

  You can use the following statements to analyze the source countries of CC attacks on a website:

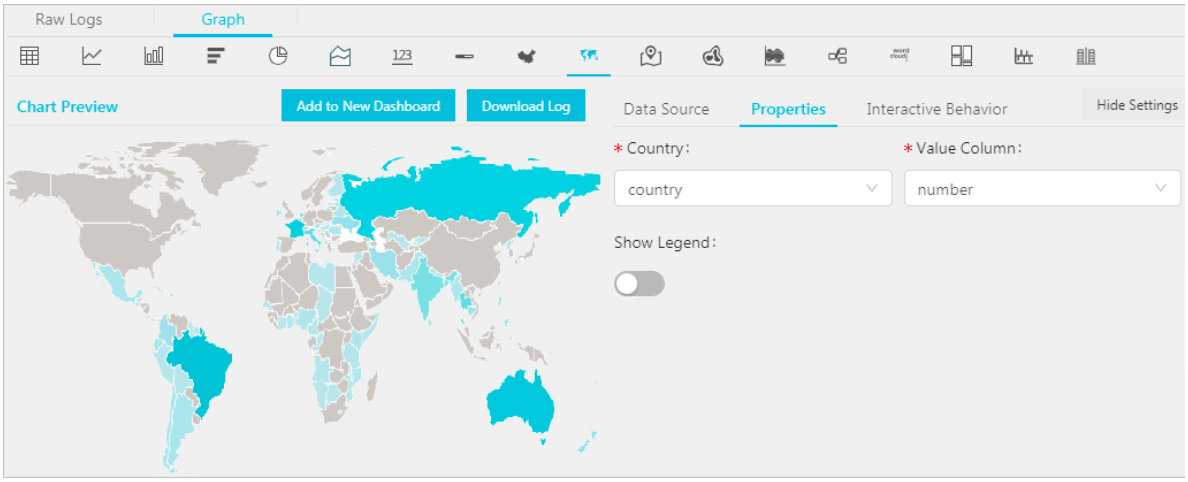
```
 matched_ho  st :  www . aliyun . com    and   cc_blocks :  1
|  SELECT   ip_to_coun  try ( if ( real_clien  t_ip ='-',
 remote_add  r ,  real_clien  t_ip ))  as   country ,
          count ( 1 )  as  " number   of   attacks "
          group   by   country
```

  Note:
  In the preceding example, the `if ( condition , option1 , option2 )` function returns the real client IP address. If `real_clien t_ip` is `-`, the function returns the value of `remote_add r` . Otherwise, the function returns

`real_clien t_ip` . The `ip_to_coun try` 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 4-42: World map



- Distribution of visitors by province

  The following example shows how to use the `ip_to_prov ince` function to obtain the distribution of visitors by province.

```
matched_ho st : www . aliyun . com and cc_blocks : 1
| SELECT ip_to_prov ince ( if ( real_clien t_ip ='-',
remote_add r , real_clien t_ip )) as province ,
        count ( 1 ) as " number of attacks "
        group by province
```

  Note:
  In the preceding example, the `ip_to_prov ince` 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.

```
matched_ho st : www . aliyun . com and cc_blocks : 1
| SELECT ip_to_geo ( if ( real_clien t_ip ='-', remote_add r
, real_clien t_ip )) as geo ,
        count ( 1 ) as " number of attacks "
        group by geo
```

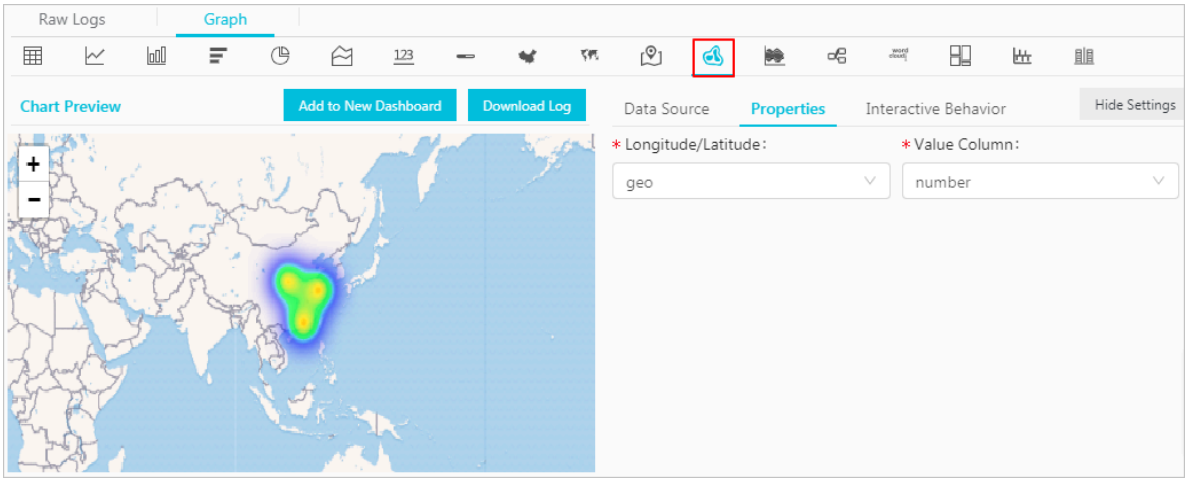
```
limit   10000
```

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 4-43: Analysis results: distribution of attackers in heatmap

The analysis results can be displayed in a Amap.

Figure 4-44: Amap



For more information about IP-based functions, see #unique_177. For example, you can use the `ip_to_prov ider` function to obtain the provider of IP addresses. You can use the `ip_to_doma in` function to determine whether an IP address is public or private.

4.6.6 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.

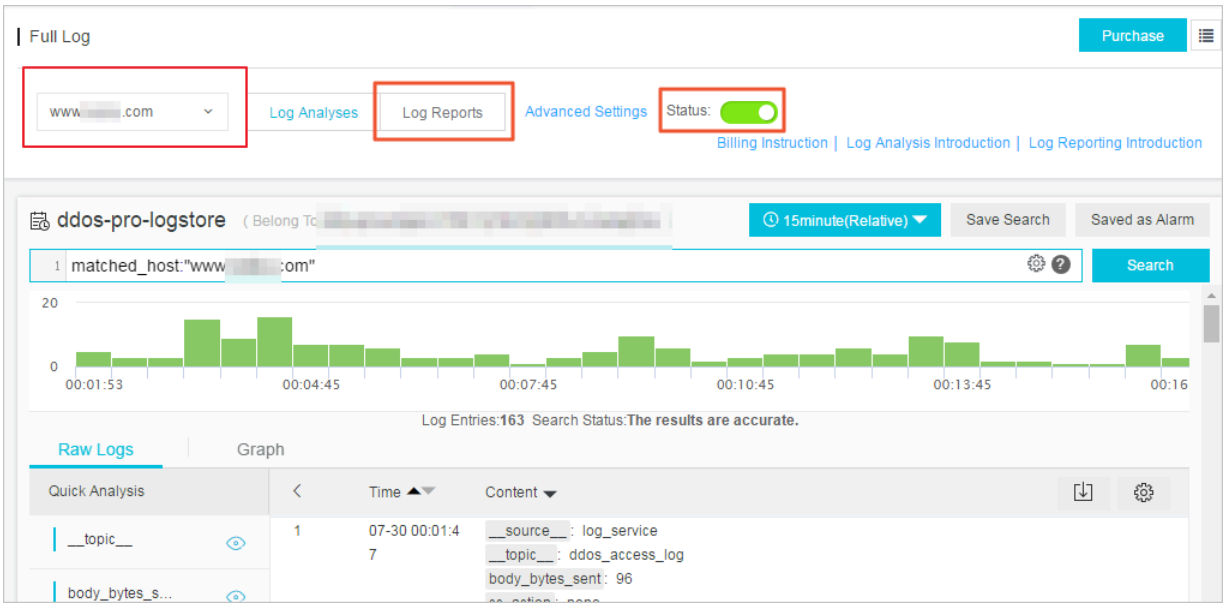
View log reports

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_ho st :" 0523 . yuanya . aliyun . com "`.

Figure 4-45: View log reports

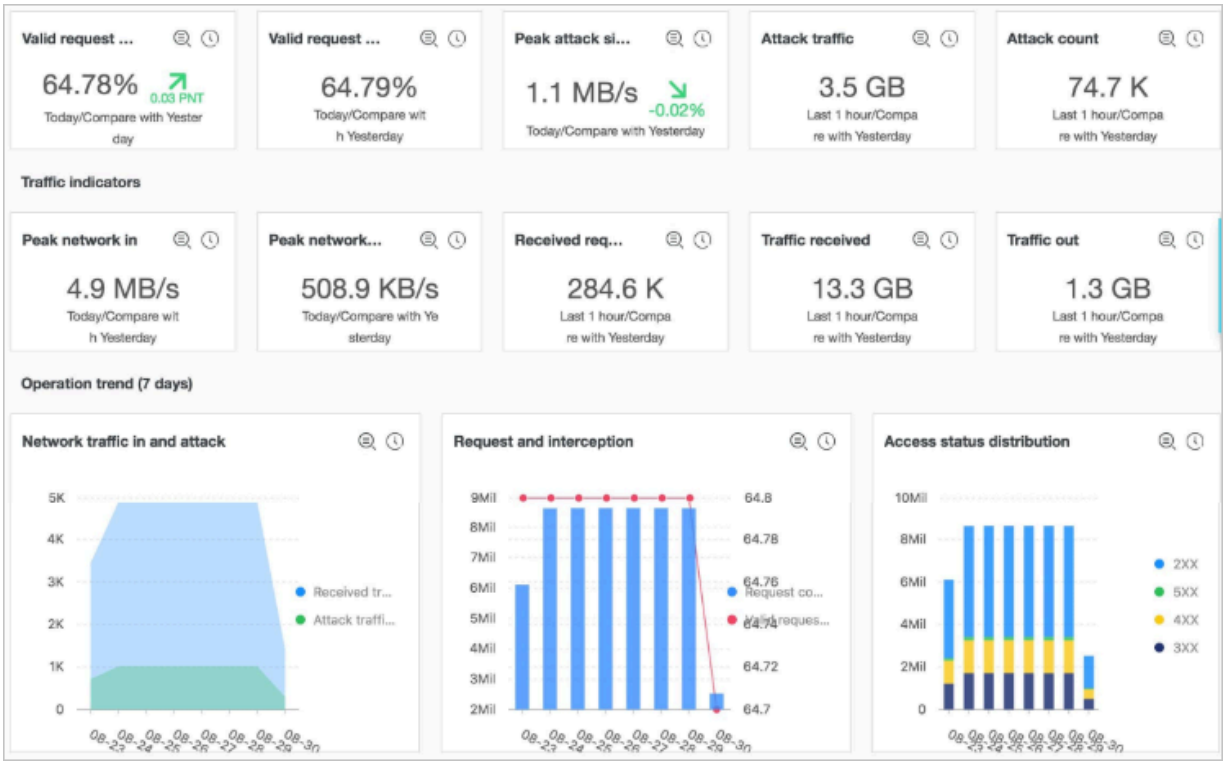


After you enable the log collection feature of Anti-DDoS Pro for your website, Log Service automatically creates two default dashboards: Operation Center and Access Center.

Dashboard name	Description
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.

Dashboard name	Description
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.

Figure 4-46: Default dashboards



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 4-47: Specify a time range

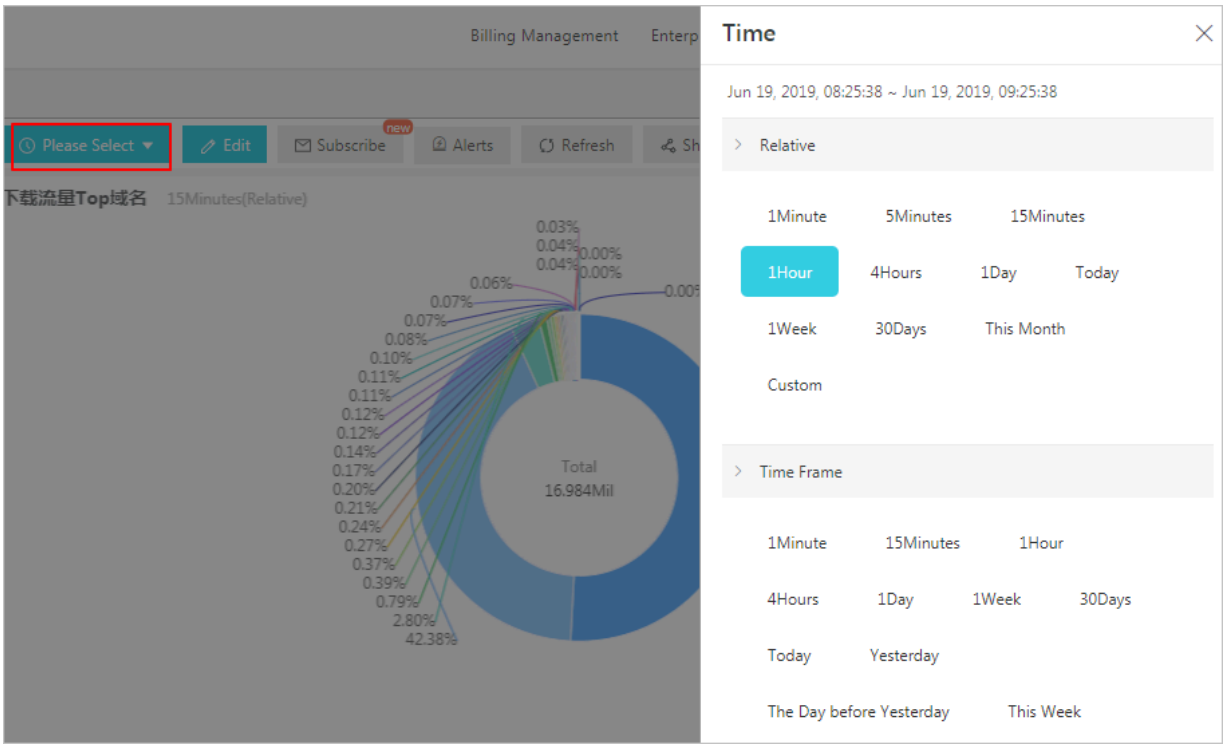


Chart types

The dashboard displays multiple types of charts based on a predefined layout. For more information about chart types, see #unique_180.

Chart type	Description
Single value chart	Displays key indicators, such as valid request ratio and attack peaks.
Line chart and area chart	Displays trends of key indicators within a time range, such as inbound traffic, attacks, and interception.
Map	Displays the geographical distribution of visitors and attackers, such as CC attack source and access heatmap .

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.

Chart name	Type	Default time range	Description	Example
Valid request ratio	Single value chart	1 hour (relative)	The ratio of valid requests to all requests. Valid requests are requests except CC attack requests and 400 bad requests.	95%
Valid request traffic ratio	Single value chart	1 hour (relative)	The ratio of the traffic incurred by valid requests to the traffic incurred by all requests.	95%
Traffic received	Single value chart	1 hour (relative)	The total inbound traffic incurred by valid requests. Unit: MB.	300 MB

Chart name	Type	Default time range	Description	Example
Attack traffic	Single value chart	1 hour (relative)	The total inbound traffic incurred by CC attacks. Unit: MB.	30 MB
Traffic out	Single value chart	1 hour (relative)	The total outbound traffic that is generated by valid requests. Unit: MB.	300 MB
Peak network in	Single value chart	1 hour (relative)	The maximum inbound throughput of the website's requests. Unit: Bytes/s.	100 Bytes/s
Peak network out	Single value chart	1 hour (relative)	The maximum outbound throughput of the website's requests. Unit: Bytes/s.	100 Bytes/s
Received requests	Single value chart	1 hour (relative)	The total number of received valid requests (not CC attacks).	30,000
Attack count	Single value chart	1 hour (relative)	The total number of requests initiated by CC attacks.	100
Peak attack size	Single value chart	1 hour (relative)	The maximum number of CC attack requests per minute.	100 per minute

Chart name	Type	Default time range	Description	Example
Network traffic in and attack	Two-line chart	1 hour (time frame)	The traffic incurred by valid requests and attack requests per minute. Unit: KB/s.	N/A
Request and interception	Two-line chart	1 hour (time frame)	The total number of requests and intercepted CC attack requests per minute.	N/A
Valid request ratio	Two-line chart	1 hour (time frame)	The ratio of valid requests (excluding CC attack requests and 400 bad requests) to all requests every minute.	N/A
Access status distribution	Flow chart	1 hour (time frame)	The distribution of requests with different status codes (such as 400, 304, and 200) per minute.	N/A
Attack source (world)	World map	1 hour (relative)	The distribution of CC attacks in origin countries.	N/A
Attack source (China)	China map	1 hour (relative)	The distribution of CC attacks that originate in the provinces of China.	N/A

Chart name	Type	Default time range	Description	Example
Attacker list	Table	1 hour (relative)	The information about top 100 attackers, including IP addresses, countries, cities, network, attack count, and attack throughput.	N/A
Attack access line distribution	Pie chart	1 hour (relative)	The distribution of ISP lines accessed by CC attacks, such as lines of China Telecom, China Unicom, and BGP.	N/A
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.

Chart name	Type	Default time range	Description	Example
PV	Single value chart	1 hour (relative)	The total number of page views (PVs).	100,000
UV	Single value chart	1 hour (relative)	The total number of unique visitors (UVs).	100,000

Chart name	Type	Default time range	Description	Example
Traffic in	Single value chart	1 hour (relative)	The total inbound traffic. Unit: MB.	300 MB
Peak network in traffic	Single value chart	1 hour (relative)	The maximum inbound throughput of the website's requests. Unit: Bytes/s.	100 Bytes/s
Peak network out traffic	Single value chart	1 hour (relative)	The maximum outbound throughput of the website's requests. Unit: Bytes/s.	100 Bytes/s
Traffic network trend	Two-line chart	1 hour (time frame)	The trend of inbound and outbound website traffic per minute. Unit: Bytes/s.	N/A
Request and interception	Two-line chart	1 hour (time frame)	The total number of requests and intercepted CC attack requests per minute.	N/A
PV/UV trends	Two-line chart	1 hour (time frame)	The trends of PV and UV per minute.	N/A
Access source	World map	1 hour (relative)	The distribution of visitors in origin countries.	N/A

Chart name	Type	Default time range	Description	Example
Access heatmap	Amap	1 hour (relative)	The heatmap that represents the geographical locations of visitors.	N/A
Traffic in source (world)	World map	1 hour (relative)	The distribution of inbound traffic in origin countries. Unit: MB.	N/A
Traffic in source (China)	China map	1 hour (relative)	The distribution of inbound traffic that originates in the provinces of China. Unit: MB.	N/A
Access line distribution	Doughnut chart	1 hour (relative)	The distribution of ISP lines accessed by visitors, such as lines of China Telecom, China Unicom, and BGP.	N/A
Network provider source	Doughnut chart	1 hour (relative)	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.	N/A

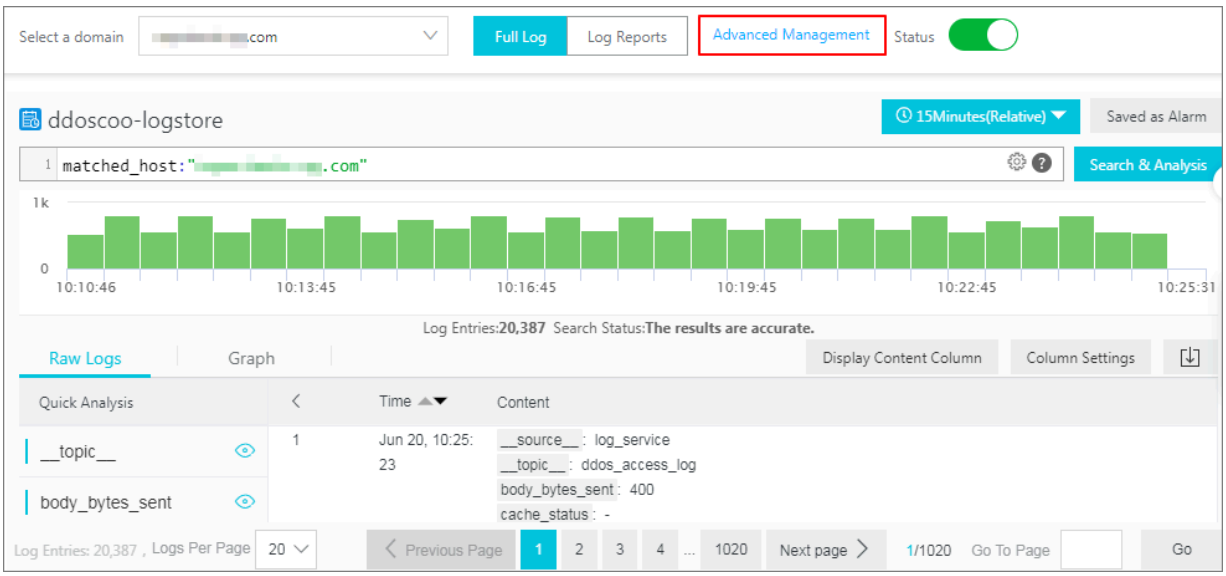
Chart name	Type	Default time range	Description	Example
Top clients	Table	1 hour (relative)	The information about the top 100 most visited clients. The information includes IP addresses, countries, cities, network, request method distribution, inbound traffic, the number of invalid requests, and the number of intercepted CC attacks.	N/A
Accessed websites	Doughnut chart	1 hour (relative)	The domain names of the top 20 most visited websites.	N/A
Referer	Table	1 hour (relative)	The top 100 most used referer URLs, redirection target hosts, and the number of redirections.	N/A
PC client distribution	Doughnut chart	1 hour (relative)	The top 20 most used user agents, such as iPhone, iPad, Windows IE, and Chrome.	N/A

Chart name	Type	Default time range	Description	Example
Request content type distribution	Doughnut chart	1 hour (relative)	The top 20 most requested content types, such as HTML, Form, JSON, and streaming data.	N/A

4.6.7 Advanced settings

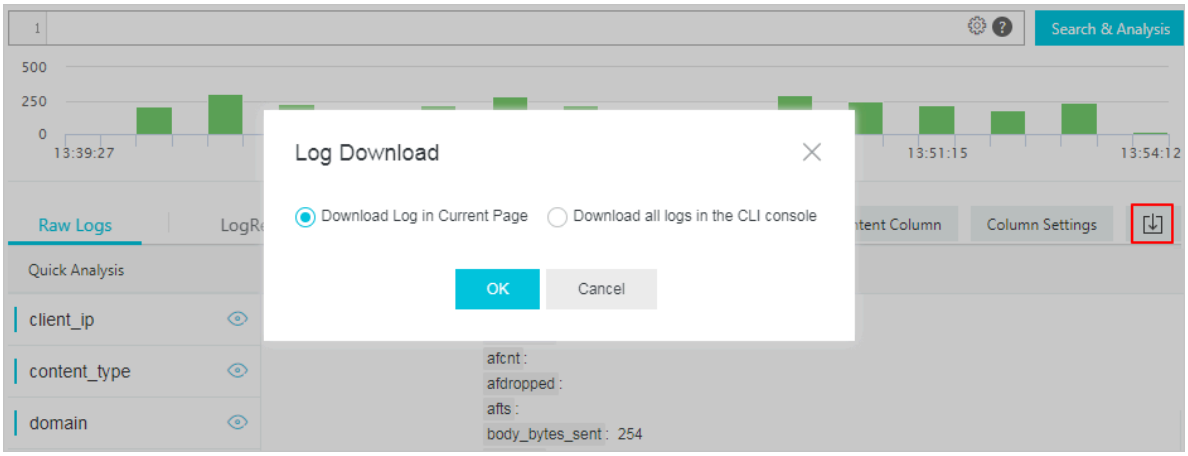
The Anti-DDoS Pro log collection function supports advanced management. You can click Advanced Management to go to the Logstores page in the Log Service console. You can perform advanced operations on Anti-DDoS Pro logs, including real-time subscription and consumption, data shipping, and other visualization 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.



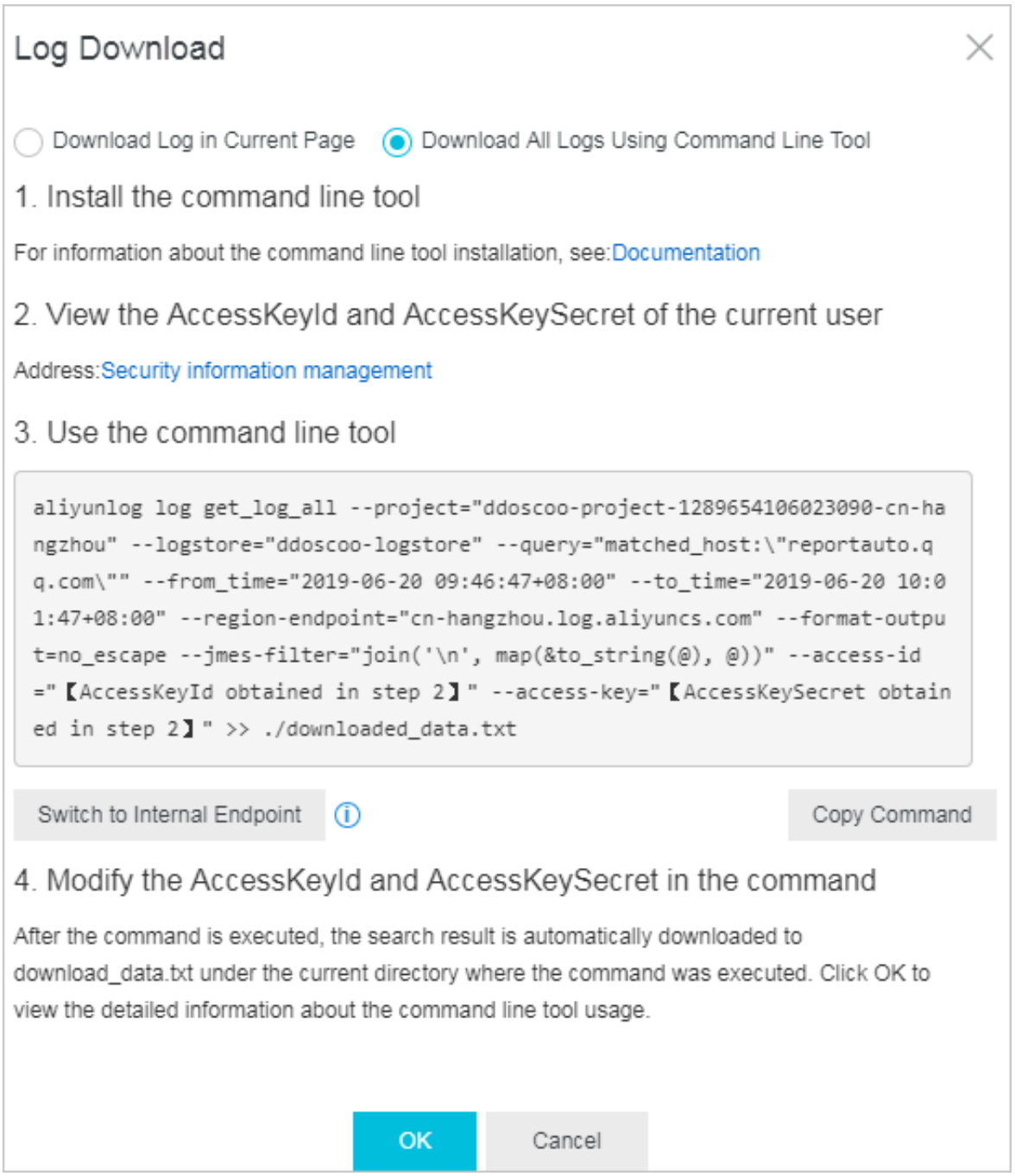
Export log data

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.



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 `AccessKeyI d obtained in step 2` and `AccessKeyS ecret obtained in step 2` with the AccessKey ID 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

4.7 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.

Features

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 posses 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 Logstrores, 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 s case, you need to 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.

4.8 WAF logs

4.8.1 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-commence, 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.

4.8.2 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.

Log storage period	Log storage size	Recommended scenarios	For International region instances		For Mainland China region instances	
			Monthly subscription	Yearly subscription	Monthly subscription	Yearly subscription
180 days	3 TB	Average daily QPS is up to 80.	USD 450	USD 5,400	USD 225	USD 2,700
	5 TB	Average daily QPS is up to 120.	USD 750	USD 9,000	USD 375	USD 4,500
	10 TB	Average daily QPS is up to 260.	USD 1,500	USD 18,000	USD 750	USD 9,000
	20 TB	Average daily QPS is up to 500.	USD 3,000	USD 36,000	USD 1,500	USD 18,000
	50 TB	Average daily QPS is up to 1,200.	USD 7,500	USD 90,000	USD 3,000	USD 36,000
	100 TB	Average daily QPS is up to 2,600.	USD 15,000	USD 180,000	USD 7,500	USD 90,000

Log storage period	Log storage size	Recommended scenarios	For International region instances		For Mainland China region instances	
			Monthly subscription	Yearly subscription	Monthly subscription	Yearly subscription
360 days	5 TB	Average daily QPS is up to 60.	USD 750	USD 9,000	USD 375	USD 4,500
	10 TB	Average daily QPS is up to 120.	USD 1,500	USD 18,000	USD 750	USD 9,000
	20 TB	Average daily QPS is up to 260.	USD 3,000	USD 36,000	USD 1,500	USD 18,000
	50 TB	Average daily QPS is up to 600.	USD 7,500	USD 90,000	USD 3,000	USD 36,000
	100 TB	Average daily QPS is up to 1,200.	USD 15,000	USD 180,000	USD 7,500	USD 90,000

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.

4.8.3 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.

4.8.4 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.

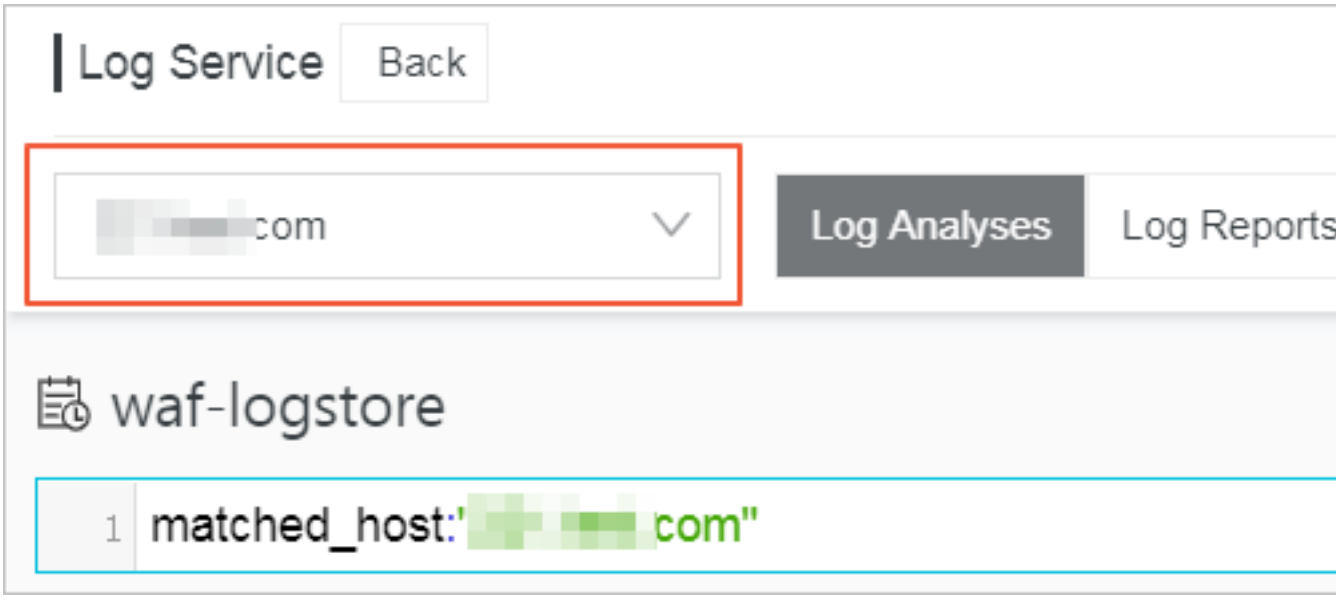
Procedure

1. Log on to the Web Application Firewall console.

2. Choose App Market > App Management, and click Real-time Log Query and Analysis Service.

   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

Default configuration item	Description
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	Description
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, seeGrant log query and analysis permissions to a RAM user.

4.8.5 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_ho st : " 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.

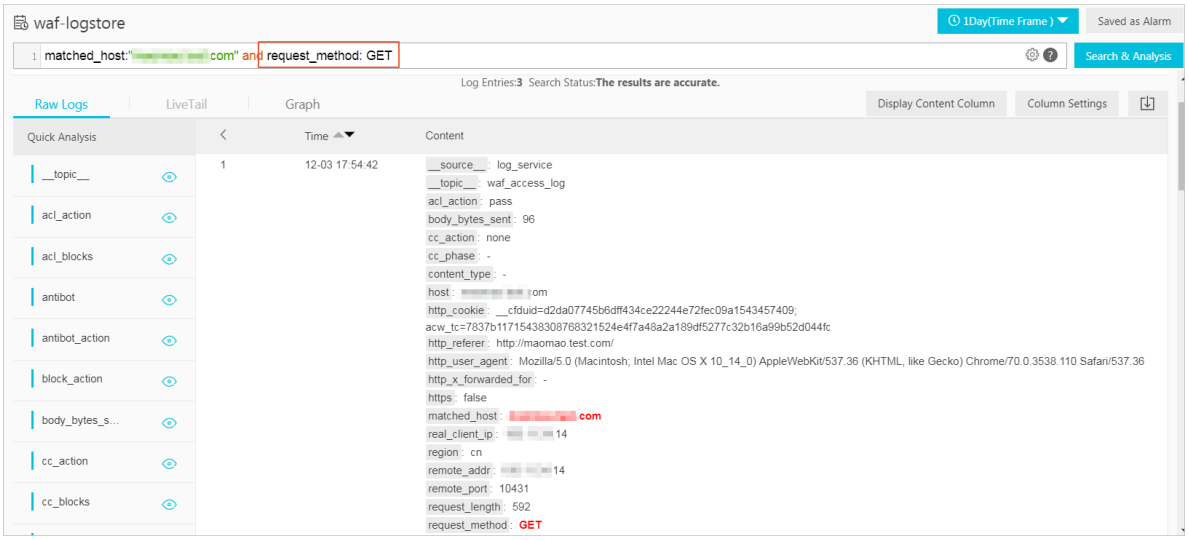
> **Note:**
> You can hold down the left mouse button and drag the histogram to select a shorter period. The `time picker` automatically updates the time period, and the query results are also updated based on the updated time period.

- 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.

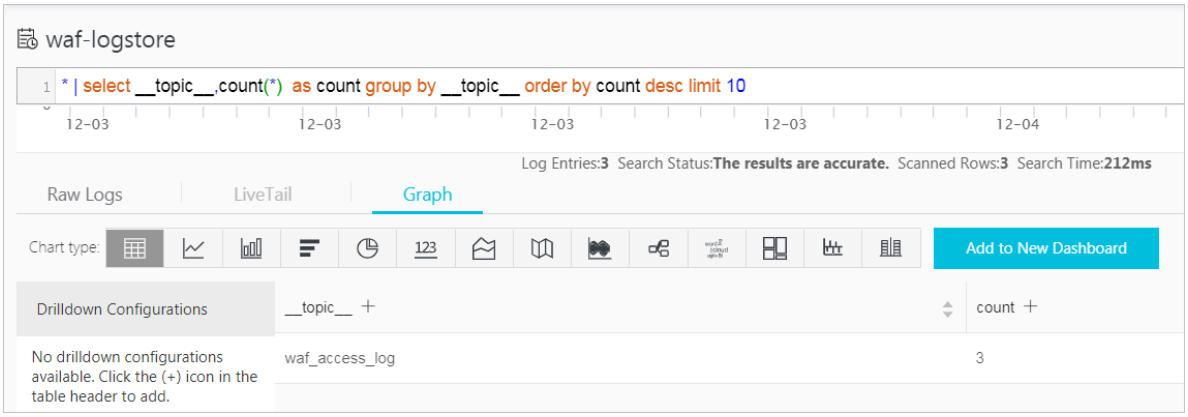
  Additionally, you can click a value or a property name to add a query criterion to the query box. For example, if you click the value `GET` in the `request_me thod : GET` filed, the query statement in the query box is updated to:

```
< The original query statement > and request_me thod :
 GET
```



- View analysis graphs

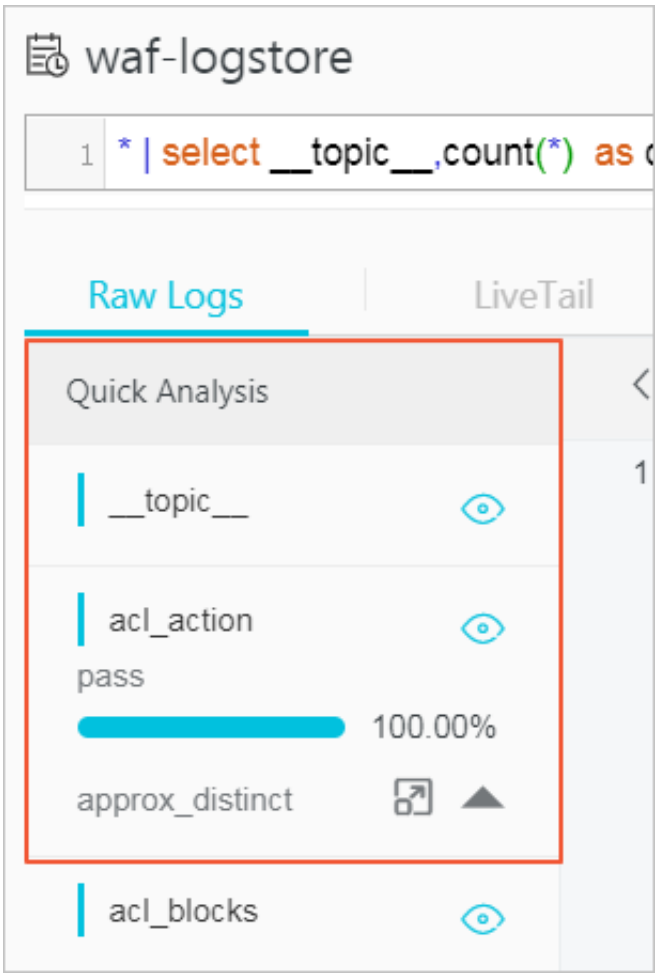
  Log Service enables you to display the analysis results in graphs. You can select the graph type as needed in the Graph tab. For more information, see Analysis graph.



- 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
```

Type	Description
Query (Search)	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.
Analysis (Analytics)	Performs computing and statistics to the query results or all log entries.

 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_ *
    ```

     Note:

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_ho st : 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_clien t_ip : 1 . 2 . 3 . 4 and matched_ho st : www .
    aliyun . com and status : 404
    ```

     Note:

In this example, the `matched_ho st` , `cc_blocks` , `real_clien t_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_ti me_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_ti me_msec in ( 5000 10000 ]
  ```

  Note:
  The following query statement has the same function.

  ```
  request_ti me_msec > 5000 and request_ti me_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, seeIndex 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, seeSyntax description.

Note:

- 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.

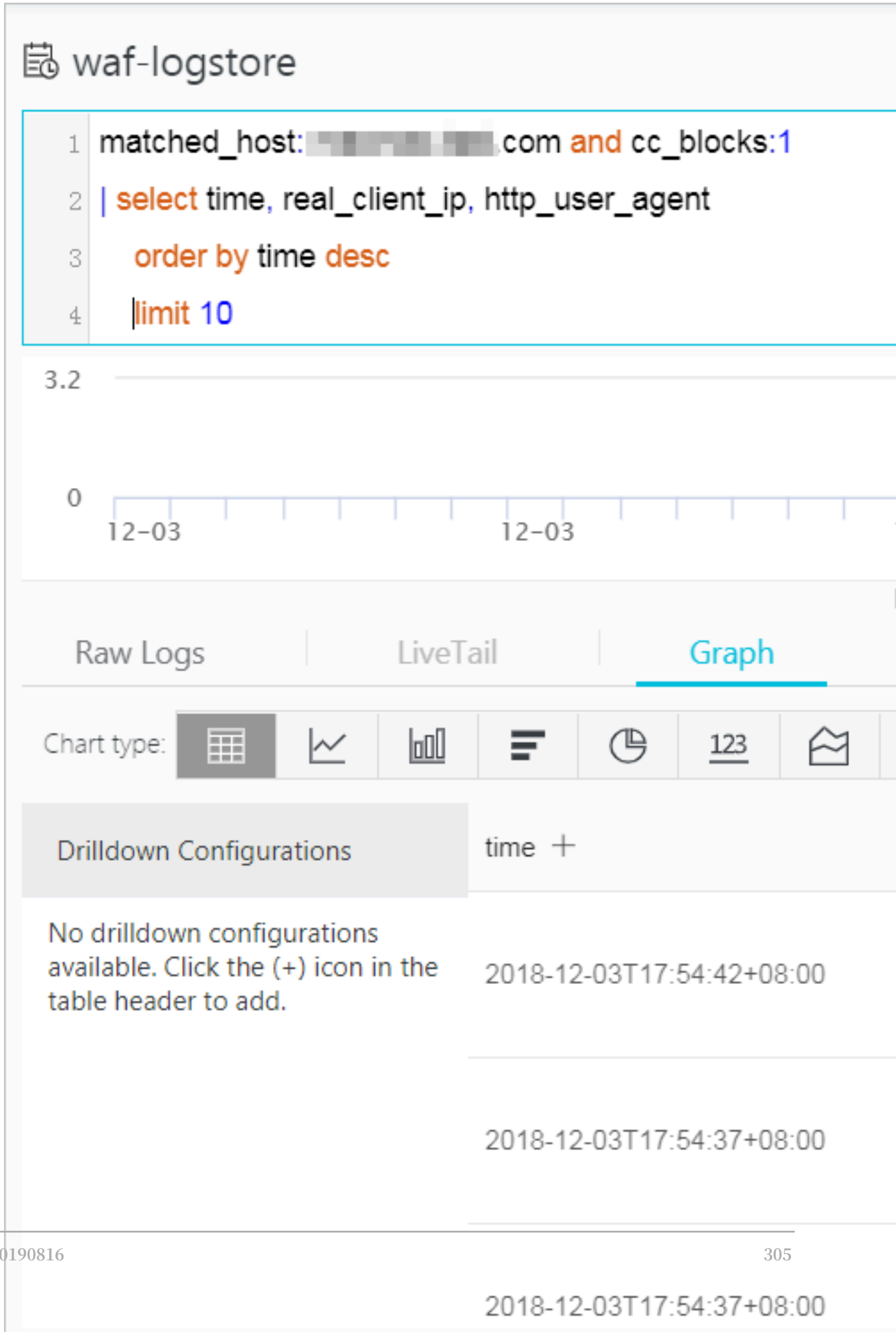
In addition, each log entry has a built-in field `__time__`, which is also used to indicate the time when the log entry is generated. This field is used for calculation when performing statistics. The format of this field is a Unix timestamp, and the value of this field indicates the number of seconds 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.

```
matched_ho st : www . aliyun . com and cc_blocks : 1
| select time , real_clien t_ip , http_user_ agent
order by time desc
```

```
limit   10
```


waf-logstore
1 matched_host: com and cc_blocks:1
2 | select time, real_client_ip, http_user_agent
3 order by time desc
4 limit 10
3.2
0
12-03
12-03
Raw Logs
LiveTail
Graph
Chart type:
123
Drilldown Configurations
time +
No drilldown configurations available. Click the (+) icon in the table header to add.
2018-12-03T17:54:42+08:00
2018-12-03T17:54:37+08:00
2018-12-03T17:54:37+08:00

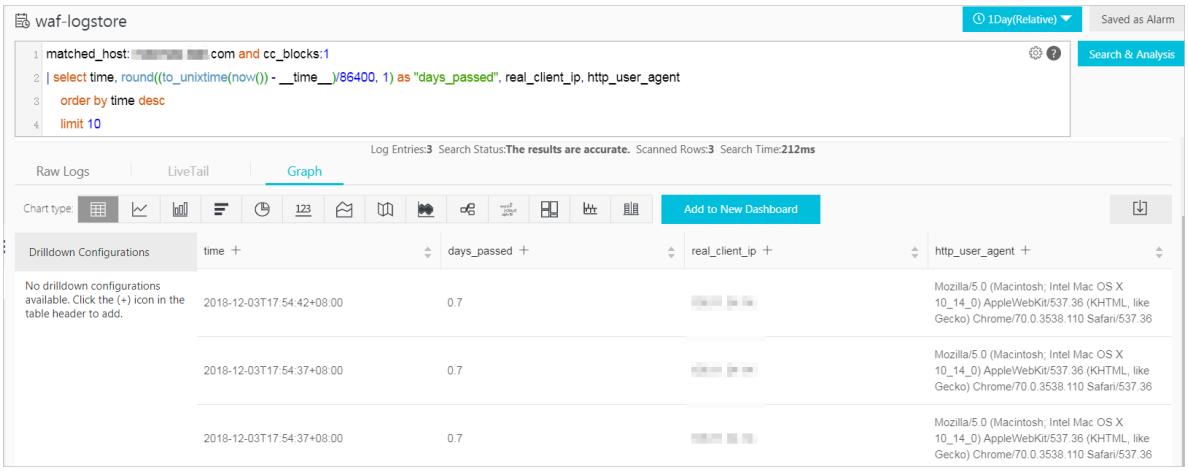
- Calculate using time.

  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.

```
matched_ho  st :  www . aliyun . com   and   cc_blocks :  1
round (( to_unixtim  e ( now ()) -  __time__ )/ 86400 ,  1 )  as  "
days_passe  d ",  real_clien  t_ip ,  http_user_  agent
      order   by   time   desc
      limit   10
```

  Note:

  In this example, `round (( to_unixtim  e ( 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_unixtim  e` . 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.

```
 matched_ho  st :  www . aliyun . com   and   cc_blocks :  1
|  select   date_trunc (' day ',  __time__ )  as   dt ,  count ( 1
 )  as   PV
      group   by   dt
```

```
        order   by   dt
```

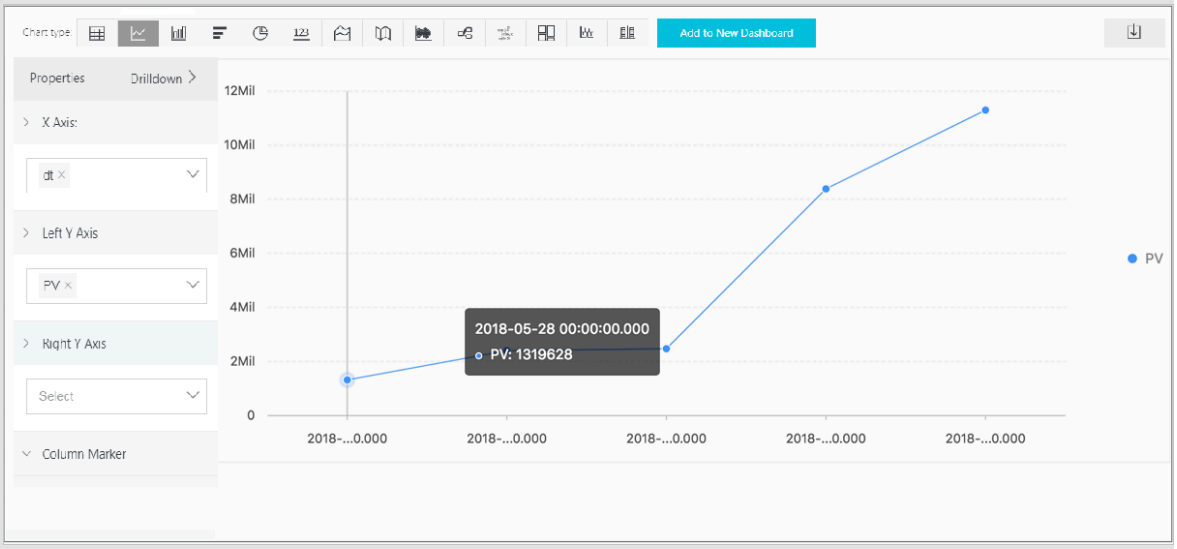
Note:

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.

dt	PV
2018-12-03 00:00:00.000	3

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_ho  st :  www . aliyun . com   and   cc_blocks :  1
|  select   from_unixt  ime ( __time__  -  __time__ %  300 )  as
 dt ,
         count ( 1 )  as   PV
```

```
group   by   dt
order   by   dt
limit   1000
```

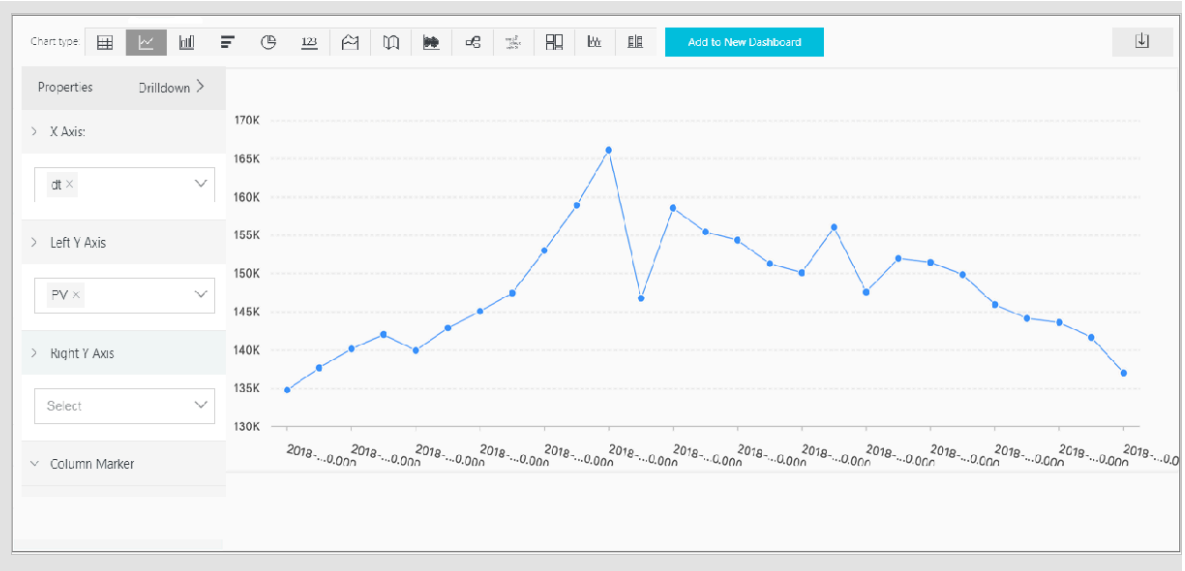
Note:

In this example, the built-in field is used for aligning the time by using the formula `__time__ - __time__ % 300`, and the `from_unixt ime` 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.

dt	PV
2018-05-31 21:30:00.000	134795
2018-05-31 21:35:00.000	137691
2018-05-31 21:40:00.000	140171
2018-05-31 21:45:00.000	142037
2018-05-31 21:50:00.000	139958
2018-05-31 21:55:00.000	142906
2018-05-31 22:00:00.000	145093
2018-05-31 22:05:00.000	147474

Note:

You can also display the results with a line graph.



The `date_parse` and `date_forma t` 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_clien t_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_add r` 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.

```
 matched_ho  st :  www . aliyun . com   and   cc_blocks :  1
|  SELECT   ip_to_coun  try ( if ( real_clien  t_ip ='-',
 remote_add  r ,  real_clien  t_ip ))  as   country ,
          count ( 1 )  as  " number   of   attacks "
          group   by   country
```

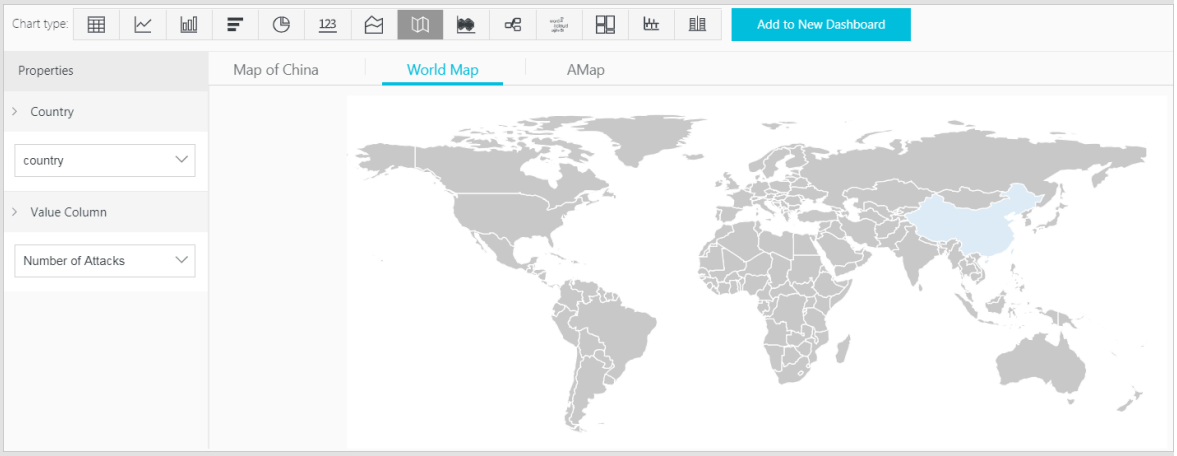
Note:

In this example, the function `if ( condition , option1 , option2 )` returns the real client IP address. If `real_clien t_ip` is `-`, the function returns the value of `remote_add r`. Otherwise, the function returns

`real_clien t_ip` . Then, use the `ip_to_coun try` to get the country information from the IP address of the client.

Note:

You can also display the results with a world map.



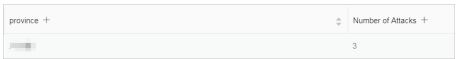
- Distribution of visitors by province

  If you want to get the distribution of visitors by province, you can use the `ip_to_prov ince` function to get the province information from the IP addresses.

```
matched_ho st : www . aliyun . com and cc_blocks : 1
| SELECT ip_to_prov ince ( if ( real_clien t_ip ='-',
remote_add r , real_clien t_ip )) as province ,
count ( 1 ) as " number of attacks "
group by province
```

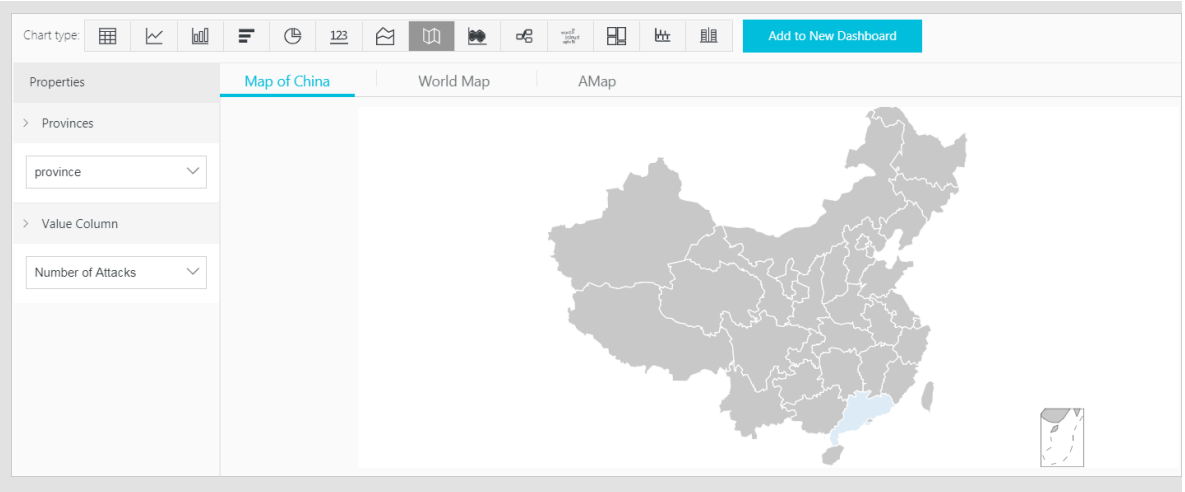
Note:

In this example, the `ip_to_prov ince` 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_ho  st :  www . aliyun . com   and   cc_blocks :  1
|  SELECT   ip_to_geo ( if ( real_clien  t_ip ='-',  remote_add  r
 ,  real_clien  t_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_prov ider` function can be used to get the IP provider name, and the `ip_to_doma in` 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.

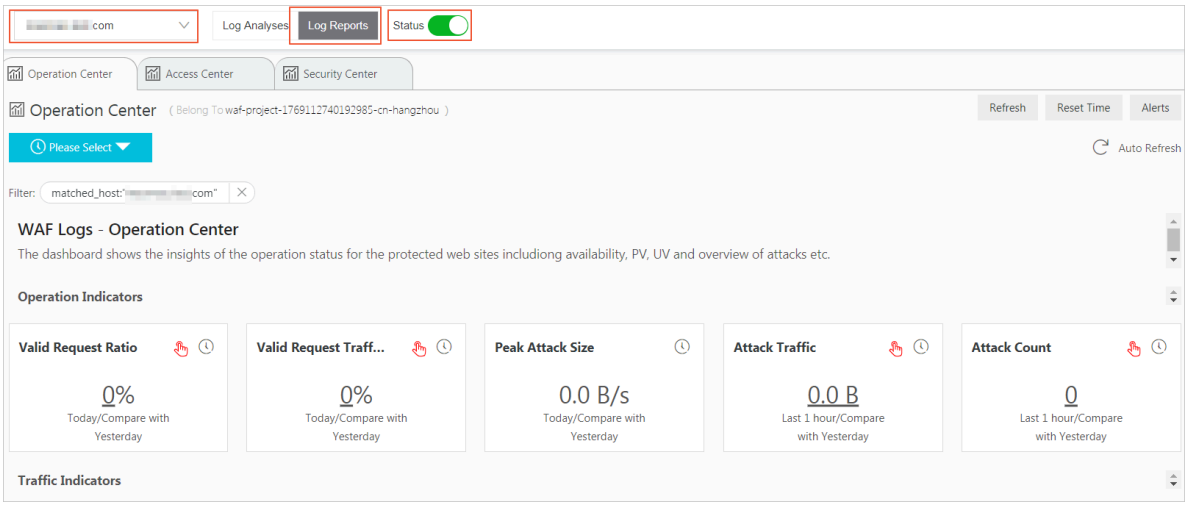
4.8.6 Log Reports

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.

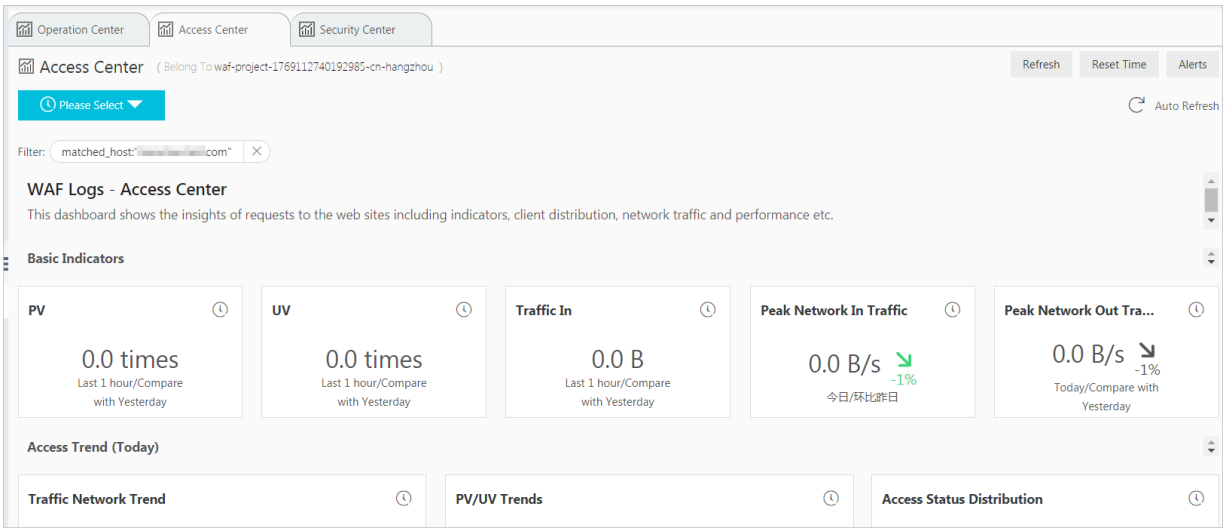
   The page that appears is integrated with the Dashboard page of Log Service. A filter is automatically added to display all log entries that are recorded for the domain you selected. In this example, the filter is `matched_ho st : www . aliyun . com` .



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.

Dashboard	Description
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, operation trends, attack overview, and other information.
Access Center	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.
Security Center	Displays basic index information of attacks, attack types, attack trend, attacker distribution, and other information.



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.

Type	Description
Number	Graphs of this type display important metrics, such as the valid request ratio and the peak of attacks.

Type	Description
Line chart and area chart	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.
Map	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.
Pie chart	Graphs of this type display a distribution, such as the distribution of attackers and the distribution of client types.
Table	Graphs of this type display a table that contains information, such as information of attackers.
Map	Graphs of this type display the geographical distribution of data.

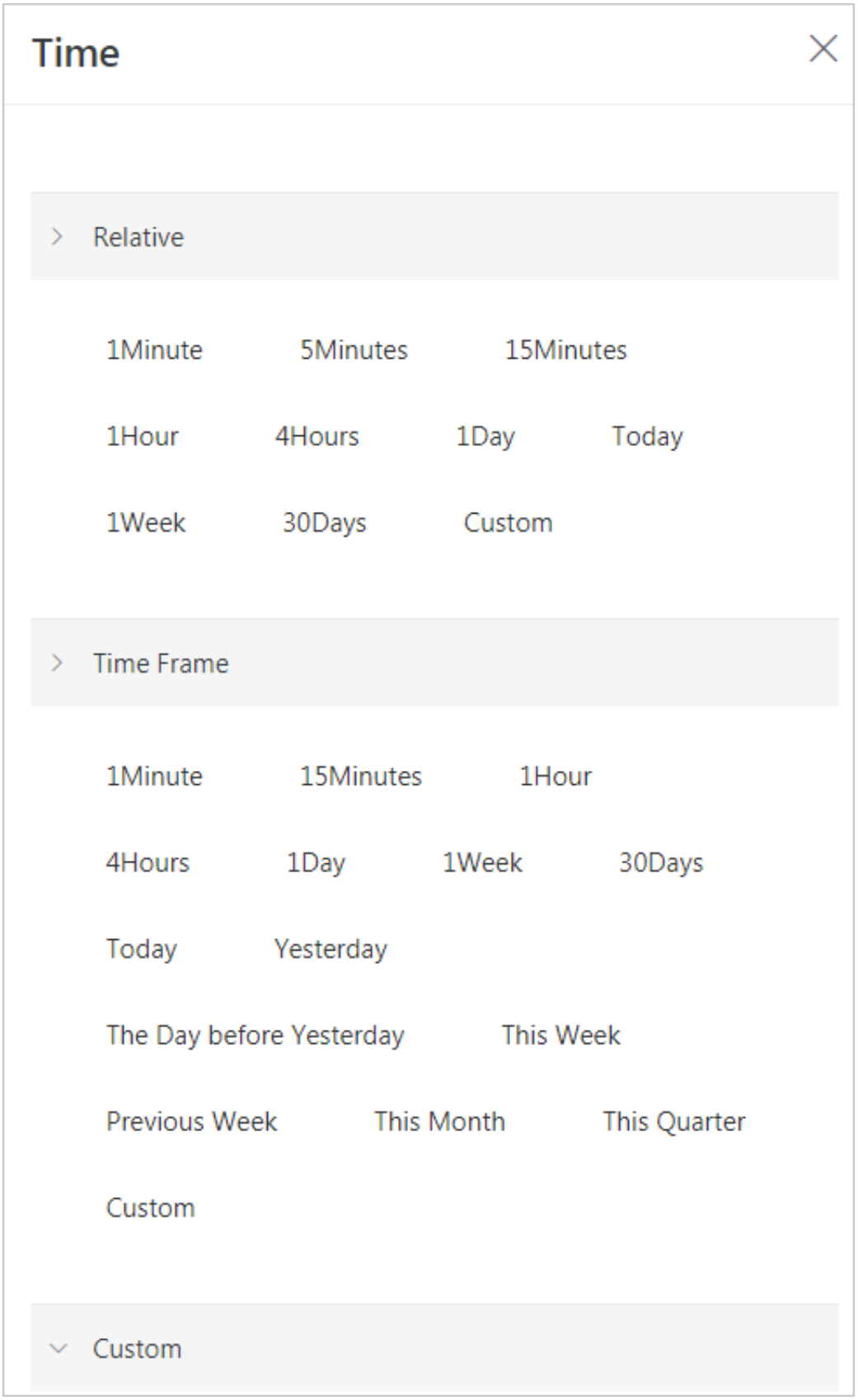
Time selector

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.

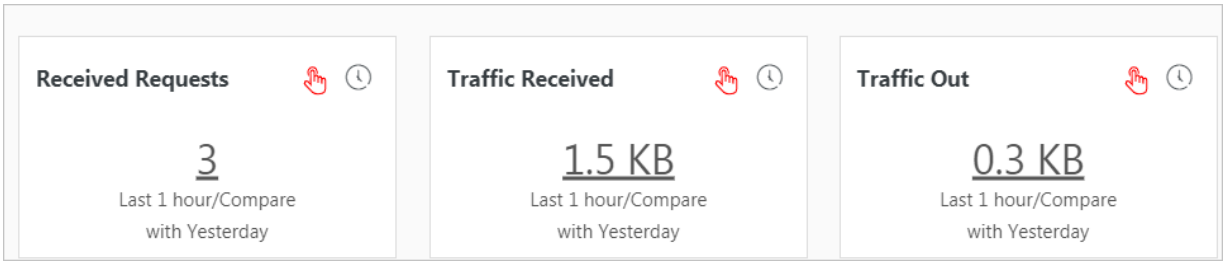
Note:

- 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
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.

Graph	Type	Default time range	Description	Example
Valid Request Ratio	Single value	Today (time frame)	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.	95%

Graph	Type	Default time range	Description	Example
Valid Request Traffic Ratio	Single value	Today (time frame)	Displays the percentage of the traffic generated by valid requests in the traffic generated by all requests.	95%
Peak Attack Size	Single value	Today (time frame)	Displays the peak of attack traffic, which is measured in Bps.	100 B/s
Attack Traffic	Single value	1 hour (relative)	Displays the total attack traffic, which is measured in B.	30 B
Attack Count	Single value	1 hour (relative)	The total number of attacks.	100
Peak Network In	Single value	Today (time frame)	Displays the peak inbound throughput , which is measured in KB/s.	100 KB/s
Peak Network Out	Single value	Today (time frame)	Displays the peak outbound throughput, which is measured in KB/s.	100 KB/s
Received Requests	Single value	1 hour (relative)	Displays the total number of valid requests.	7,800
Received traffic	Single value	1 hour (relative)	Displays the total inbound traffic that is generated by valid requests, which is measured in MB.	1.4 MB
Traffic Out	Single value	1 hour (relative)	Displays the total outbound traffic that is generated by valid requests, which is measured in MB.	3.8 MB

Graph	Type	Default time range	Description	Example
Network Traffic In And Attack	Area chart	Today (time frame)	Displays the trends of throughput generated by valid requests and attacks, which is measured in Kbit/s.	-
Request And Interception	Line chart	Today (time frame)	Displays the trends of valid requests and requests that are intercepted, which is measure in Kbit/h.	-
Access Status Distribution	Flow chart	Today (time frame)	Displays the trends of requests with different status codes (404, 304, 200, and other status codes), which is measured in Kbit/h.	-
Attack Source (World)	World map	1 hour (relative)	Displays the distribution of attackers by country.	-
Attack Source (China)	Map of China	1 Hour (Relative)	Displays the distribution of attackers in China by province.	-
Attack Type	Pie chart	1 hour (relative)	Displays the distribution of attacks by attack type.	-
Attacked Hosts	Tree map	1 hour (relative)	Displays the domains that are attacked and the number of attacks.	-

- 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.

Graph	Type	Default time range	Description	Example
PV	Single value	1 hour (relative)	Displays the total number of PV.	100,000
UV	Single value	1 hour (relative)	Displays the total number of UV.	100
Traffic In	Single value	1 hour (relative)	Displays the total inbound traffic, which is measured in MB.	300 MB
Peak Network In Traffic	Single value	Today (time frame)	Displays the peak inbound throughput , which is measured in KB/s.	0.5 KB/s
Peak Network Out Traffic	Single value	Today (time frame)	Displays the peak outbound throughput, which is measured in KB/s.	1.3 KB/s
Traffic Network Trend	Area chart	Today (time frame)	Displays the trends of inbound and outbound throughput, which are measured in KB/ s.	-
PV/UV Trends	Line chart	Today (time frame)	Displays the trends of PV and UV, which is measured in Kbit/ h.	-
Access Status Distribution	Flow chart	Today (time frame)	Displays the trends of requests with different status codes (404, 304, 200 , and other status code), which is measured in Kbit/h.	-

Graph	Type	Default time range	Description	Example
Access Source	World map	1 hour (relative)	Displays the distribution of attackers by country.	-
Traffic In Source (World)	World map	1 hour (relative)	Displays the distribution (by country) of inbound traffic from requests.	-
Traffic In Source (China)	Map of China	1 hour (relative)	Displays the distribution (by province) of inbound traffic from requests in China.	-
Access Heatmap	Amap	1 hour (relative)	Displays the heat map that indicates the source distribution of requests by geographical position.	-
Network Provider Source	Pie chart	1 hour (relative)	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.	-
Referer	Table	1 hour (relative)	Displays the first 100 referer URLs which the hosts are most often redirected from, and displays the information of hosts and redirection frequency.	-

Graph	Type	Default time range	Description	Example
Mobile Client Distribution	Pie chart	1 hour (relative)	Displays the distribution of requests from mobile clients, by client type.	-
PC Client Distribution	Pie chart	1 hour (relative)	Displays the distribution of requests from PC clients, by client type .	-
Request Content Type Distribution	Pie chart	1 hour (relative)	Displays the distribution of request sources by content type, such as HTML, form, JSON, and streaming data.	-
Accessed Sites	Tree map	1 Hour (Relative)	Displays the addresses of 30 domains that are visited most.	-
Top Clients	Table	1 hour (relative)	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.	-

Graph	Type	Default time range	Description	Example
URL With Slowest Response	Table	1 hour (relative)	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 informatio n.	-

- Security Center: Displays basic details of attacks, attack types, the attack trend, the distribution of attackers, and other information.

Chart	Type	Default time range	Description	Example
Peak Attack Size	Single value	1 hour (relative)	Displays the peak of the throughput when your website is suffering attacks, which is measured in Bps.	100 B/s
Attacked Hosts	Single value	Today (time frame)	Displays the number of domains that are attacked.	3
Source Country Of Attack	Single value	Today (time frame)	Displays the number of countries that are attack sources.	2
Attack Traffic	Single value	1 hour (relative)	Displays the total amount of traffic that is generated by attacks, which is measured in B.	1 B
Attacker UV	Single value	1 hour (relative)	Displays the number of unique clients that are attack sources.	40

Chart	Type	Default time range	Description	Example
Attack type distribution	Flow chart	Today (time frame)	Displays the distribution of attacks by attack type.	-
Intercepted Attack	Single value	1 hour (relative)	Displays the number of attacks that are intercepted by WAF.	100
HTTP flood attack Interception	Single value	1 hour (relative)	Displays the number of HTTP flood attacks that are intercepted by WAF.	10
Web Attack Interception	Single value	1 hour (relative)	Displays the number of Web applicatio n attacks that are intercepted by WAF.	80
Access Control Event	Single value	1 hour (relative)	Displays the number of requests that are intercepted by the HTTP ACL policies of WAF.	10
HTTP flood attack (World)	World map	1 hour (relative)	Displays the distribution of HTTP flood attackers by country.	-
HTTP flood attack (China)	China map	1 hour (relative)	Displays the distribution of HTTP flood attackers by province in China.	-
Web Attack (World)	World map	1 Hour (Relative)	Displays the distribution of Web application attacks by country.	-
Web Attack (China)	Map of China	1 hour (relative)	Displays the distribution of Web application attacks by province in China .	-

Chart	Type	Default time range	Description	Example
Access Control Attack (World)	World Map	1 hour (relative)	Displays the distribution by country of requests that are intercepted by the HTTP ACL policies of WAF.	-
Access Control Attack (China)	Map of China	1 Hour (Relative)	Displays the distribution by province in China of requests that are intercepted by the HTTP ACL policy of WAF.	-
Attacked Hosts	Tree map	1 hour (relative)	Displays the websites that are attacked most.	-
HTTP flood attack Strategy Distribution	Pie chart	1 hour (relative)	Displays the distribution of security policies being activated for HTTP flood attacks.	-
Web Attack Type Distribution	Pie chart	1 hour (relative)	Displays the distribution of Web attacks by attack type.	-
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.	-

Chart	Type	Default time range	Description	Example
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.	-

4.8.7 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.

Field	Description	Example
__topic__	The topic of the log entry. The value of this field is waf_access _log, which cannot be changed.	waf_access_log
acl_action	The action generated by the WAF HTTP ACL policy to the request, such as pass, drop, and captcha. Note: If the value is null or -, it indicates that the action is pass.	pass
acl_blocks	Indicates whether the request is blocked by the HTTP ACL policy. · If the value is 1, the request is blocked. · If the value is not 1, the request is passed.	1

Field	Description	Example
antibot	The type of the Anti-Bot Service protection strategy that applies, which includes: · ratelimit: Frequency control · sdk: APP protection · intelligence: Algorithmic model · acl: HTTP ACL policy · blacklist: Blacklist	ratelimit
antibot_action	The action performed by the Anti-Bot Service protection strategy, which includes: · challenge: Verifying using an embedded JavaScript script · drop: Blocking · report: Logging the access event · captcha: Verifying using a slider captcha	challenge
block_action	The type of the WAF protection that is activated, which includes: · tmd: Protection against HTTP flood attacks · waf: Protection against Web application attacks · acl: HTTP ACL policy · geo: Blocking regions · antifraud: Risk control for data · antibot: Blocking Web crawlers	tmd
body_bytes_sent	The size of the body in the access request, which is measured in Bytes.	2
cc_action	Protection strategies against HTTP flood attacks, such as none, challenge, pass, close, captcha, wait, login, and n.	close

Field	Description	Example
cc_blocks	Indicates whether the request is blocked by the CC protection. · If the value is 1, the request is blocked. · If the value is not 1, the request is passed.	1
cc_phase	The CC protection strategy that is activated, which can be seccookie, server_ip_blacklist, static_whitelist, server_header_blacklist, server_cookie_blacklist, server_args_blacklist, or qps_overmax.	server_ip_blacklist
content_type	The content type of the access request.	application/x-www-form-urlencoded
host	The source website.	api.aliyun.com
http_cookie	The client-side cookie, which is included in the request header.	k1=v1;k2=v2
http_referer	The URL information of the request source, which is included in the request header. - indicates no URL information.	http://xyz.com
http_user_agent	The User Agent field in the request header, which contains information such as the client browser and the operating system.	Dalvik/2.1.0 (Linux; U; Android 7.0; EDI-AL10 Build/HUAWEIEDISON-AL10)
http_x_forwarded_for	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.	-

Field	Description	Example
https	Indicates whether the request is an HTTPS request. · true: the request is an HTTPS request. · false: the request is an HTTP request.	true
matched_host	The matched domain name (extensive domain name) that is protected by WAF. If no domain has been matched, the value is -.	*.aliyun.com
querystring	The query string in the request.	title=tm_content%3Darticle&pid=123
real_client_ip	The real IP address of the client. If the system cannot get the real IP address, the value is -.	1.2.3.4
region	The information of the region where the WAF instance is located.	cn
remote_addr	The IP address of the client that sends the access request.	1.2.3.4
remote_port	The port of the client that sends the access request.	3242
request_length	The size of the request, measured in Bytes.	123
request_method	The HTTP request method used in the access request.	GET
request_path	The relative path of the request. The query string is not included.	/news/search.php
request_time_msec	The request time, which is measured in microseconds.	44
request_traceid	The unique ID of the access request that is recorded by WAF.	7837b**********ea1f0
server_protocol	The response protocol and the version number of the origin server.	HTTP/1.1

Field	Description	Example
status	The status of the HTTP response to the client returned by WAF.	200
time	The time when the access request occurs.	2018-05-02T16:03:59+08:00
ua_browser	The information of the browser that sends the request.	ie9
ua_browser_family	The family of the browser that the sent the request.	internet explorer
ua_browser_type	The type of the browser that the sent the request.	web_browser
ua_browser_version	The version of the browser that sends the request.	9.0
ua_device_type	The type of the client device that sends the request.	computer
ua_os	The operating system used by the client that sends the request.	windows_7
ua_os_family	The family of the operating system used by the client.	windows
upstream_addr	A list of origin addresses, separated by commas. The format of an address is `IP : Port`.	1.2.3.4:443
upstream_ip	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.	1.2.3.4
upstream_response_time	The time that the origin site takes to respond to the WAF request, which is measured in seconds. "-" indicates the timeout of the request.	0.044

Field	Description	Example
upstream_status	The response status that WAF receives from the origin server. "-" indicates that no response is received. The reason can be the response timeout, or the request being blocked by WAF.	200
user_id	Alibaba Cloud account ID.	12345678
waf_action	The action from the Web attack protection policy. · If the value is block, the attack is blocked. · If the value is bypass or other values, the attack is ignored.	block
web_attack_type	The Web attack type such as xss, code_exec, webshell, sqli, lfilei, rfilei, and other.	xss
waf_rule_id	The ID of the WAF rule that is matched.	100

4.8.8 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

4.8.9 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.

Operation	Required account type and permissions
Enable Log Service (the service remains enabled after this operation)	Alibaba Cloud account
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 `AliyunLogF ullAccess` permission · RAM user that has specific permissions
Use the log query and analysis service	· Alibaba Cloud account · RAM user that has the `AliyunLogF ullAccess` permission · RAM user that has specific permissions

Grant permissions to RAM users as required.

Scenario	Permission	Procedure
Grant permissions on all Log Service operations to a RAM user.	`AliyunLogF ullAccess`	For more information, see RAM users.

Scenario	Permission	Procedure
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.	`AliyunLogR eadOnlyAcc ess`	For more information, see RAM users.
Grant the RAM user permissions on enabling and using the WAF log query and analysis service . This RAM user is not granted other administra tive permissions on Log Service.	Custom authorization policy	For more information, see the following procedure.

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.

```
{
  " Version ": " 1 ",
  " Statement ": [
   {
      " Action ": " log : GetProject ",
      " Resource ": " acs : log :*:*: project /${ Project }",
      " Effect ": " Allow "
    },
    {
      " Action ": " log : CreateProj  ect ",
      " Resource ": " acs : log :*:*: project /*",
      " Effect ": " Allow "
    },
 {
      " Action ": " log : ListLogSto  res ",
      " Resource ": " acs : log :*:*: project /${ Project }/
 logstore /*",
```

```
      " Effect ": " Allow "
    },
    {
      " Action ": " log : CreateLogS  tore ",
      " Resource ": " acs : log :*:*: project /${ Project }/
logstore /*",
      " Effect ": " Allow "
    },
{
      " Action ": " log : GetIndex ",
      " Resource ": " acs : log :*:*: project /${ Project }/
logstore /${ Logstore }",
      " Effect ": " Allow "
    },
    {
      " Action ": " log : CreateInde  x ",
      " Resource ": " acs : log :*:*: project /${ Project }/
logstore /${ Logstore }",
      " Effect ": " Allow "
    },
    {
      " Action ": " log : UpdateInde  x ",
      " Resource ": " acs : log :*:*: project /${ Project }/
logstore /${ Logstore }",
      " Effect ": " Allow "
    },
    {
      " Action ": " log : CreateDash  board ",
      " Resource ": " acs : log :*:*: project /${ Project }/
dashboard /*",
      " Effect ": " Allow "
    },
{
      " Action ": " log : UpdateDash  board ",
      " Resource ": " acs : log :*:*: project /${ Project }/
dashboard /*",
      " Effect ": " Allow "
    },
{
      " Action ": " log : CreateSave  dSearch ",
      " Resource ": " acs : log :*:*: project /${ Project }/
savedsearc  h /*",
      " Effect ": " Allow "
    },
{
      " Action ": " log : UpdateSave  dSearch ",
      " Resource ": " acs : log :*:*: project /${ Project }/
savedsearc  h /*",
      " 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.

4.8.10 Manage log storage

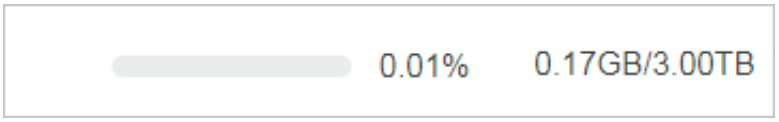
After WAF Log Service is activated, log storage is allocated for your WAF Log Service based on the specified log storage size. You can view the usage of the log storage on the Log Service page in the Web Application Firewall console.

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.

4.9 Anti-Bot logs

4.9.1 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.

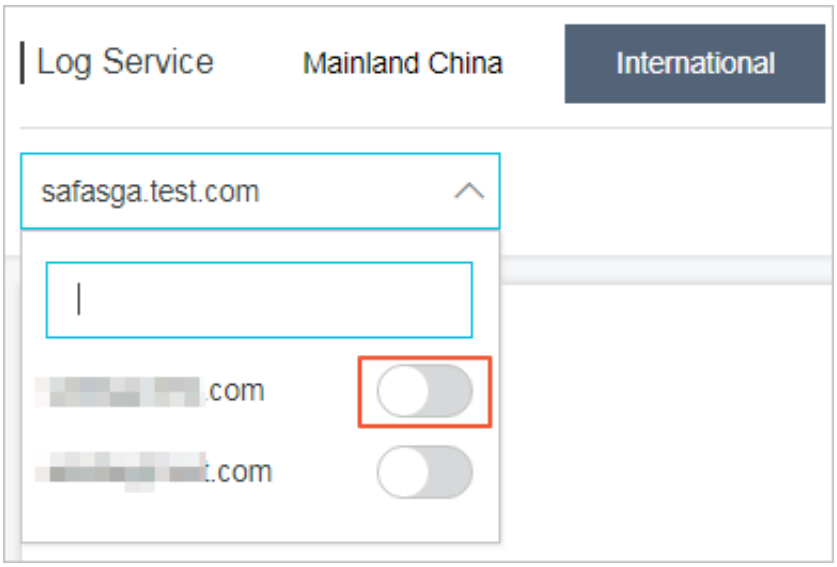
Procedure

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.

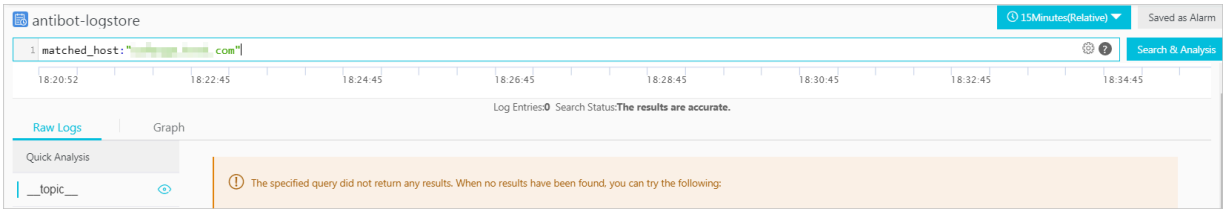
    Note:

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 automatica lly updated.
- If the RAM user requires the log query analysis function, grant the related Log Service permissions to the RAM user through RAM.

4.9.2 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.

Field	Description	Example
__topic__	The log topic. This field is invariably set to antibot_ac cess_log.	antibot_access_log
antibot	The type of the triggered Anti-Bot protection policy, including: · ratelimit: rate limiting · sdk: app protection · algorithm: algorithm pattern · intelligence: bot intelligence · acl: access control list · blacklist: blacklist	ratelimit
antibot_action	The operation specified by the Anti-Bot protection policy, including: · `challenge` : Deliver a JavaScript script for verification · `drop` : Intercept · `captcha` : Verify by dragging a slider · `report` : Monitor only	drop
antibot_rule	The ID of the triggered Anti-Bot protection rule.	5472

Field	Description	Example
antibot_verify	The result of the verification performed by Anti-Bot.  Note: This value is recorded when the antibot_action field is set to challenge or captcha. · challenge_fail: JavaScript verification fails. · challenge_pass: JavaScript verification is passed. · captcha_fail: Slide captcha verification fails. · captcha_pass: Slide captcha verification is passed.	challenge_fail
block_action	The type of the bot protection that is triggered. The value is invariably set to `antibot` .	antibot
body_bytes_sent	The size of HTTP body (in byte) sent to the client.	2
content_type	The content type of the access request.	application/x-www-form-urlencoded
host	The source website.	api.aliyun.com
http_cookie	The cookie information about the access client, which is included in the access request header.	k1=v1;k2=v2
http_referer	The source URL of the access request, which is included in the access request header. `-` is displayed if no source URL is available.	http://xyz.com
http_user_agent	The User Agent field in the access request header, which typically includes the web browser identifier and operating system identifier of the source client.	Dalvik/2.1.0 (Linux; U; Android 7.0; EDI-AL10 Build/HUAWEIEDISON-AL10)

Field	Description	Example
http_x_forwarded_for	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.	-
https	Whether the access request is an HTTPS request. Valid values: · true: The access request is an HTTPS request. · false: The access request is an HTTP request.	true
matched_host	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.	*.aliyun.com
real_client_ip	The actual IP address of the access client. - is displayed if no actual IP address is retrieved.	1.2.3.4
region	The information about the region where the Anti-Bot instance is located.	cn
remote_addr	The IP address of the client that initiates the access request.	1.2.3.4
remote_port	The port of the client that initiates the access request.	23713
request_length	The length of the access request. Unit: bytes.	123
request_method	The HTTP request method of the access request.	GET
request_path	The relative path of the request (excluding the query string).	/news/search.php
request_time_msec	The duration of the access request. Unit: milliseconds.	44

Field	Description	Example
request_traceid	The unique ID of the access request.	7837b117154103869434 37009ea1f0
server_protocol	The protocol and version of the response returned by the origin server.	HTTP/1.1
status	The status of the HTTP response that Anti-Bot returns to the client.	200
time	The occurrence time of the access request.	2018-05-02T16:03:59+08:00
ua_browser	The information about the web browser that initiates the access request.	ie9
ua_browser_family	The family of the web browser that initiates the access request.	internet explorer
ua_browser_type	The type of the web browser that initiates the access request.	web_browser
ua_browser_version	The version of the web browser that initiates the access request.	9.0
ua_device_type	The device type of the client that initiates the access request.	computer
ua_os	The operating system of the client that initiates the access request.	windows_7
ua_os_family	The operating system family of the client that initiates the access request.	windows
upstream_addr	The origin address list of Anti-Bot in the format of `IP address : Port`. Separate multiple IP addresses with commas (,).	1.2.3.4:443

Field	Description	Example
upstream_ip	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.	1.2.3.4
upstream_response_time	The time for the origin server to respond to an Anti-Bot request. Unit: seconds. The response times out if "-" is returned.	0.044
upstream_status	The status of the response that the origin server returns to Anti-Bot. No response is available if "-" is returned. For example, the request is intercepted by Anti-Bot, or the response returned by the origin server times out.	200
user_id	AliUID of the Alibaba Cloud account.	12345678
wxbb_action	If the protection type of Anti-Bot is app protection, the following actions are supported: · `close` : intercepts requests. That is, the antibot_action field is set to `drop` . · `test` : only monitors requests. That is, the antibot_action field is set to `report` . Note: This field is set to – if SDK protection is not configured.	close
wxbb_invalid_wua	For more information about app protection, consult your technical engineer.	valid wua

4.10 ActionTrail access logs

4.10.1 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 #unique_239.
- 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_22.

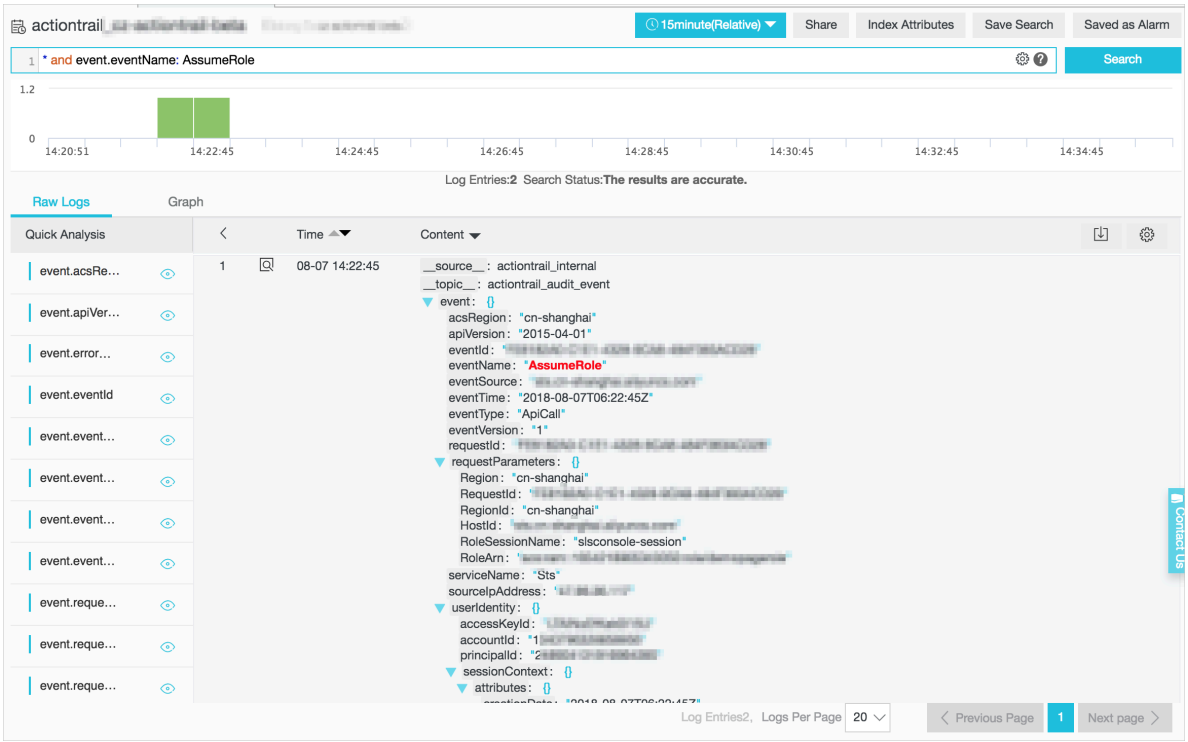
Application scenarios

- 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 4-48: View the ECS release operation log

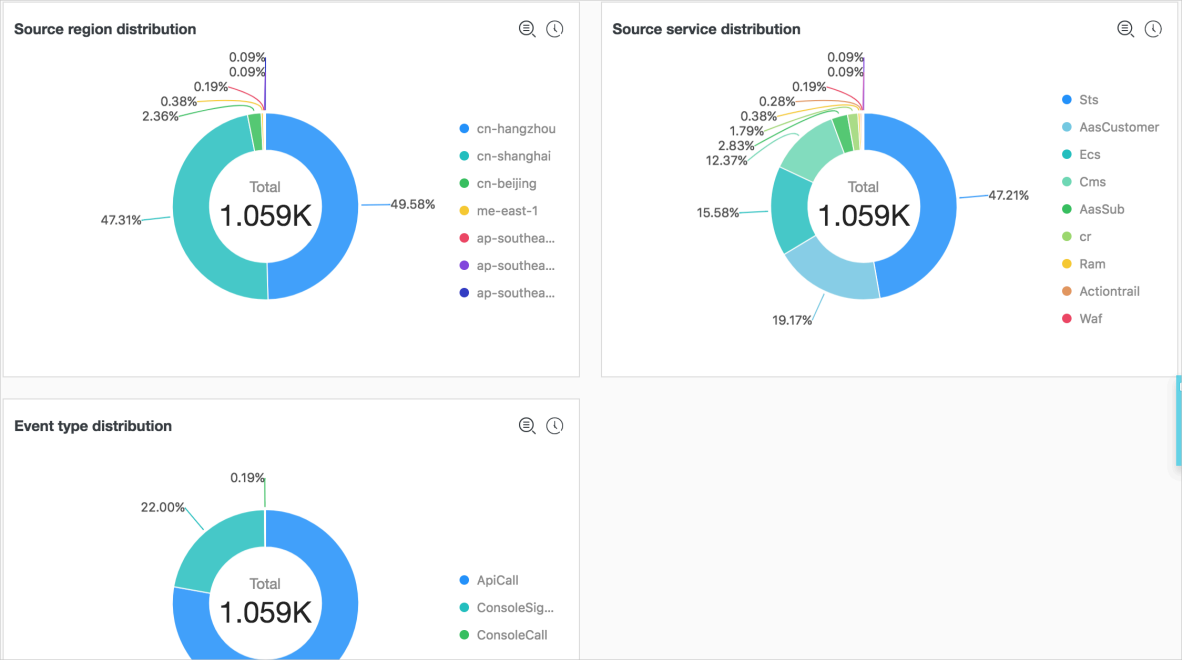


- Distribution and source tracking of important resource operations

  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 4-49: View the distribution of RDS deletion



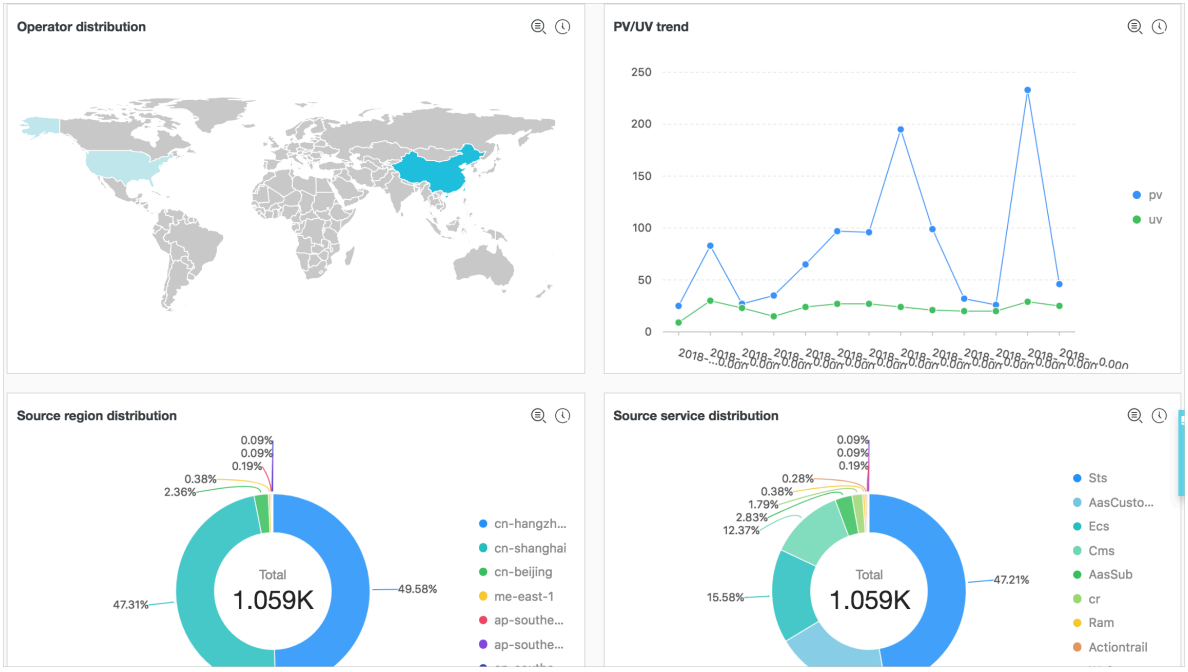
- Resource operation distribution view

  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 4-50: Trends of failed operations



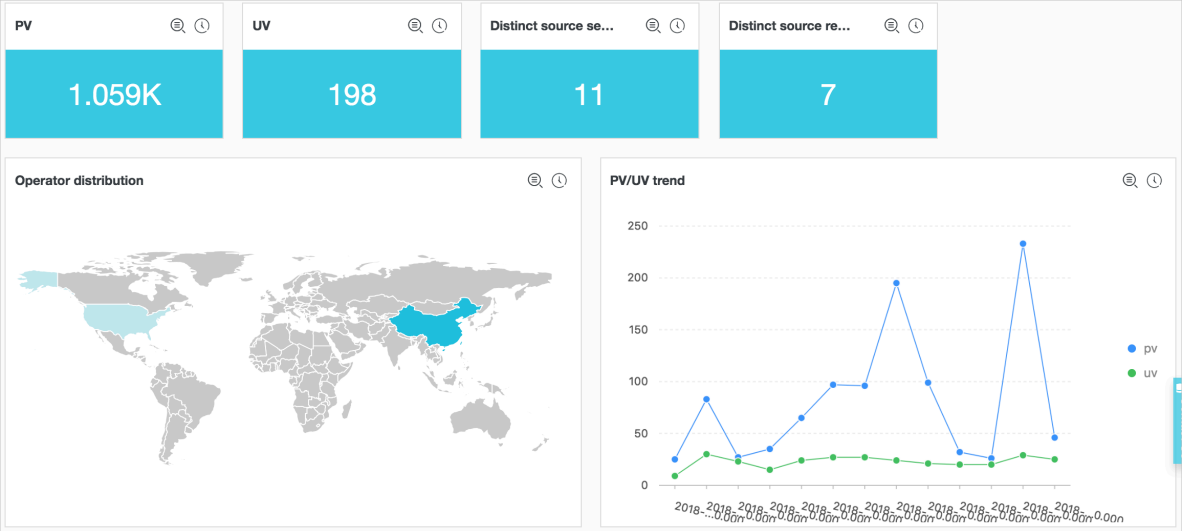
- Real-time analysis of operation data

  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 4-51: Frequency distribution of operators from network operators



4.10.2 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.

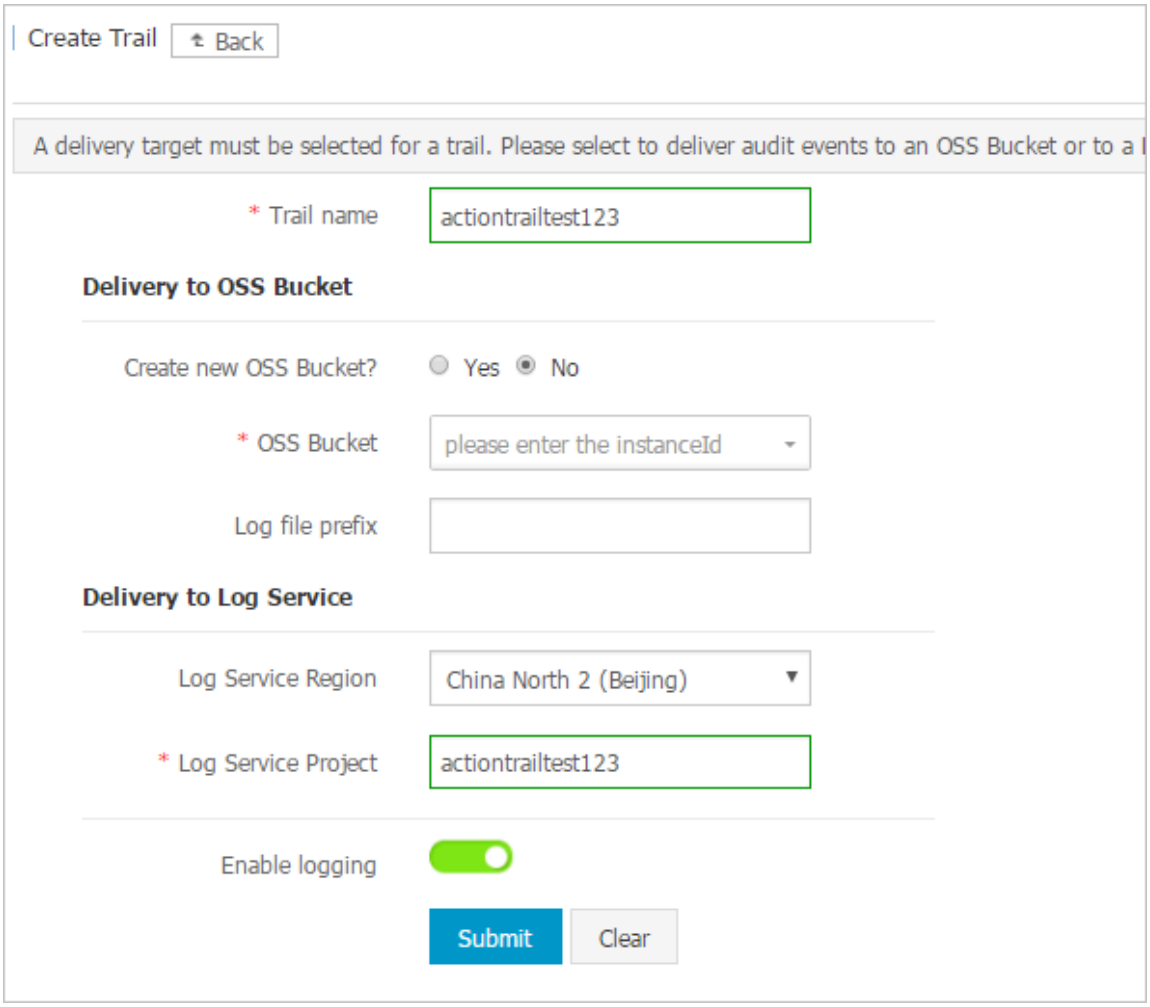
4. Configure trail parameters.
   a. Enter Trail name.
   b. Deliver audit events to an OSS Bucket (optional).

      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 4-52: Configure trail parameters.

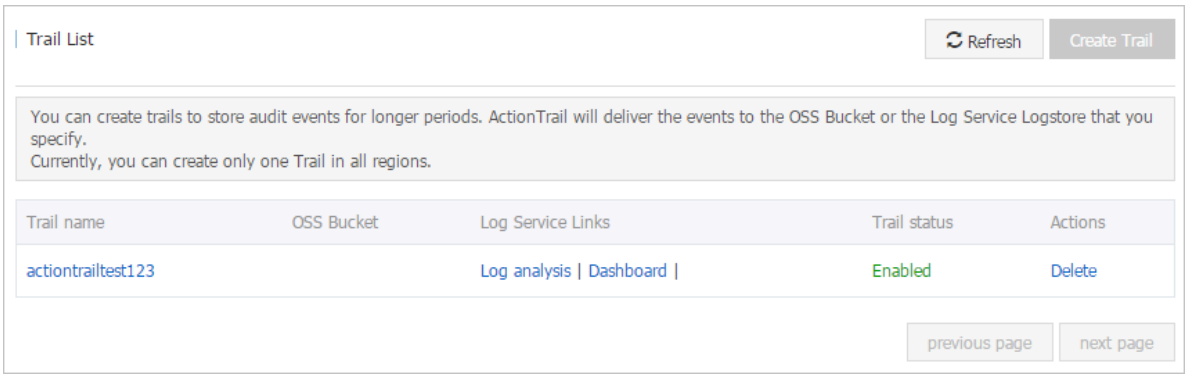


5. Click Submit to complete the configuration.

   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 4-53: 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 Logstoree 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_22.

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_241.
- 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_242.

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

Default configuration item	Configuration content
Project	A project that you select or customize when you create the trail.
Logstore	By default, Logstore is created. The Logstore name is `actiontrail_Trail name`. All logs of ActionTrail are saved in this Logstore.
Region	A region that you select when you create the trail.
Shard	By default, two shards are created and the Auto Split Shard feature is enabled.
Log storage time	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_243.
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

Field name	Name	Example
__topic__	Log topic.	This field is fixed at `actiontrai l_audit_ev ent`
event	Event body, which is in the JSON format. The content of the event body varies with the event.	event example
event.eventId	The ID of the event, which uniquely indicates the event.	07F1234-3E1D-4BFF-AC6C-12345678
event.eventName	Event name.	CreateVSwitch

Field name	Name	Example
event.eventSource	The source of the event.	http://account.aliyun.com:443/login/login_aliyun.htm
event.eventType	Event type.	ApiCallApicall
event.eventVersionEvent.eventversion	The version of the data format of ActionTrail, which is currently fixed to 1.	1
event.acsRegion	The region where the event is located.	cn-hangzhou
event.requestId	The request ID of the cloud service operation.	07F1234-3E1D-4BFF-AC6C-12345678
event.apiVersion	The version of the related API.	2017-12-04
event.errorMessage	The error message of an event failure.	unknown confidential
event.serviceName	The event-related service name.	Ecs
event.sourceIpAddress	The Source IP associated with the event.	1.2.3.4
event.userAgent	The event-related client agent.	Mozilla/5.0 (...)
event.requestParameters.HostId	The host ID in the request-related parameter.	ecs.cn-hangzhou.aliyuncs.com
event.requestParameters.Name	The name in the request-related parameter.	ecs-test
event.requestParameters.Region	The domain in the request-related parameter.	cn-hangzhou
event.userIdentity.accessKeyId	The AccessKey ID used by the request.	25 **********
event.userIdentity.accountId	The ID of the account requested.	123456
event.userIdentity.principalId	The voucher ID of the account requested.	123456
event.userIdentity.type	The type of account requested.	root-account

Field name	Name	Example
event.userIdentity.userName	The name of account requested.	root

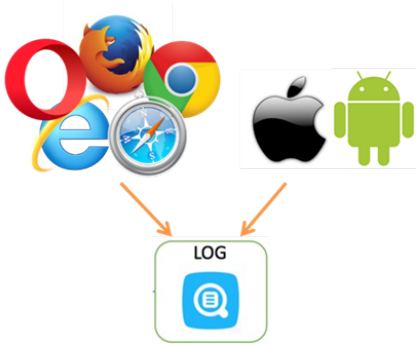
event example

```
{
  " acsRegion ": " cn - hangzhou ",
  " additional  EventData ": {
    " isMFACheck  ed ": " false ",
    " loginAccou  nt ": " test1234 @ aliyun . com "
  },
  " eventId ": " 7be1e173 - 1234 - 44a1 - b135 - 1234 ",
  " eventName ": " ConsoleSig  nin ",
  " eventSourc  e ": " http :// account . aliyun . com : 443 / login /
 login_aliy  un . htm ",
  " eventTime ": " 2018 - 07 - 12T06 : 14 : 50Z ",
  " eventType ": " ConsoleSig  nin ",
  " eventVersi  on ": " 1 ",
  " requestId ": " 7be1e173 - 1234 - 44a1 - b135 - 1234 ",
  " serviceNam  e ": " AasCustome  r ",
  " sourceIpAd  dress ": " 42 . 120 . 75 . 137 ",
  " userAgent ": " Mozilla / 5 . 0  ( Macintosh ;  Intel   Mac   OS
 X   10_13_6 )  AppleWebKi  t / 537 . 36  ( KHTML ,  like   Gecko )
 Chrome / 67 . 0 . 3396 . 99   Safari / 537 . 36 ",
  " userIdenti  ty ": {
    " accessKeyI  d ": " 25 ***************",
    " accountId ": " 1234 ",
    " principalI  d ": " 1234 ",
    " type ": " root - account ",
    " userName ": " root "
  }
}
```

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

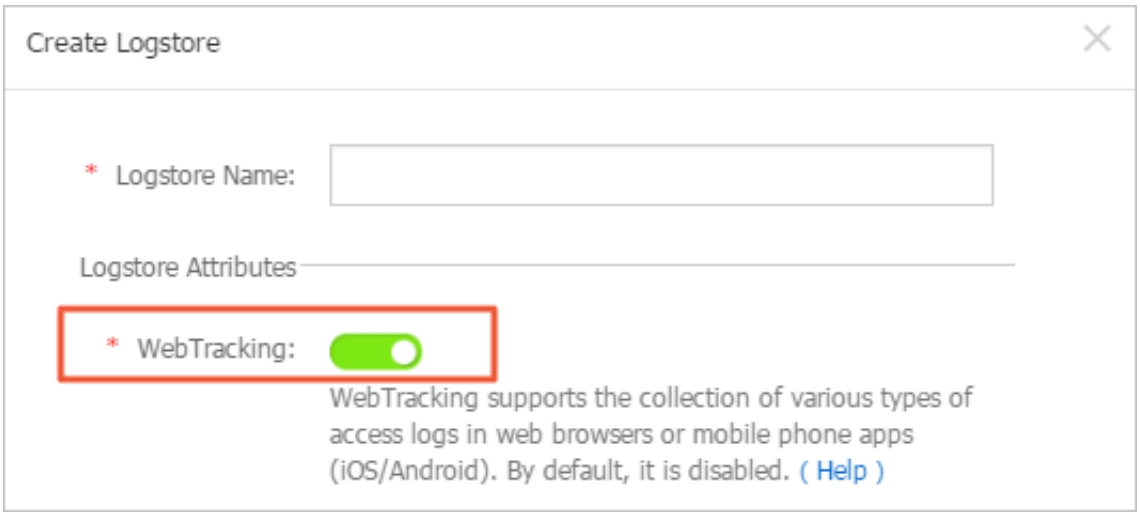
- Using Web Tracking means that this Logstore enables the anonymous write permission of the Internet without valid authentication, and dirty data may be generated.
- Only Get requests are supported. A request body exceeding 16 KB cannot be uploaded.

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

  Java SDK:

```
import com . aliyun . openservic es . log . Client ;
import com . aliyun . openservic es . log . common . LogStore ;
import com . aliyun . openservic es . log . exception .
LogExcepti on ;
public class WebTrackin g {
  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 . GetLogStor e ( project ,
logStore ). GetLogStor e ();
         client . UpdateLogS tore ( project , new LogStore (
logStore , logSt . GetTtl (), logSt . GetShardCo unt (), true
));
        // Disable the Web Tracking function .
```

```
            // client . UpdateLogS  tore ( project ,  new   LogStore (
 logStore ,  logSt . GetTtl (),  logSt . GetShardCo  unt (),  false
 ));
            // Create   a   Logstore   that   supports   the   Web
 Tracking   function .
            // client . UpdateLogS  tore ( project ,  new   LogStore (
 logStore ,  1 ,  1 ,  true ));
        }
         catch  ( LogExcepti  on   e ){
             e . printStack  Trace ();
        }
    }
}
```

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.

     ```
     < 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:

```
this . httpReques  t_ . open (" GET ",  url ,  true )
```

Replace the last parameter to send requests synchronously:

```
this . httpReques  t_ . open (" GET ",  url ,  false )
```

2. **Create a Tracker object.**

```
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:

Field	Definition
`${ host }`	The domain name of the region where your Log Service is located.
`${ project }`	The name of the project created in Log Service.
`${ logstore }`	The name of the Logstore with the Web Tracking function enabled under `${ project }`.

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.

Field	Definition
`${ project }`	The name of the project created in Log Service.
`${ host }`	The domain name of the region where your Log Service is located.
`${ logstore }`	The name of the Logstore with the Web Tracking function enabled under `${ project }`.
`APIVersion = 0 . 6 . 0`	The reserved field, which is required.
`__topic__ = yourtopic`	Specify the log topic, reserved fields (optional).
`key1 = val1 , key2 = val2`	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.

- Use the HTML IMG tag

```
< 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.

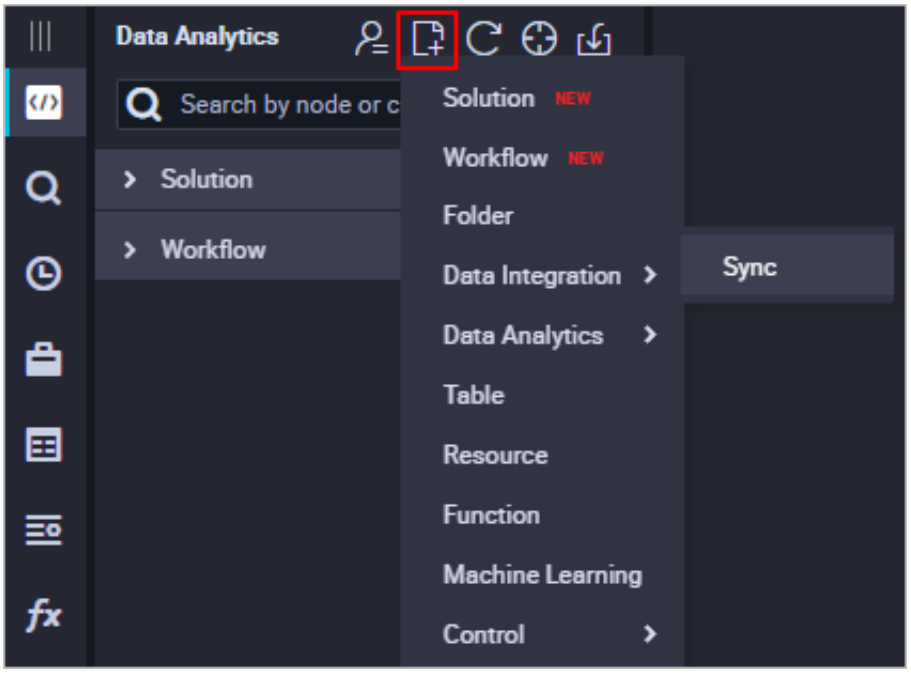
Procedure

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_248/unique_248_Connect_42_section_nkh_hnf_vdb.

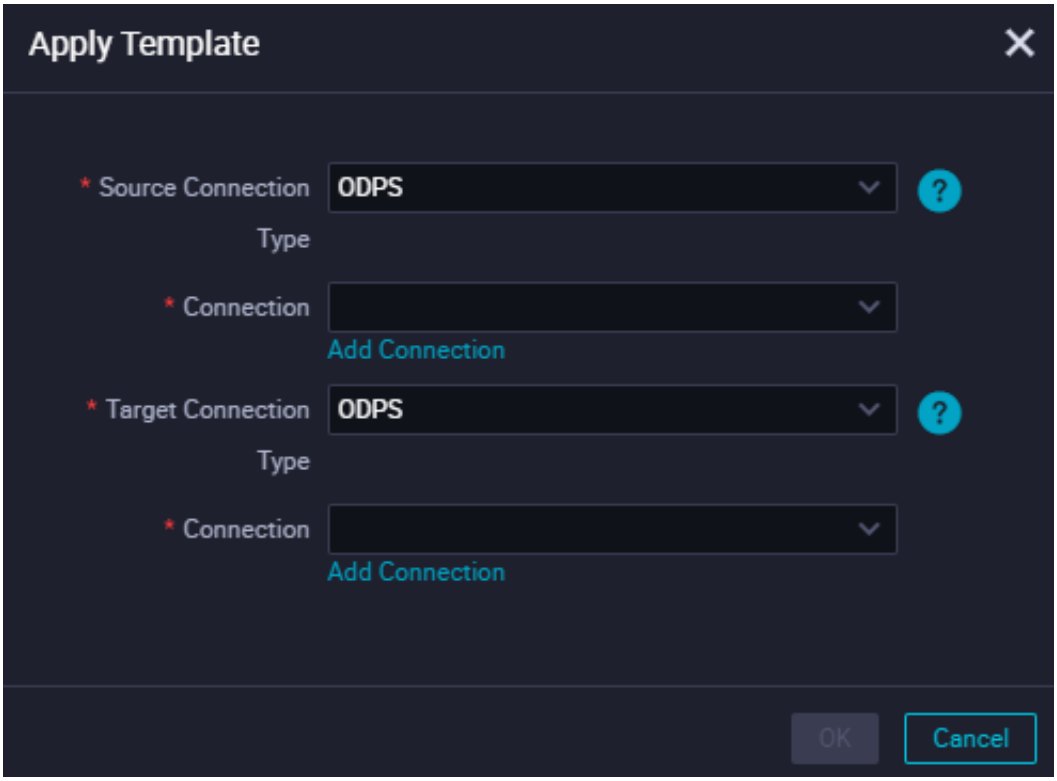
2. Create a synchronization task in script mode.
   a. Click Sync Tasks in the left-side navigation pane, and click Script Mode to configure the synchronization task.

Figure 5-1: Script Mode



b. Specify the parameters in the import template.

Figure 5-2: Import template



Parameter	Description
Source type	Select `ODPS` as the type of your data source.
data sources	The name of your data source. You can also click New Source to create a data source.
Type of objective	Select `LogHub` as the type of the shipping destination.
data sources	The name of the shipping destination. Select the LogHub data destination created in step 1, or click New Source to create a data destination.

Then, click confirmation to configure the synchronization task.

c. Enter your configuration.

The example is as follows:

```
{
  " type ": " job ",
```

```
  " version ": " 1 . 0 ",
  " configurat  ion ": {
    " setting ": {
      " errorLimit ": {
        " record ": " 0 "
      },
      " speed ": {
        " mbps ": " 1 ",
        " concurrent ":  1 ,
        " dmu ":  1 ,
        " throttle ":  false
      }
    },
    " reader ": {
      " plugin ": " odps ",
      " parameter ": {
              " accessKey ":"*****",
          " accessId ":"*****",
          " column ":["*"],
          " isCompress ":" false ",
          " partition ":[" pt = 20161226 "],
          " project ":" aliyun_acc  ount ",
          " table ":" ak_biz_log  _detail "
      }
    },
    " 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 "
      }
    }
  }
}
```

Parameter	Required	Description
endpoint	Yes	The endpoint of Log Service. For more information, see #unique_17.

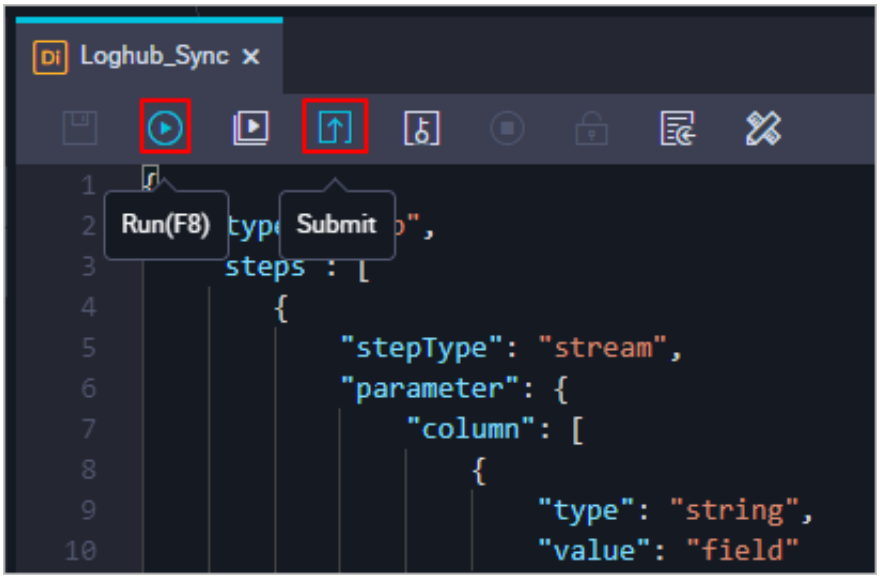
Parameter	Required	Description
accessKeyId	Yes	The AccessKeyId of your Alibaba Cloud account or RAM user.
accessKeySecret	Yes	The AccessKeyId of your Alibaba Cloud account or RAM user.
project	Yes	The name of the destination project in Log Service.
logstore	Yes	The name of the destination Logstore in Log Service.
topic	No	The field in MaxCompute that you specify as the topic field in Log Service. It is an empty string by default.
batchSize	No	The number of entries that LogHub Writer transfers at a time. It is 1024 by default.
column	Yes	The column name in each entry.  Note: The columns that are not specified in the column parameter are dirty data.
time	No	The name of the time field.  Note: If the time field is not specified, the system time is used as the log time by default.

Parameter	Required	Description
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 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.

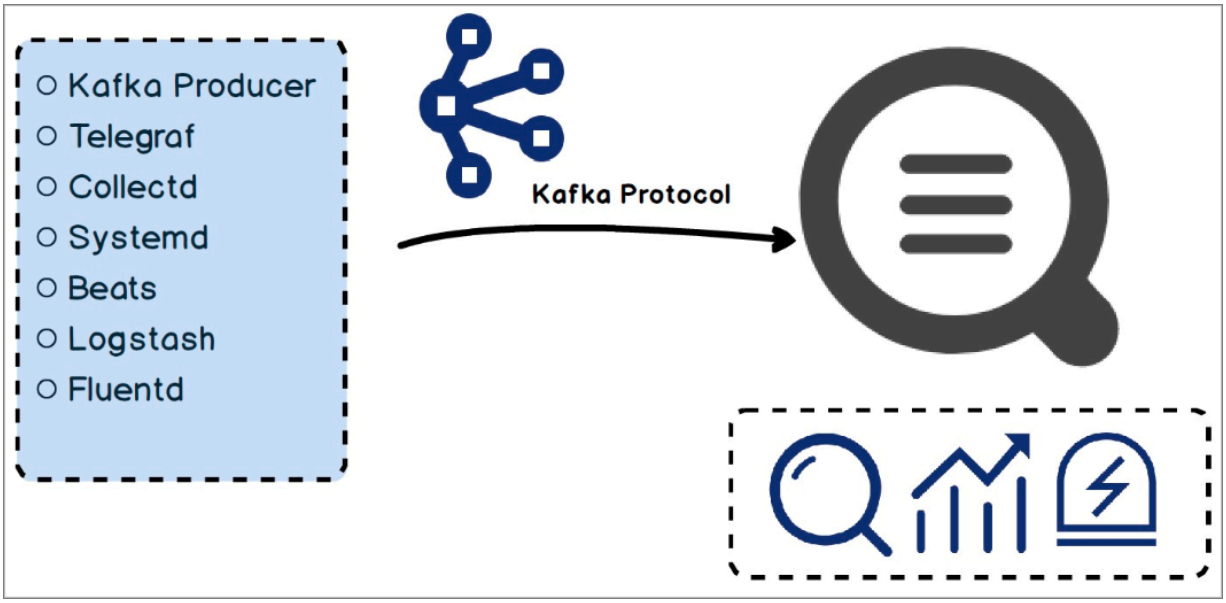
Data types

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	Data type imported to LogHub
Long	String
Double	String
String	String
Data	String
Boolean	String
Bytes	String

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.

Parameter	Description	Example
Connection protocol	The connection protocol for secure data transmission. You must use SASL_SSL.	SASL_SSL
hosts	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_28.	· cn-hangzhou-intranet.log.aliyuncs.com:10011 · cn-hangzhou.log.aliyuncs.com:10012
topic	The mapped Logstore name in Log Service. You must create a Logstore in advance.	test-logstore-1
username	The mapped project name in Log Service.	<yourusername>

Parameter	Description	Example
password	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.	<yourpassword>
Certificate	The directory of the certificate. Each domain name in Log Service has a CA certificate. You only need to use the default root certificate.	/etc/ssl/certs/ca-bundle.crt

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.

Error code	Description	Solution
NetworkException	The error message returned because a network error has occurred.	Wait for 1 second and try again.
TopicAuthorizationException	The error message returned because the authentication fails. Generally, your AccessKey is invalid or has no	Enter a valid AccessKey and ensure that it has the required write permission.

Error code	Description	Solution
	permission to write data into the corresponding project or Logstore.	
UnknownTopicOrPartitionException	The error message returned because either of the following errors has occurred: · The corresponding project or Logstore does not exist. · The region where the project is located is different from the region indicated by the endpoint that you entered.	1. Create a project and a Logstore in advance. 2. Ensure that the region where the project is located is the same as the region indicated by the endpoint that you entered.
KafkaStorageException	The error message returned because a server error has occurred.	Wait for 1 second and try again.

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 `< yourAccess KeyId >`, and the AccessKey Secret is `< yourAccess KeySecret >`.

· Example 1: Use Beats software to write data into Log Service

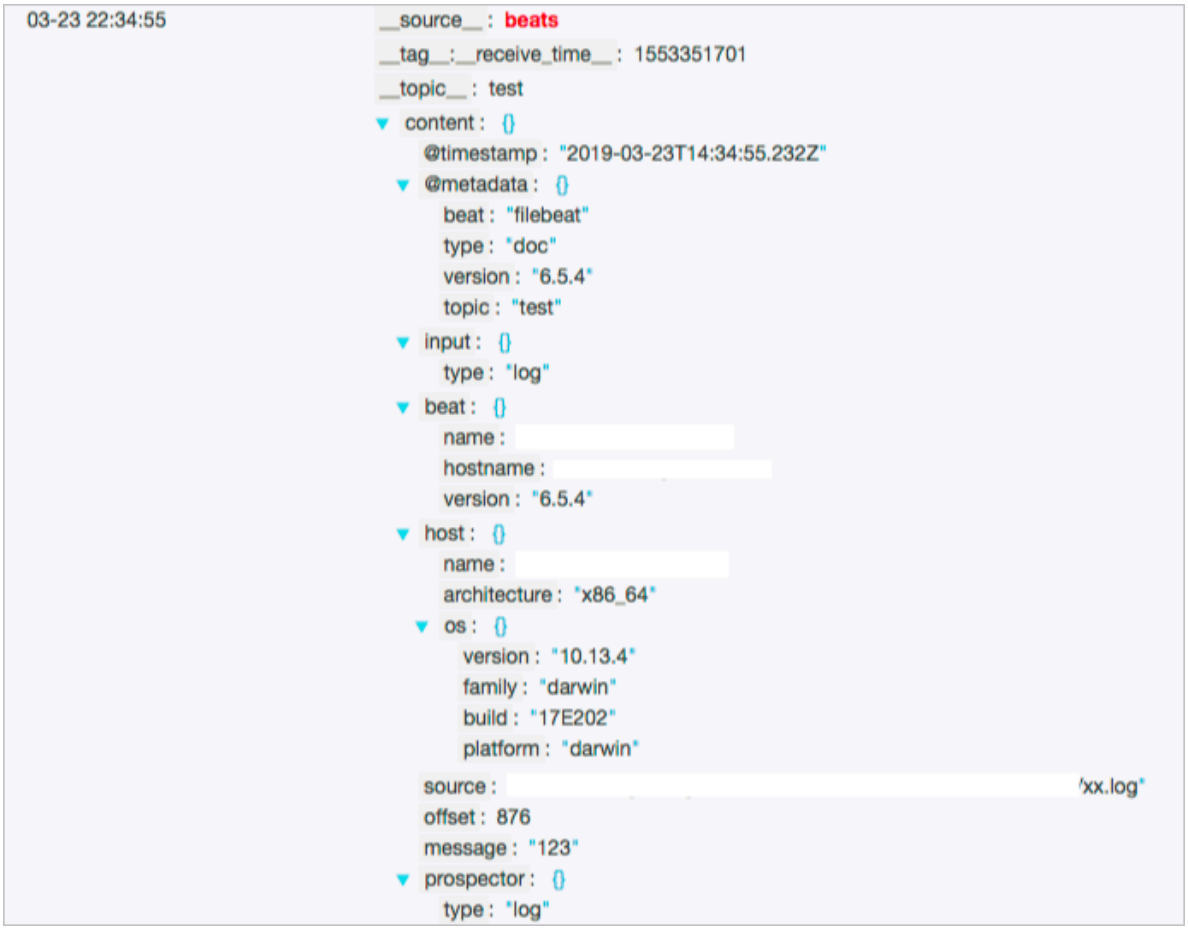
You can export the collected data to Kafka by using Beats software such as Metricbeat, Packetbeat, Winlogbeat, Auditbeat, Filebeat, and Heartbeat. For more information, see Configure the Kafka output. The sample code is as follows:

```
output . kafka :
 #  initial   brokers   for   reading   cluster   metadata
```

```
  hosts : [" cn - hangzhou . log . aliyuncs . com : 10012 "]
  username : "< youruserna  me >"
  password : "< yourpasswo  rd >"
  ssl . certificat  e_authorit  ies :
#  message   topic   selection  +  partitioni  ng
  topic : ' test - logstore - 1 '
  partition . round_robi  n :
    reachable_  only :  false

  required_a  cks :  1
  compressio  n :  gzip
  max_messag  e_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_250. The following figure shows a log sample.



- 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:

```
< Plugin write_kafk a >
  Property " metadata . broker . list " " cn - hangzhou . log .
aliyuncs . com : 10012 "
  Property " security . protocol " " sasl_ssl "
  Property " sasl . mechanism " " PLAIN "
  Property " sasl . username " "< youruserna me >"
  Property " sasl . password " "< yourpasswo rd >"
  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.

- 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:

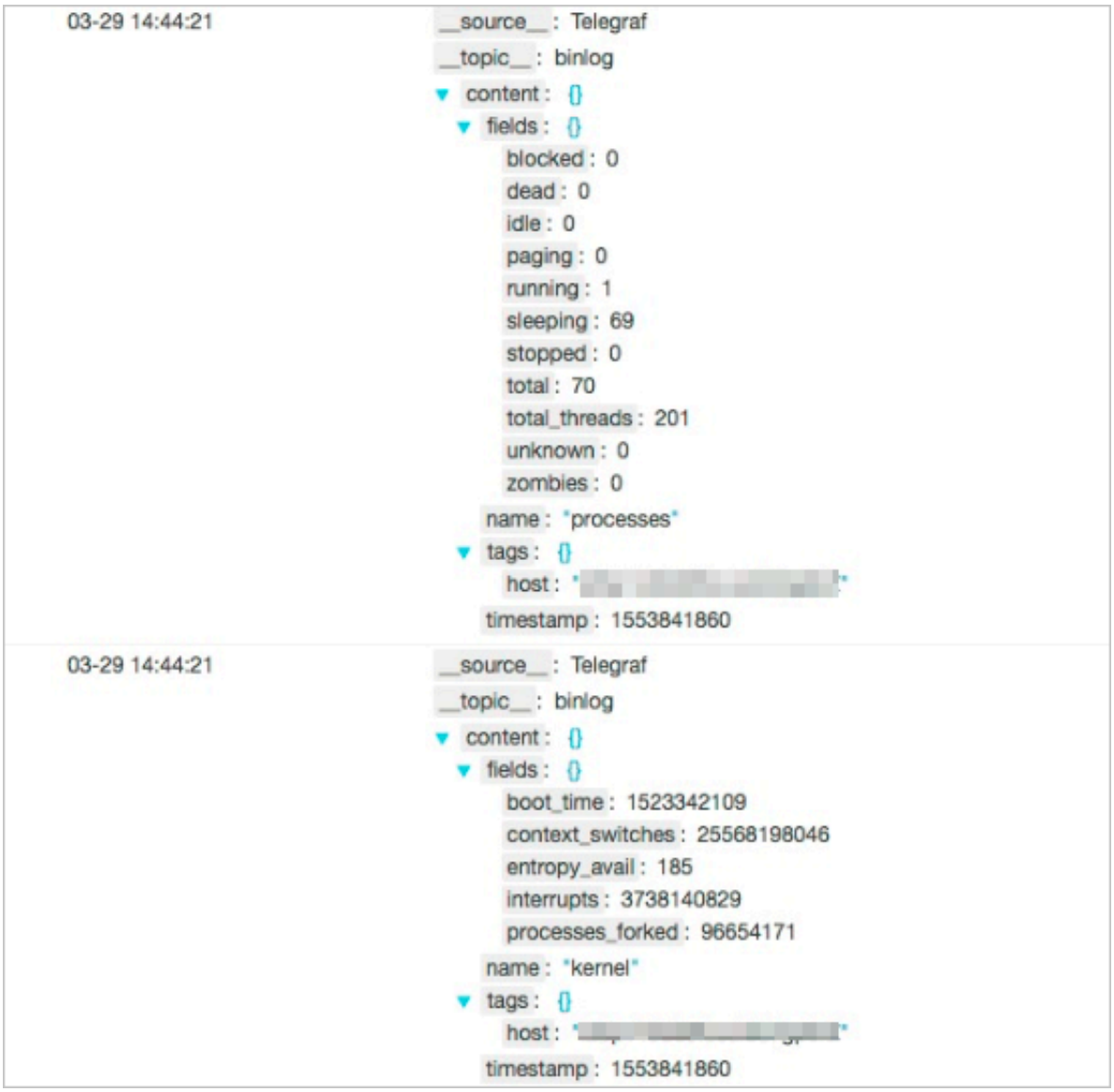
```
[[ outputs . kafka ]]
  ##  URLs   of   kafka   brokers
   brokers  = [" cn - hangzhou . log . aliyuncs . com : 10012 "]
  ##  Kafka   topic   for   producer   messages
   topic  = " test - logstore - 1 "
   routing_ke  y  = " content "
  ##  Compressio  nCodec   represents   the   various   compressio
 n   codecs   recognized   by
  ##  Kafka   in   messages .
  ##  0  :  No   compressio  n
  ##  1  :  Gzip   compressio  n
  ##  2  :  Snappy   compressio  n
  ##  3  :  LZ4   compressio  n
   compressio  n_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   verificati  on
  #  insecure_s  kip_verify  =  false
  ##  Optional   SASL   Config
   sasl_usern  ame  = "< youruserna  me >"
   sasl_passw  ord  = "< yourpasswo  rd >"
  ##  Data   format   to   output .
  ##  https :// github . com / influxdata / telegraf / blob / master
 / docs / DATA_FORMA  TS_OUTPUT . md
   data_forma  t  = " 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.

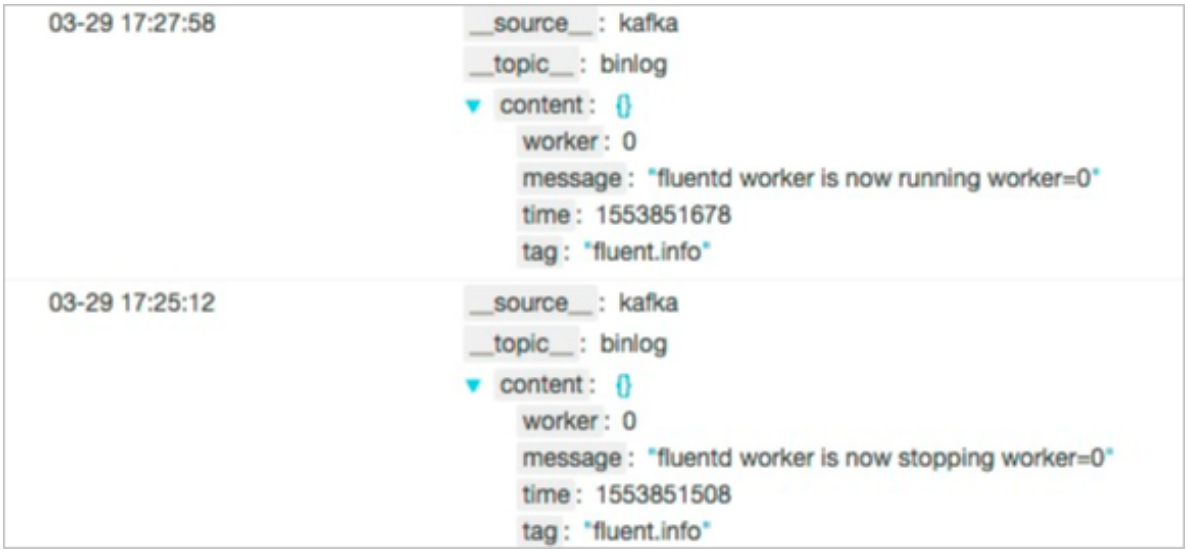
Fluentd provides many input, processing, and output plug-ins. Specifically, the Kafka plug-in can help Fluentd export data to Kafka. You only need to install and configure this plug-in.

The sample code is as follows:

```
< match  **>
  @ type   kafka
  #  Brokers :  You   can   choose   either   brokers   or
 zookeeper .
   brokers        cn - hangzhou . log . aliyuncs . com : 10012
   default_to  pic   test - logstore - 1
   default_me  ssage_key   content
   output_dat  a_type   json
   output_inc  lude_tag   true
   output_inc  lude_time   true
   sasl_over_  ssl   true
   username  < youruserna  me >
   password  < yourpasswo  rd >
   ssl_ca_cer  ts_from_sy  stem   true
  #  ruby - kafka   producer   options
   max_send_r  etries   10000
   required_a  cks   1
   compressio  n_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_clie nt_jaas . conf .*

     ```
     KafkaClien t {
       org . apache . kafka . common . security . plain . PlainLogin
     Module   required
       username ="< youruserna me >"
       password ="< yourpasswo rd >";
     };
     ```

  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:

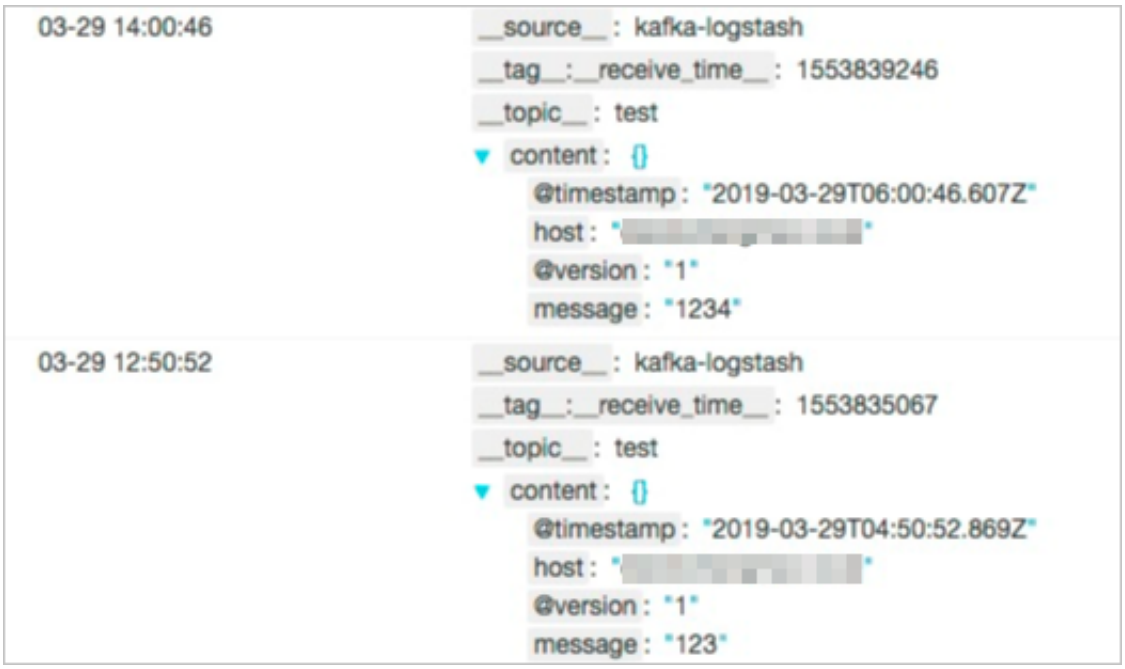
     ```
     input { stdin { } }
     output {
       stdout { codec => rubydebug }
       kafka {
         topic_id => " test - logstore - 1 "
         bootstrap_ servers => " cn - hangzhou . log . aliyuncs .
     com : 10012 "
         security_p rotocol => " SASL_SSL "
         ssl_trusts tore_locat ion => "/ etc / client - root .
     truststore . jks "
         ssl_trusts tore_passw ord => " 123456 "
         jaas_path => "/ etc / kafka_clie nt_jaas . conf "
         sasl_mecha nism => " PLAIN "
         codec => " json "
     ```

```
    client_id  => " kafka - logstash "
  }
}
```

Note:
The configuration in the preceding sample code is used for a connectivity test. In actual applications, we recommend that you remove the stdout output configuration.

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 Use the Syslog protocol to collect data to Log Service

This topic describes how to use the Syslog protocol to collect data to Log Service without using the collection agent to forward collected data. The following sections contain information about the limits, required configurations, sample logs, and application examples detailing possible uses of the Syslog protocol.

Limits

- **The Syslog protocol must be the standard RFC5424 protocol.**

- Each log can be up to 64 KB.
- You must use TLS 1.2 (based on TCP) to guarantee data transmission security.

Required configurations

- You must associate Syslog with Log Service.

  Specifically, the port number of Syslog (10009) and the service endpoint of the target Project are required. For example, `cn-hangzhou-intranet.log.aliyuncs.com:10009` associates Syslog with a Project. The service endpoint of a Project in Log Service is a URL that is used to access the Project. For more information, see #unique_28.
- You must set the `STRUCTURED-DATA` field (contains the parameters related to Log Service) of Syslog as follows.

Configuration	Description	Example
`STRUCTURED-DATA`	This field requires a name . This name must be Logservice.	Logservice
`Project`	The user name is mapped to a project name.	test-project-1
`Logstore`	The topic is mapped to a Logstore name. You must create a Logstore before you can collect data.	test-logstore-1
`access-key-id`	Indicates your AccessKey ID. We recommend that you use the AK of a RAM user. For more information, see Grant a RAM user the permissions to access Log Service.	<yourAccessKeyId>

Configuration	Description	Example
`access-key-secret`	Indicates your AccessKey Secret. We recommend that you use the AK of a RAM user. For more information, see Grant a RAM user the permissions to access Log Service.	<yourAccessKeySecret>

Sample logs

Log Service automatically resolves the received Syslog data. Furthermore, Log Service removes the Logservice field to avoid the AccessKey from being leaked. The following are the fields uploaded to Log Service (for more information about fields, see RFC5424 protocol).

Field	Description
__source__	Indicates the hostname field in Syslog.
__topic__	This field is fixed as syslog-forwarder.
__facility__	Indicates the facility (device or module) information in Syslog.
__program__	Indicates a process name.
__serverity__	Indicates the log severity.
__priority__	Indicate the log priority.
__unixtimestamp__	Indicates a log timestamp in seconds.
content	Indicates the msg field in Syslog.

Application examples

For each of the following application examples, the required parameters are set as follows to have the Syslog logs be directly collected on a host to Log Service:

- The project in Log Service is set as `test-project-1`.
- The Logstore in Log Service is set as `test-logstore-1`.
- The region where the project is located is set as `cn-hangzhou`.

- The AccessKey ID of a RAM user that has the write permissions is set as `< yourAccess KeyId >`, and the AccessKey Secret of the RAM user is set as `< yourAccess KeySecret >`.
- Application example 1: Use Rsyslog to forward system logs to Log Service

  Rsyslog is installed in a Linux host to process system logs. You can use Rsyslog to forward the system logs to Log Service. The Rsyslog configuration file format varies by Rsyslog versions. You can run the `man rsyslogd` command to view which Rsyslog version is being used.

  Note:

  Make sure that you have installed the gnutls module in Rsyslog. If the module is not installed in Rsyslog, you can run the `sudo apt - get install rsyslog - gnutls` command or the `sudo yum install rsyslog - gnutls` command to install the module.

  - Rsyslog V8 and later

    `$ DefaultNet streamDriv erCAFile` indicates the directory where the system root certificate is located.

```
#  Setup   disk   assisted   queues
        $ WorkDirect  ory  / var / spool / rsyslog  #  where   to
   place   spool   files
        $ ActionQueu  eFileName   fwdRule1      #  unique   name
   prefix   for   spool   files
        $ ActionQueu  eMaxDiskSp  ace   1g       #  1gb   space
   limit  ( use   as   much   as   possible )
        $ ActionQueu  eSaveOnShu  tdown   on      #  save
 messages   to   disk   on   shutdown
        $ ActionQueu  eType   LinkedList        #  run
 asynchrono  usly
        $ ActionResu  meRetryCou  nt  - 1        #  infinite
 retries   if   host   is   down
        $ ActionSend  TCPRebindI  nterval   100   #  close   and
   re - open   the   connection   to   the   remote   host   every
   100   of   messages   sent .
        # RsyslogGnu  TLS   set   to   default   ca   path
        $ DefaultNet  streamDriv  erCAFile  / etc / ssl / certs /
 ca - bundle . crt
         template ( name =" LogService  Format "  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 =\"< yourAccess  KeyId >\"  access - key
 - secret =\"< yourAccess  KeySecret >\"] % msg %\ n "
        )
        #  Send   messages   to   Loggly   over   TCP   using
 the   template .
```

```
          action ( type =" omfwd "  protocol =" tcp "  target
=" cn - hangzhou . log . aliyuncs . com "  port =" 10009 "
template =" LogService  Format "  StreamDriv  er =" gtls "
StreamDriv  erMode =" 1 "  StreamDriv  erAuthMode =" x509 / name
"  StreamDriv  erPermitte  dPeers ="*. log . aliyuncs . com ")
```

- Rsyslog V7 and earlier

  `$ DefaultNet streamDriv erCAFile` indicates the directory where the system root certificate is located.

```
#  Setup   disk   assisted   queues
        $ WorkDirect  ory  / var / spool / rsyslog        #
 where   to   place   spool   files
        $ ActionQueu  eFileName   fwdRule1            #  unique
  name   prefix   for   spool   files  $ ActionQueu  eMaxDiskSp
 ace   1g             #  1gb   space   limit  ( use   as   much
  as   possible ) $ ActionQueu  eSaveOnShu  tdown   on
 #  save   messages   to   disk   on   shutdown
        $ ActionQueu  eType   LinkedList             #  run
 asynchrono  usly
        $ ActionResu  meRetryCou  nt  - 1              #
 infinite   retries   if   host   is   down  $ ActionSend
 TCPRebindI  nterval   100         #  close   and   re - open
 the   connection   to   the   remote   host   every   100   of
 messages   sent .
        #  RsyslogGnu  TLS   set   to   default   ca   path
        $ DefaultNet  streamDriv  erCAFile  / etc / ssl / certs /
 ca - bundle . crt
        $ ActionSend  StreamDriv  er   gtls
        $ ActionSend  StreamDriv  erMode   1
        $ ActionSend  StreamDriv  erAuthMode   x509 / name
        $ ActionSend  StreamDriv  erPermitte  dPeer   cn -
 hangzhou . log . aliyuncs . com
         template ( name =" LogService  Format "  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 =\"< yourAccess  KeyId >\"  access - key
- secret =\"< yourAccess  KeySecret >\"] % msg %\ n ")
        *.*  action ( type =" omfwd "  protocol =" tcp "  target
=" cn - hangzhou . log . aliyuncs . com "  port =" 10009 "
template =" LogService  Format ")
```

1. Select the required version of Rsyslog, set either of the preceding configuration template according to your needs, and add the template to the end of the Rsyslog configuration file.

   Note:

Generally, the directory to store the Rsyslog configuration file is `/etc/rsyslog.conf`.

2. Run the `sudo service rsyslog restart` and `sudo /etc/init.d/syslog-ng restart` commands or the `systemctl restart rsyslog` command to restart Rsyslog.
3. If no system logs are generated, run the `logger` command to generate a few logs for test. For example, `logger hello world!`.

- Example 2: Use Syslog-ng to forward system logs to Log Service

Syslog-ng is an open-source daemon for log management that helps Unix and Unix-like operating systems support the Syslog protocol. You can run the `sudo yum install syslog-ng` command or the `sudo apt-get install syslog-ng` command to install Syslog-ng.

```
### Syslog-ng Logging Config for LogService ###
        template LogService Format {
        template ("<${ PRI }> 1 ${ ISODATE } ${ HOST :--} ${
PROGRAM :--} ${ PID :--} ${ MSGID :--} [ logservice project =\"
test-project-1 \" logstore =\" test-logstore-1 \" access
-key-id =\"< yourAccess KeyId >\" access-key-secret =\"<
yourAccess KeySecret >\"] $ MSG \ n "); template_e scape ( no
);
       };
        destinatio n d_logservi ce {
        tcp (" cn-hangzhou . log . aliyuncs . com " port ( 10009
)
        tls ( peer-verify ( required-untrusted ))
        template ( LogService Format ));
       };
        log {
        source ( s_src ); # default use s_src
        destinatio n ( d_logservi ce );
       };
       ### END Syslog-ng Logging Config for LogService
 ###
```

Note:
By default, Rsyslog is installed in an ECS instance to process system logs. If you want to use Syslog-ng, you need to uninstall Syslog-ng first because Rsyslog and Syslog-ng cannot work together.

1. Set the preceding configuration template as needed, and add the template to the end of the Syslog-ng configuration file.

Note:

The typical directory to store 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`, and `sudo systemctl restart syslog-ng` commands to restart Syslog-ng.
3. If no system logs are generated, run the `logger` command to generate a few logs for test. For example, `logger hello world!`.

Troubleshoot exceptions

- Manually upload a log to check the network connectivity

  You can use the ncat command to simulate uploading system logs. This method can check the network connectivity and whether the AccessKey has the permissions to upload logs. If ncat is not installed in your server, you can run the `sudo yum install nmap-ncat` command to install it. For example, the following command is run to send a system log to Log Service:

```
[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
```

  The project in Log Service is set as `test-project-1` and the Logstore is set as `test-logstore-1`. The region where the project is located is set as cn-hangzhou. The AccessKey ID of a RAM user that has the write permissions is set as `<yourAccessKeyId>`, and the AccessKey Secret of the RAM user is set as `<yourAccessKeySecret>`.

  Note:

  - When you run the ncat command to upload a log, you must specify your current time for the log in ISO 8601 format. For example, if you upload a log at 2019-03-28T11:00:15.003, you must convert it to 2019-03-28T03:00:15.003Z before you add it to the ncat command.

- The ncat commands cannot identify network connection interrupt. Therefore, you need to enter messages and press Enter within 30 seconds after you run a ncat command.

After you run the command, you can preview logs in the Log Service console. For more information, see Preview log data.

- Diagnose log collection errors

  If you fail to manually upload logs, you can diagnose log collection errors to view error information. For more information, see #unique_253.

- Check Rsyslog error logs

  Rsyslog logs are saved in the `/ var / log / message` directory be default. You can run a vim command to view the directory.

  - Rsyslog error

    ```
    dlopen : / usr / lib64 / rsyslog / lmnsd_gtls . so :  cannot
      open   shared   object   file :  No   such   file   or
    directory
    ```

    This error occurred because the gnutls module is not installed. You can run the `sudo apt - get install rsyslog - gnutls` command or the `sudo yum install rsyslog - gnutls` command to install and restart Rsyslog.

  - Rsyslog error

    ```
    unexpected   GnuTLS   error   - 53   -   this   could   be   caused
      by   a   broken   connection .  GnuTLS   reports : Error   in
    the   push   function
    ```

    This error occurred because the TCP connection was idle for long time and then was forcibly shutdown. Rsyslog can automatically connect to Log Service.

- Check Syslog-ng error logs

  Syslog-ng logs are saved in the Journal logs by default. You can run the `systemctl status syslog - ng . service` and `journalctl - xe` command to check detailed logs.

  Syslog-ng initialization error

  ```
  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
```

You must check that the Syslog-ng configuration file format is valid and that the settings in the file do not conflict (for example, multiple internal(); are not set).

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.

  Note:
  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.

Parameters	Description
endpoint	Log Service endpoint. Example: `http :// regionid . example . com` . For more information, see Log Service endpoint.
project	The project name of Log Service.
logstore	The Logstore name.
topic	The log topic name. The default value is null.
source	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.

Parameters	Description
access_key_id	The AccessKey ID of the Alibaba Cloud account.
access_key_secret	The AccessKey Secret of the Alibaba Cloud account.
max_send_retry	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.

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 #unique_101.
   - 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 #unique_257.
   - 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 #unique_258.

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.

   ```
   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 #unique_259.

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

```
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

```
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.
- multiline: Customizes complex types of multiline logs.
- kv: Structurally parses logs of key-value pair type.

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.

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 manipulate d` 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 project of Apache, which allows you to set the log output destination to console, file, GUI component, socket server, NT event recorder, or UNIX Syslog daemon. You can also set the output format and level of each log to control log generation with a finer granularity. These configurations can be performed flexibly by using a configuration file without modifying application codes.

Alibaba Cloud Log4j Appender allows you to set the log output destination to Alibaba Cloud Log Service. For more information about download link and user guide, refer to Github.

5.6.3 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:

- C Producer Library (recommended for servers)
- C Producer Lite Library (recommended for IOT and smart devices)

5.6.4 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 Apache log

The Apache log format and directory are generally in the `/etc/apache2/httpd.conf` configuration file.

Log format

By default, the Apache log configuration file defines two print formats: combined format and common format. You can also create your own customized log print format as needed.

- Combined format:

```
LogFormat  "% h  % l  % u  % t  \"% r \" %> s  % b  \"%{ Referer }
i \" \"%{ User - Agent } i \""  combined
```

- Common format:

```
LogFormat  "% h  % l  % u  % t  \"% r \" %> s  % b "
```

- Customized format:

```
LogFormat  "% h  % l  % u  % t  \"% r \" %> s  % b  \"%{ Referer }
i \" \"%{ User - Agent } i \" % D  % f  % k  % p  % q  % R  % T  %
I  % O "  customized
```

You need to specify the print format, log file path, and log name of the current log in the Apache log configuration file. For example, the following log configuration file indicates that combined print format is used, and the log path and name is displayed as /var/log/apache2/access_log.

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

Field description

Format	Key name	Description
%a	client_addr	Client IP address.
%A	local_addr	Local private IP address.
%b	response_size_bytes	Size of response in bytes. When the size of response is null, this field is a hyphen (-).
%B	response_bytes	Size of response in bytes. When the size of response is null, this field is a hyphen (-).
%D	request_time_msec	Request time, in microseconds.
%h	remote_addr	Remote hostname.
%H	request_protocol_supple	Request protocol.
%l	remote_ident	Client log name from identd.
%m	request_method_supple	Request method.
%p	remote_port	Server port.
%P	child_process	Child process ID.

%q	request_query	Query string. If no query string exists, this field is an empty string.
"%r"	request	Request content, including the request method name, address, and HTTP protocol.
%s	status	HTTP status code.
%>s	status	Final HTTP status code.
%f	filename	Filename.
%k	keep_alive	Number of keepalive requests.
%R	response_handler	Handler on the server.
%t	time_local	Server time.
%T	request_time_sec	Request time, in seconds.
%u	remote_user	Client username.
%U	request_uri_supple	Requested URL path. No query is included in the path.
%v	server_name	Server name.
%V	server_name_canonical	Server name conforming to the UseCanonicalName setting.
%I	bytes_received	Number of bytes received by the server. You must enable the mod_logio module.
%O	bytes_sent	Number of bytes sent by the server. You must enable the mod_logio module.
"%{User-Agent}i"	http_user_agent	Client information.
"%{Rererer}i"	http_referer	Source page.

Sample log

```
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 )  AppleWebKi  t / 537 . 36  ( KHTML ,  like
Gecko )  Chrome / 48 . 0 . 2564 . 97  Safari / 537 . 36 "
```

Configure a Logtail client to collect Apache logs

1. On the Logstores page, click the Data Import Wizard icon.

2. Select a data type.

   Select APACHE Access Log.

3. Configure data source.

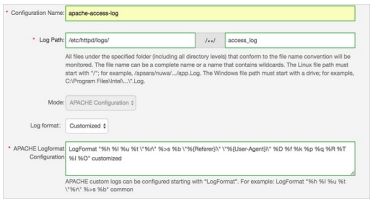
   a. Enter the Configuration Name and Log Path.

   b. Select a Log format.

   c. Enter APACHE Logformat Configuration if you select the customized log format.

      Enter the log format configuration fields of the standard APACHE configuration file. Generally, the configuration file starts with LogFormat.

      > **Note:**
      > If you select common or combined from the Log format drop-down list, the configuration fields of the corresponding log format are automatically added here. You need to confirm whether the added configuration fields are consistent with the format defined in the local Apache configuration file.

   d. Confirm APACHE Key Name.

      Log Service automatically reads your Apache keys. Confirm the Apache key names on the current page.

   e. (Optional) Configure Advanced options.

Parameter	Description
Upload Raw Log	Specifies whether to upload the raw log. If you turn on this switch, the raw log content is uploaded as the __raw__ field with the parsed log content.

Parameter	Description
Topic Generation Mode	· Null - Do not generate topic: The default value, which specifies that the topic is set to a null string. You can query logs without entering the topic. · Machine Group Topic Attributes: sets the topic based on a machine group to differentiate log data generated on different frontend servers. · File Path RegEx: uses Custom RegEx to extract a part of the log path as the topic. This mode is used to differentiate log data generated by different users or instances.
Custom RegEx	The custom regular expression specified if you set Topic Generation Mode to File Path RegEx.
Log File Encoding	· utf8: specifies UTF-8 encoding. · gbk: specifies GBK encoding.
Maximum Directory Monitoring Depth	The maximum depth of the monitored directory when logs are collected from the log source, that is, at most how many levels of directories can be monitored. Valid values : [0, 1000]. A value of 0 indicates that only the current directory is monitored.
Timeout	Specifies whether the system considers that a log file has timed out if the file is not updated within the specified period. You can set Timeout as follows: · Never: specifies that all log files are continuously monitored without timeout. · 30 Minute Timeout: specifies that 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 .

Parameter	Description
Filter Configuration	The filter conditions that logs must completely meet before they can be collected. For example: · Collect logs that meet a condition: Set a condition `Key : level  Regex : WARNING | ERROR`, which indicates that only logs whose level is WARNING or ERROR are collected. · Filter logs that do not meet a condition: - Set a condition `Key : level  Regex :^(?!.*( INFO | DEBUG )).*`, which indicates that logs whose level is INFO or DEBUG are not collected. - Set a condition `Key : url  Regex :.*^(?!.*( healthchec  k )).*`, which indicates that logs with healthcheck in url are not collected. For example, logs in which the key is url and the value is `/ inner / healthchec  k / jiankong . html` are not collected. For more examples, see regex-exclude-word and regex-exclude-pattern.

4. Click Next.
5. Select a machine group and then click Apply to Machine Group.

   If you have not created any machine group, click +Create Machine Group to create one.

   After you apply the Logtail configuration to the machine group, Log Service collects Apache logs according to the configuration. You can configure indexes and log shippers by following the steps of the Data Import Wizard.

5.7.2 Nginx logs

The Nginx log format and directory are generally in the configuration file `/ etc / nginx / nginx . conf .`

Nginx log format

The log configuration file defines the print format of Nginx logs, that is, the main format:

```
log_format   main  '$ remote_add  r  - $ remote_use  r  [$ time_local
] "$ request " '
                '$ request_ti  me  $ request_le  ngth  '
                '$ status  $ body_bytes  _sent  "$ http_refer  er " '
                '"$ http_user_  agent "';
```

The declaration uses the main log format and the written file name.

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

Field Description

Field name	Definition
remoteaddr	The IP address of the client.
remote_user	The username of the client.
request	The requested URL and HTTP protocol.
status	The request status.
bodybytessent	The number of bytes (not including the size of the response header) sent to the client. The total number of bytes for this variable is the same as that sent to the client by bytes_sent in modlogconfig of the Apache module.
connection	The connection serial number.
connection_requests	The number of requests received by using a connection.
msec	The log write time, which is which is measured in seconds and precise to milliseconds.
pipe	Whether or not requests are sent by using the HTTP pipeline. p indicates requests are sent by using the HTTP pipeline. Otherwise, the value is . .

Field name	Definition
httpreferer	Web page link from which the access is directed.
“http_user_agent”	Information about the browser on the client. http_user_agent must be enclosed in double quotation marks.
requestlength	The length of a request, including the request line, request header, and request body.
Request_time	The request processing time, which is measured in seconds and precise to milliseconds. The time starts when the first byte is sent to the client and ends when the logs are written after the last character is sent to the client.
[$time_local]	he local time in the general log format . This variable must be enclosed in brackets.

Log sample

```
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 "
```

Configure Logtail to collect Nginx logs

1. Click the Data Import Wizard chart in the Logstore list page to enter the data import wizard.
2. Select a data source.

   Select the text file and click Next.

3. Select the data source.

   a. Enter the Configuration Name, and Log Path.

   b. Enter the nNginx log format.

      Complete the standard Nginx profile log configuration section, typically beginning with the `log_format`. Log Service automatically reads your Nginx key.

   c. Set Advanced Options according to your requirements. Click Next after completing the configurations.

      For more information about advanced options, see Advanced options.

   After configuring Logtail, apply the configuration to the machine group to start collecting Nginx logs standardly.

5.7.3 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 constructi on method of formatter needs two parameters: message format string and message date string. Both of the parameters are optional.

Python log format:

```
import   logging
import   logging . handlers
LOG_FILE  = ' tst . log '
handler  =  logging . handlers . RotatingFi  leHandler ( LOG_FILE ,
maxBytes  =  1024 * 1024 ,  backupCoun  t  =  5 ) #  Instantiat  e
the   handler
```

```
fmt  = '%( asctime ) s  - %( filename ) s :%( lineno ) s  - %( name )
s  - %( message ) s '
formatter  =  logging . Formatter ( fmt )   #  Instantiat  e   the
formatter
handler . setFormatt  er ( 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:

Format	Meaning
%(name)s	The logger name of the generated log.
%(levelno)s	The log level in numeric format, including DEBUG, INFO, WARNING, ERROR, and CRITICAL.
%(levelname)s	The log level in text format, including DEBUG, INFO, WARNING, ERROR, and CRITICAL.
%(pathname)s	The full path of the source file where the statement that outputs the log resides (if available).
%(filename)s	The file name.
%(module)s	The name of the module where the statement that outputs the log resides.
%(funcName)s	The name of the function that calls the log output.
%(lineno)d	The code line where the function statement that calls the log output resides (if available).
%(created)f	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.

Format	Meaning
%(relativeCreated)d	The interval (in milliseconds) between the log created time and the time that the logging module is loaded.
%(asctime)s	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).
%(msecs)d	The log creation time in the milliseconds.
%(thread)d	The thread ID (if available).
%(threadName)s	The thread name (if available).
%(process)d	The process ID (if available).
%(message)s	The log message.

Log sample

Log sample

```
2015 - 03 - 04   23 : 21 : 59 , 682  -  log_test . py : 16  -  tst
-  first   info   message
2015 - 03 - 04   23 : 21 : 59 , 682  -  log_test . py : 17  -  tst
-  first   debug   message
```

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  - %( levelno ) 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
  ```

```
1398659966  87072   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 +(\ d +)\ s +(\ w +)\ s +-\ s +(. *)
```

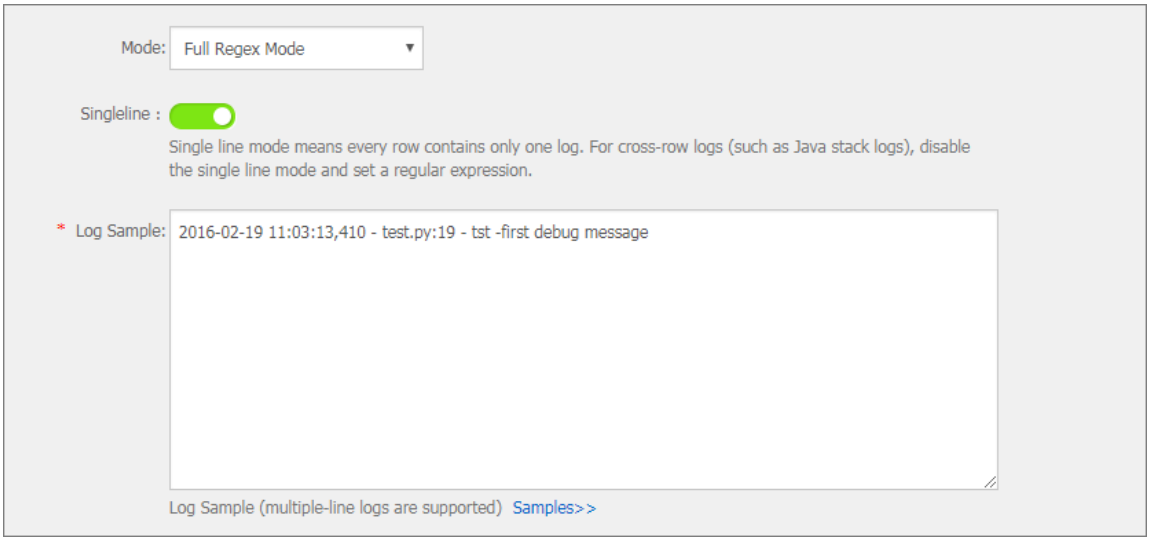
Configure Logtail to collect Python logs

For the detailed procedures of collecting Python logs by using Logtail, see #unique_272. Select the corresponding configuration based on your network deployment and actual situation.

1. Create a project and a Logstore. For detailed procedures, see #unique_112.
2. On the Logstores page, click the Data Import Wizard icon.
3. Select a data source.

   Select the Text File.

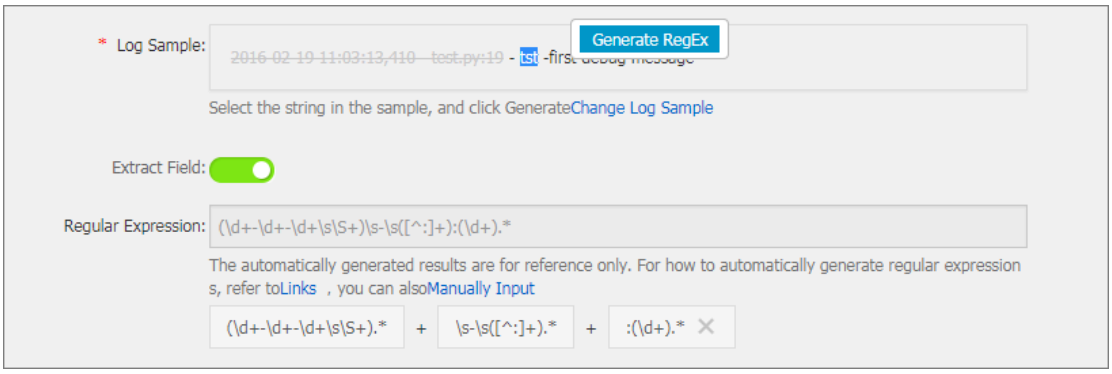
4. Configure the data source.
   a. Enter the Configuration Name and Log Path, and then select the Full Regex Mode from the mode drop-down list.
   b. Turn on the Singleline switch.
   c. Enter Log Sample.



   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.

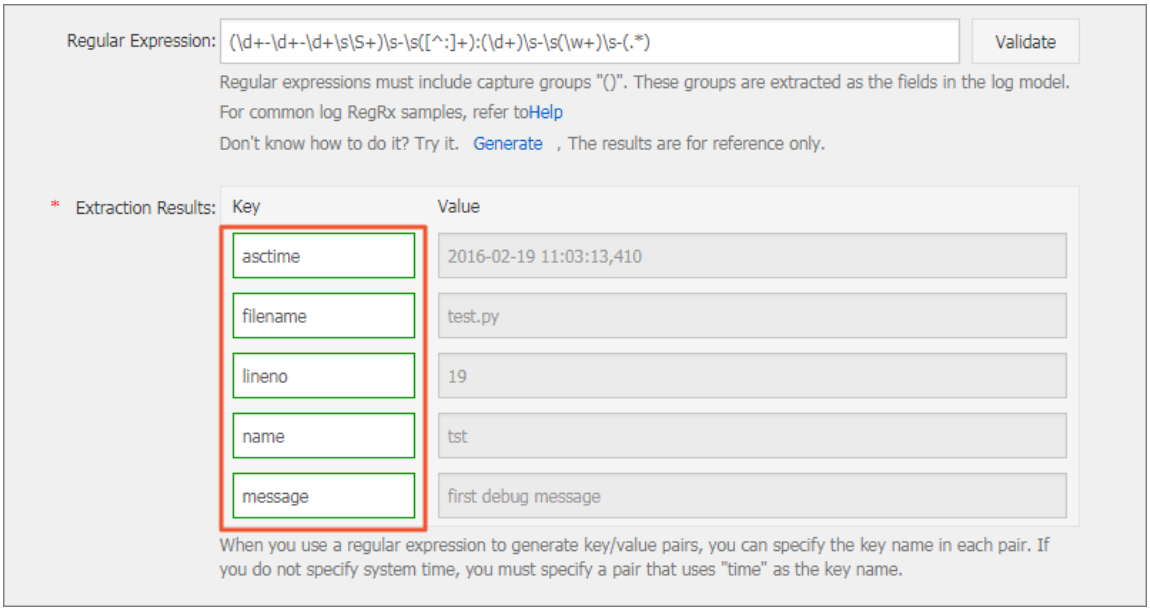
C. Validate the regular expression.

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.

5.7.4 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 #unique_274.

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.

```
< Configurat  ion   status =" WARN ">
  < Appenders >
    < Console   name =" Console "  target =" SYSTEM_OUT ">
      < PatternLay  out   pattern ="% d { yyyy - MM - dd   HH : mm :
 ss : SSS   zzz } [% t ] %- 5level  % logger { 36 } - % msg % n "/>
    </ Console >
  </ Appenders >
  < Loggers >
    < Logger   name =" com . foo . Bar "  level =" trace ">
      < AppenderRe  f   ref =" Console "/>
    </ Logger >
    < Root   level =" error ">
      < AppenderRe  f   ref =" Console "/>
    </ Root >
  </ Loggers >
</ Configurat  ion >
```

For how to configure Logtail to collect Log4j logs, see #unique_275. 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 e 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 . PermanentT  airDaoImpl  -  Fail   to   Read   Permanent
Tair , key : e : 4702173193  19741_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 .
```

The regular expression used to extract log information:

```
(\ d +-\ d +-\ d +\ s \ d +:\ d +:\ d +,\ d +)\ s \[([^\]]*)\]\ s (\ S
 +)\ s +(\ S +)\ s -\ s (.
```

Time conversion format:

```
% Y -% m -% d  % H :% M :% S
```

Extraction results of the log sample:

Key	value
time	2013-12-25 19:57:06,954
ip	10.207.37.161
level	WARN
class	impl.PermanentTairDaoImpl
message	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]

5.7.5 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.

```
 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 #unique_275. 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_lo  g ?
 ignoreErro  r = true   HTTP / 1 . 1 "  304  - " http ://
 aliyun . com /" " Mozilla / 5 . 0  ( Macintosh ;  Intel   Mac
 OS   X   10_10_3 )  AppleWebKi  t / 537 . 36  ( KHTML ,  like
 Gecko )  Chrome / 48 . 0 . 2564 . 97   Safari / 537 . 36 "
```

  - Regular expression type

```
\[([^]]+)]\ s \[(\ w +)]\ s (\ w +)\ s -\ s (\ S +)\ s -\ s -\ s
 "([^"]+)"\ s (\ d +)[^"]+("[^"]+)"\ s "([^"]+). *
```

  - Extracted fields:

`time` , `level` 、, `loggerName` , `ip` , `request` , `status` , `referer` and `user_agent` .

5.7.6 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 .
c4a1a0aecd  b194316955  5231dcc4ad  fb7 . cn - hangzhou . alicontain
er . com / wp - admin / install . php " " Mozilla / 5 . 0  (
Macintosh ;  Intel   Mac   OS   X   10_10_5 )  AppleWebKi  t / 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:

Key	Value
ip	10.10.10.1
time	07/Jan/2016:21:06:39 +0800
method	GET
url	/wp-admin/js/password-strength-meter.min.js? ver=4.4 HTTP/1.0
status	200
length	776
ref	http://wordpress.c4a1a0aecdb1943169555231dcc4adfb7.cn-hangzhou.alicontainer.com/wp-admin/install.php
user-agent	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

5.7.7 Delimiter logs

Log introduction

Delimiter logs use line breaks as the boundary. Each line is a log. The fields of each log are delimited by a fixed delimiter. A delimiter can contain a single character or multiple characters, including the tab (\t), space, vertical bar (|), comma (,), and semicolon (;). Fields that contain the delimiter must be enclosed in a quote, which is double quotation marks (" ").

Comma-separated values (CSV) logs and tab-separated values (TSV) logs are common delimiter logs.

Log Service supports a delimiter in single-character or multi-character mode to delimit fields in each delimiter log.

Single-character mode

In single-character mode, you must specify a delimiter. You can also specify a quote as required.

- Delimiter: The fields of each log are delimited by a single-character delimiter, such as the tab (`\ t`), the vertical bar (`|`), the space, the comma (`,`), the semicolon (`;`), or a non-printable character.

Note:

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

If a double quotation mark (") is included in a log but not used as a quote, it must be escaped and processed as double quotation marks (""). Log Service automatically restores 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 double quotation marks ("") in the content of a field. If the use of the double quotation mark (") does not comply with the format definitions of delimiter logs, you can consider using other methods such as the simple mode or full regex mode to parse fields.

For example, when the comma (,) is used as the delimiter and a log field contains the double quotation mark (") and comma (,), this field must be enclosed in the quote and the double quotation mark (") in the content of this field must be escaped into double quotation marks (""). The log format after processing is `1999 , Chevy ," Venture "" Extended Edition , Very Large ""","", 5000 . 00`. The log can be parsed into five fields as follows: `1999`, `Chevy`, `Venture " Extended Edition , Very Large "`, null field, and `5000 . 00`.

- Quote: When log fields contain the delimiter, you must specify a quote to enclose such fields and ensure that they can be parsed properly. Log Service parses the content enclosed in the quote as a complete field. Only the delimiter can exist between fields.

  Note:
  If any characters other than the delimiter, such as the space or tab (\t), exist between fields, you need to modify the log format.

  The quote can contain a single character, such as the tab (`\ t`), the vertical bar (|), the space, the comma (,), the semicolon (;), or a non-printable character.

  For example, when the comma (,) is used as the delimiter and double quotation marks (" ") are used as the quote, the log format is `1997 , Ford , E350 ," ac , abs , moon ", 3000 . 00`. The log 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 the quote 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`.

Multi-character mode

In multi-character mode, a delimiter can contain two or three characters, such as `||`, `&&&`, or `^_^`. In this mode, Log Service parses logs based on the delimiter only. You do not need to use a quote to enclose log fields.

Note:

You need to ensure that log fields do not contain the delimiter, otherwise Log Service may incorrectly parse these fields.

For example, if the delimiter is set to `&&`, the log `1997 && Ford && E350 && ac & abs & moon && 3000 . 00` can be parsed into five fields as follows: `1997`, `Ford`, `E350`, `ac & abs & moon`, and `3000 . 00`.

Sample logs

- Logs whose fields are delimited by a single-character delimiter

```
05 / May / 2016 : 13 : 30 : 28 , 10 . 10 . *. *," POST  /
PutData ?  Category = YunOsAccou  ntOpLog & AccessKeyI  d
=****************& 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 = YunOsAccou  ntOpLog & AccessKeyI  d
=****************& Date = Fri % 2C % 2028 % 20Jun % 202013
% 2006 % 3A53 % 3A30 % 20GMT & Topic = raw & Signature
=******************************  HTTP / 1 . 1 ", 401 , 23472 ,
aliyun - sdk - java
```

- Logs whose fields are delimited by a multi-character delimiter

```
05 / May / 2016 : 13 : 30 : 28 && 10 . 200 . **. **&& POST
/ PutData ?  Category = YunOsAccou  ntOpLog & AccessKeyI  d
=****************& Date = Fri % 2C % 2028 % 20Jun % 202013 % 2006
% 3A53 % 3A30 % 20GMT & Topic = raw & Signature = pD12XYLmGx  KQ %
2Bmkd6x7hA  gQ7b1c % 3D   HTTP / 1 . 1 && 200 && 18204 && aliyun -
sdk - java
05 / May / 2016 : 13 : 31 : 23 && 10 . 200 . **. **&& POST
/ PutData ?  Category = YunOsAccou  ntOpLog & AccessKeyI  d
=****************& Date = Fri % 2C % 2028 % 20Jun % 202013
% 2006 % 3A53 % 3A30 % 20GMT & Topic = raw & Signature
=******************************  HTTP / 1 . 1 && 401 && 23472 &&
aliyun - sdk - java
```

Configure Logtail to collect delimiter logs

For more information about how to configure Logtail to collect delimiter logs, see Collect text logs. You can select the corresponding configuration based on your network deployment and actual situation.

1. On the Logstores page of the target project in the Log Service console, click the Data Import Wizard icon of the target Logstore.
2. Select a data source.

   Select Text File and go to the next step.

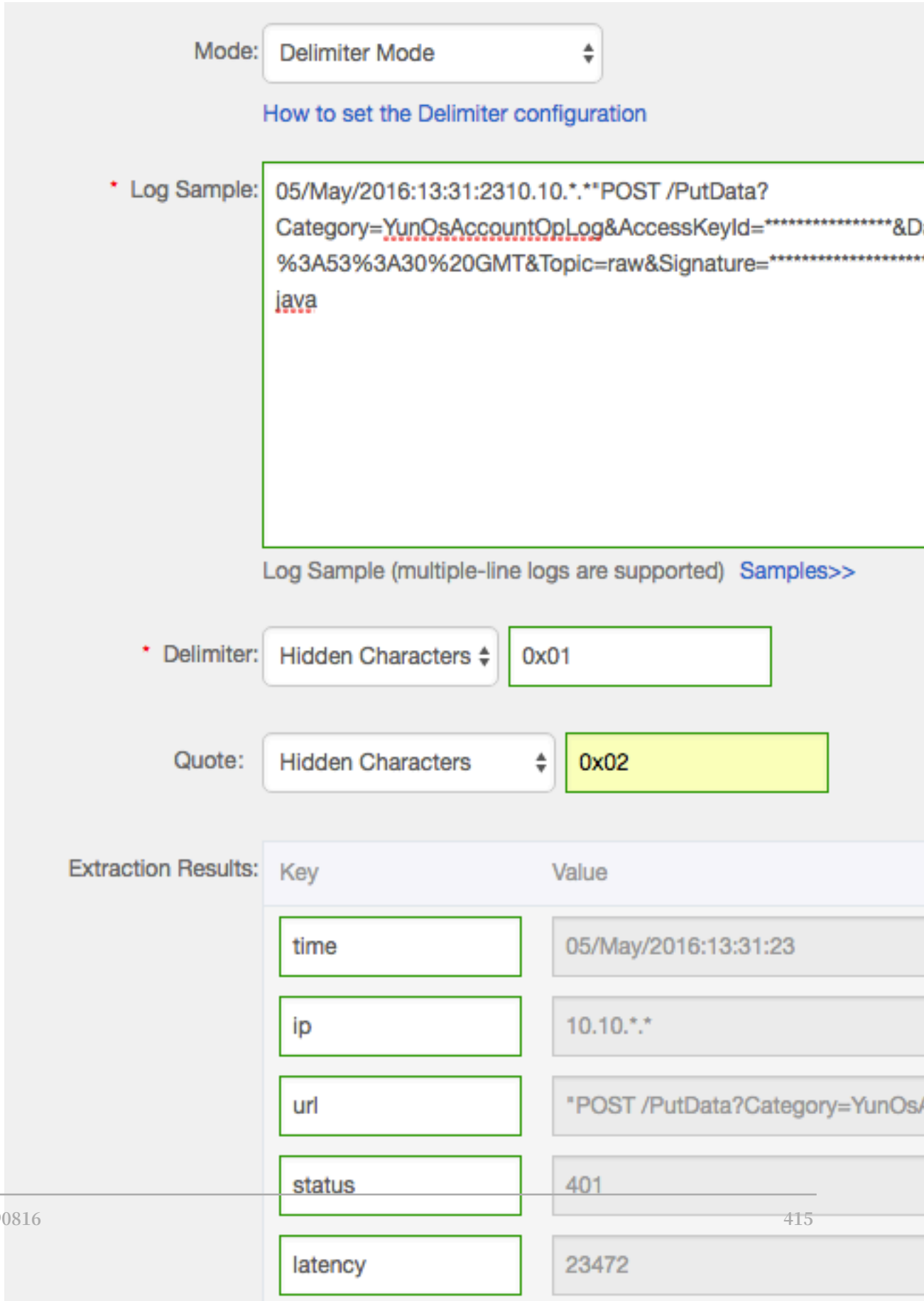
3. Configure the data source.
   a. Enter the configuration name and log path. Then, select Delimiter Mode as the log collection mode.
   b. Enter the log sample and select the delimiter and quote.

      Select the appropriate delimiter and quote based on the log format. Otherwise, Log Service may fail to parse logs.

       Note:
      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`.

Figure 5-4: Configure the data source



c. Specify the keys in the log extraction results.

After you enter a log sample and select a delimiter, Log Service extracts log fields according to your selected delimiter, and defines them as values. You must specify the key for each value.

In the preceding log sample, the non-printable character `0x01` is used as the delimiter and `0x02` is used as the quote. The log is parsed into six fields. Enter time, ip, url, status, latency, and user-agent as the keys of six fields, respectively.

d. Determine whether to upload a log with some fields missing.

Configure whether to upload a log whose number of parsed fields is less than the number of configured keys. If you enable Incomplete Entry Upload, the log is uploaded. If you disable Incomplete Entry Upload, the log is discarded.

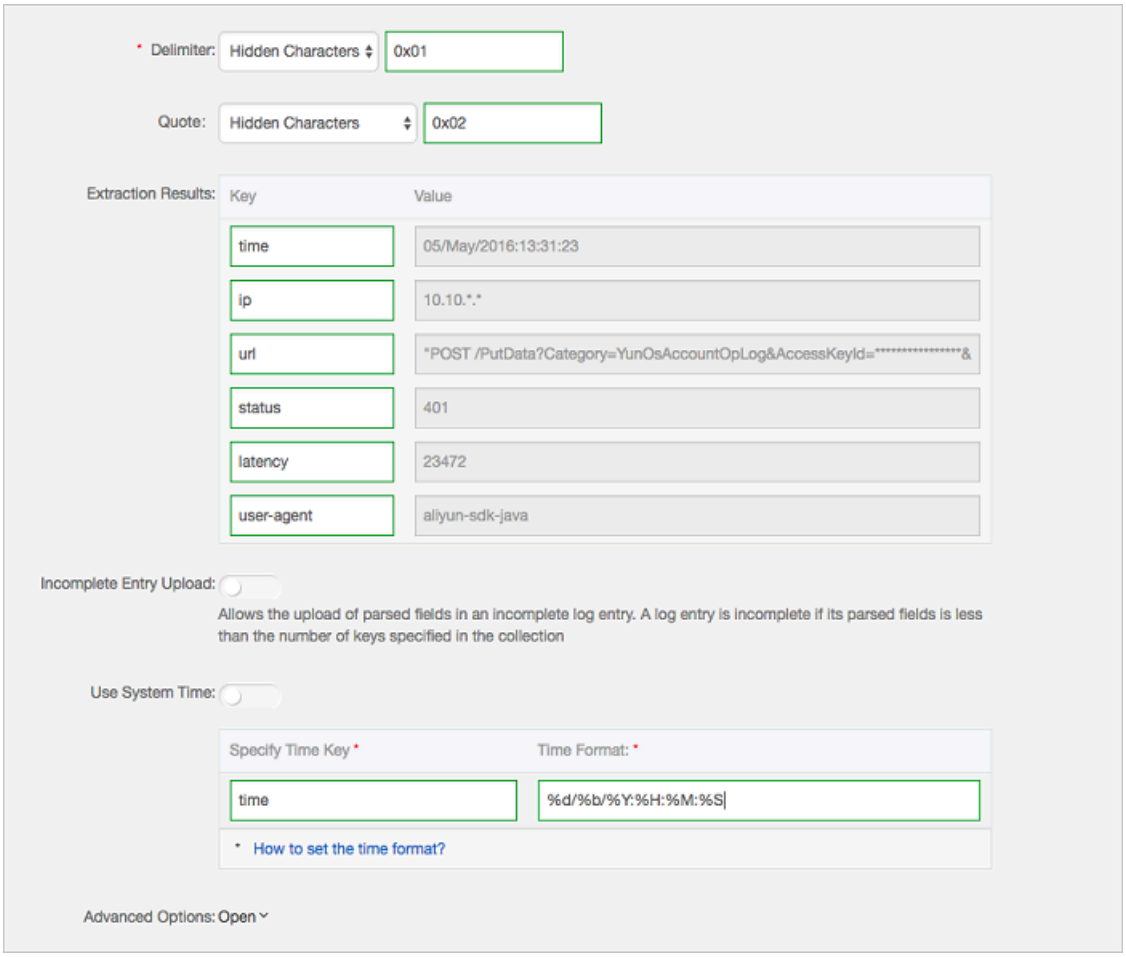
For example, if you set the delimiter to the vertical bar (|), the log sample `11 | 22 | 33 | 44 | 55` can be parsed into the following fields: `11`, `22`, `33`, `44`, and `55`. You can set their keys to `A`, `B`, `C`, `D`, and `E`, respectively. If you enable Incomplete Entry Upload and Log Service collects the log `11 | 22 | 33 | 55`, the `55` field is uploaded as the value of the `D` key.

If you disable Incomplete Entry Upload, Log Service discards the log because the fields and keys do not match.

e. Specify the log time.

You can use the system time or a log field (for example, the time field 05/May/2016:13:30:29) as the log time. For more information about how to configure the time format, see #unique_100.

Figure 5-5: Specify the log time



f. Preview logs in the console to confirm whether logs are collected.

5.7.8 JSON logs

JSON logs are constructed in two structures:

- Object: A collection of key/value pairs.
- Array: An ordered list of values.

Logtail supports JSON logs of the object type. Logtail automatically extracts the keys and values from the first layer of an object as the names and values of fields

respectively. The field value can be the object, array, or basic type, for example, a string or number. `\ n` is used to separate the lines of JSON logs. Each line is extracted as a single log.

Logtail does not support automatic parsing of non-object data (for example, JSON arrays). You can use regular expressions for field extraction or use the simple mode for log collection by line.

Log sample

```
{" url ": " POST  / PutData ?  Category = YunOsAccou  ntOpLog &
 AccessKeyI  d = U0Ujpek ********& Date = Fri % 2C % 2028 % 20Jun
 % 202013 % 2006 % 3A53 % 3A30 % 20GMT & Topic = raw & Signature =
 pD12XYLmGx  KQ % 2Bmkd6x7hA  gQ7b1c % 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 = YunOsAccou  ntOpLog &
 AccessKeyI  d = U0Ujpek ********& Date = Fri % 2C % 2028 % 20Jun
 % 202013 % 2006 % 3A53 % 3A30 % 20GMT & Topic = raw & Signature =
 pD12XYLmGx  KQ % 2Bmkd6x7hA  gQ7b1c % 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 "}
```

Configure Logtail to collect JSON logs

For the complete process of collecting JSON logs by using Logtail, see #unique_272. This document shows the detailed configuration Log Collection Mode of Logtail.

1. On the Logstore List, click the Data Import Wizard.
2. Select the data type.

   Select the text file and click Next.

3. Configure the data source.
   a. Enter the configuration name, Log Path, and select log collection mode as JSON mode.
   b. Select whether to use the system time as the log time according to your requirements. You can enable or disable the Use System Time function.
      - Enable Use System Time function

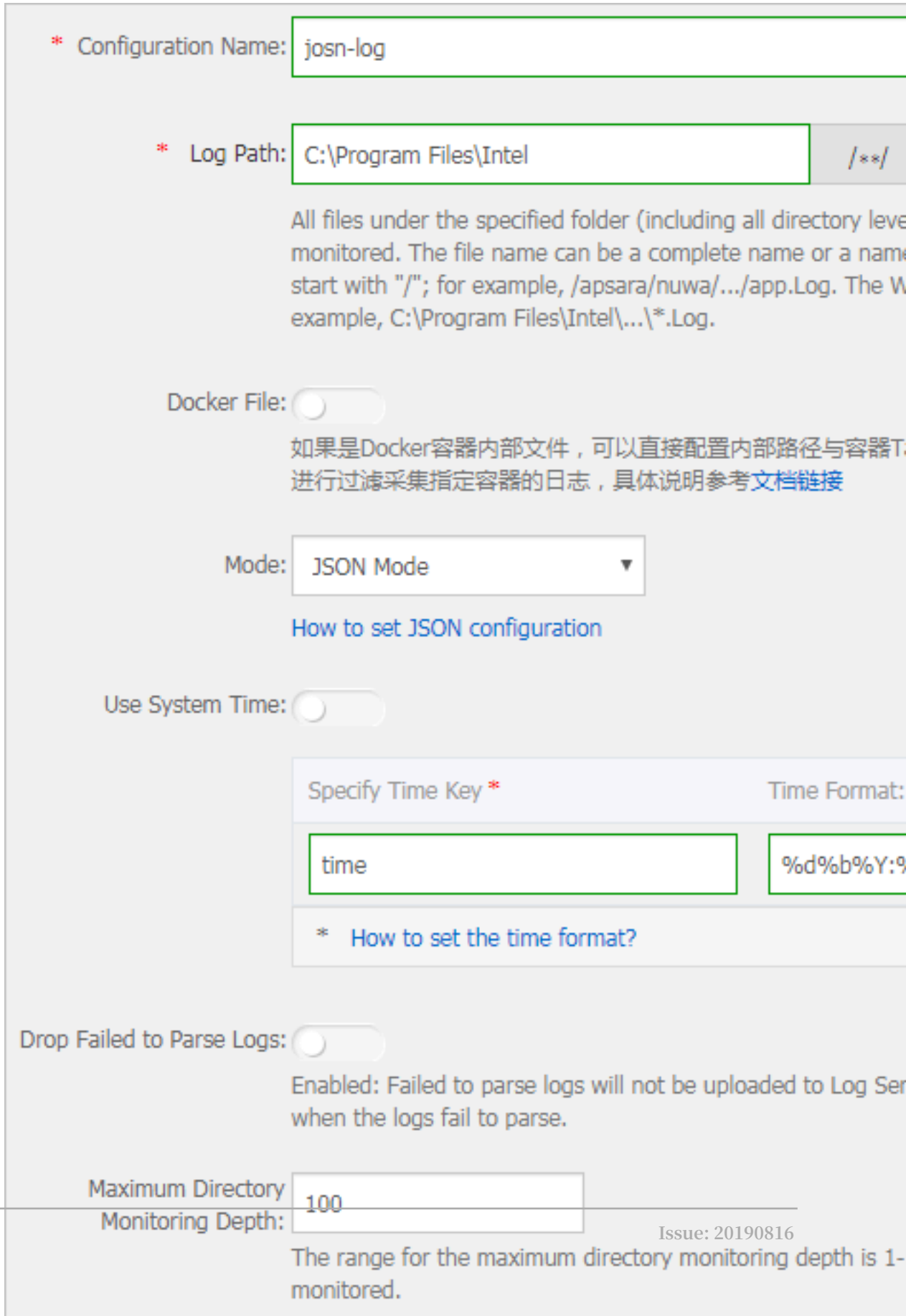
        Enabling this function means to use the time when Log Service collects the log as the log time, instead of extracting the time fields in the log.
      - Disable Use System Time function

        Disabling this function means to extract the time fields from the log as the log time.

        If you select to disable the Use System Time function, you must define the key of the extracted time field, and the time conversion format. For example, the

`time` field (05/May/2016:13:30:29) in JSON Object can be extracted as log time. For how to configure the date format, see #unique_100.

Figure 5-6: JSON logs



5.7.9 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
 Think \ Log :: record (' D   method   instantiat  ion   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 \ BuildLiteB  ehavior  [  RunTime : 0 .
 000014s  ]
 INFO : [  app_init  ] -- END -- [  RunTime : 0 . 000091s  ]
 Info : [ app_begin ] --  start  --
 INFO :  Run   Behavior \ ReadHtmlCa  cheBehavio  r  [  RunTime : 0 .
 000038s  ]
 INFO : [  app_begin  ] -- END -- [  RunTime : 0 . 000076s  ]
 INFO : [  view_parse  ] -- START --
 INFO :  Run   Behavior \ ParseTempl  ateBehavio  r  [  RunTime : 0 .
 000068s  ]
 INFO : [  view_parse  ] -- END -- [  RunTime : 0 . 000104s  ]
 INFO : [  view_filte  r  ] -- START --
 INFO :  Run   Behavior \ WriteHtmlC  acheBehavi  or  [  RunTime : 0 .
 000032s  ]
 INFO : [  view_filte  r  ] -- END -- [  RunTime : 0 . 000062s  ]
 INFO : [  app_end  ] -- START --
 INFO :  Run   Behavior \ ShowPageTr  aceBehavio  r  [  RunTime : 0 .
 000032s  ]
 INFO : [  app_end  ] -- END -- [  RunTime : 0 . 000070s  ]
 ERR :  D   method   instantiat  ion   does   not   find   the   model
   class
```

Configure Logtail to collect ThinkPHP logs

For the complete process of collecting ThinkPHP logs by using Logtail, see #unique_275. 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 +)[^:]+:\ d +\ s ]\ s +(\ S +)\ s
 (\ S +)\ s +(.
```

Time expression:

```
% Y -% m -% dT % H :% M :% S
```

5.7.10 Use Logstash to collect IIS logs

You need to modify the configuration file to parse the IIS log fields before you use logsturg 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_posi  tion  => " beginning "
  }
}
 filter  {
   if  [ type ] == " iis_log_1 " {
  # ignore   log   comments
   if  [ message ] =~ "^#" {
     drop  {}
  }
   grok  {
    #  check   that   fields   match   your   IIS   log   settings
     match  => [" message ", "%{ TIMESTAMP_  ISO8601 : log_timest
 amp } %{ IPORHOST : site } %{ WORD : method } %{ URIPATH : page } %{
 NOTSPACE : querystrin  g } %{ NUMBER : port } %{ NOTSPACE : username
 } %{ IPORHOST : clienthost } %{ NOTSPACE : useragent } %{ NUMBER
 : response } %{ NUMBER : subrespons  e } %{ NUMBER : scstatus } %{
 NUMBER : time_taken }"]
  }
     date  {
```

```
      match  => [ " log_timest  amp ", " YYYY - MM - dd   HH : mm : ss
 " ]
        timezone  => " Etc / UTC "
  }
   useragent  {
     source => " useragent "
     prefix => " browser "
  }
   mutate  {
     remove_fie  ld  => [ " log_timest  amp "]
  }
  }
}
 output  {
   if  [ type ] == " iis_log_1 " {
   logservice  {
         codec  => " json "
         endpoint  => "***"
         project  => "***"
         logstore  => "***"
         topic  => ""
         source  => ""
         access_key  _id  => "***"
         access_key  _secret  => "***"
         max_send_r  etry  =>  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 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 grok.

Restart Logstash to apply configurations

Create a configuration file in the *conf* directory and restart Logstash to apply the file. See #unique_259 for more information.

5.7.11 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

```
10 . 116 . 14 . 201 ,-, 2 / 25 / 2016 , 11 : 53 : 17 , W3SVC7
, 2132 , 200 , 0 , GET , project / shenzhen - test / logstore /
logstash / detail , C :\ test \ csv \ test_csv . log
```

- Collection configuration

```
 input  {
   file  {
     type  => " csv_log_1 "
     path  => [" C :/ test / csv /*. log "]
     start_posi  tion  => " 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_r  etry  =>  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 #unique_259 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 , W3SVC7
, 1332 , 200 , 0 , GET , project / shenzhen - test / logstore /
logstash / detail , C :\ test \ csv \ test_csv_w  ithtime . log
```

- Collection configuration

```
input  {
  file  {
    type  => " csv_log_2 "
    path  => [" C :/ test / csv_withti  me /*. log "]
    start_posi  tion  => " 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_r  etry  =>  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 #unique_259 as a Windows service.

5.7.12 Use Logstash to collect other logs

You can modify the configuration file to parse log fields before you use logsturg 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
2016 - 02 - 25   15 : 37 : 11  [ main ]  ERROR   com . aliyun . sls
. test_log4j  -  catch   exception  !
 java . lang . Arithmetic  Exception : /  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 /:?]
2016 - 02 - 25   15 : 38 : 02  [ main ]  INFO   com . aliyun . sls
. test_log4j  -  normal   log
```

- Collection configuration

```
input  {
  file  {
    type  => " common_log  _1 "
    path  => [" C :/ test / multiline /*. log "]
    start_posi  tion  => " beginning "
    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_r  etry  =>  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 #unique_259.

5.7.13 Unity3D logs

Context

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#unique_8 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 #unique_8.

2. Register Unity3D LogHandler

   Create a C# file `LogOutputH andler . 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 #unique_17.

```
using UnityEngin e ;
using System . Collection s ;
public class LogOutputH andler : MonoBehavi our
{
    // Register the HandleLog function on scene start
to fire on debug . log events
    public void OnEnable ()
    {
        Applicatio n . logMessage Received += HandleLog ;
    }
    // Remove callback when object goes out of scope
    public void OnDisable ()
    {
        Applicatio n . logMessage Received -= HandleLog ;
    }
    string project = " your project name ";
    string logstore = " your logstore name ";
    string serviceAdd r = " 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_Trac e =" + WWW . EscapeURL (
stackTrace );
```

```
        parameters  += "&";
       // Add   any   User ,  Game ,  or   Device   MetaData   that
  would   be   useful   to   finding   issues   later
        parameters  += " Device_Mod  el =" +  WWW . EscapeURL (
SystemInfo . deviceMode  l );
        string   url  = " http ://" +  project  + "." +
serviceAdd  r  + "/ logstores /" +  logstore  + "/ track ?
APIVersion = 0 . 6 . 0 &" +  parameters ;
        StartCorou  tine ( SendData ( url ));
    }
     public   IEnumerato  r   SendData ( string   url )
    {
        WWW   sendLog  =  new   WWW ( url );
        yield   return   sendLog ;
    }
}
```

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   UnityEngin  e ;
using   System . Collection  s ;
public   class   LogglyTest  :  MonoBehavi  our  {
    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 . LogExcepti on* . 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.