

Alibaba Cloud Log Service

Quick Start

Issue: 20190412

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 <code>cd / d C :/ windows</code> command to enter the Windows system folder.
<i>Italics</i>	It is used for parameters and variables.	<code>bae log list --instanceid Instance_ID</code>
[] or [a b]	It indicates that it is an optional value, and only one item can be selected.	<code>ipconfig [-all -t]</code>

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

Contents

Legal disclaimer.....	I
Generic conventions.....	I
1 5-minute quick start.....	1
2 ECS ログの収集.....	12
3 Collect Kubernetes logs.....	18
4 Access - Log4j/Logback/Producer Lib.....	24
5 Analysis - Nginx access logs.....	31
6 Analyze IIS access logs.....	47
7 Analysis - Apache access logs.....	56

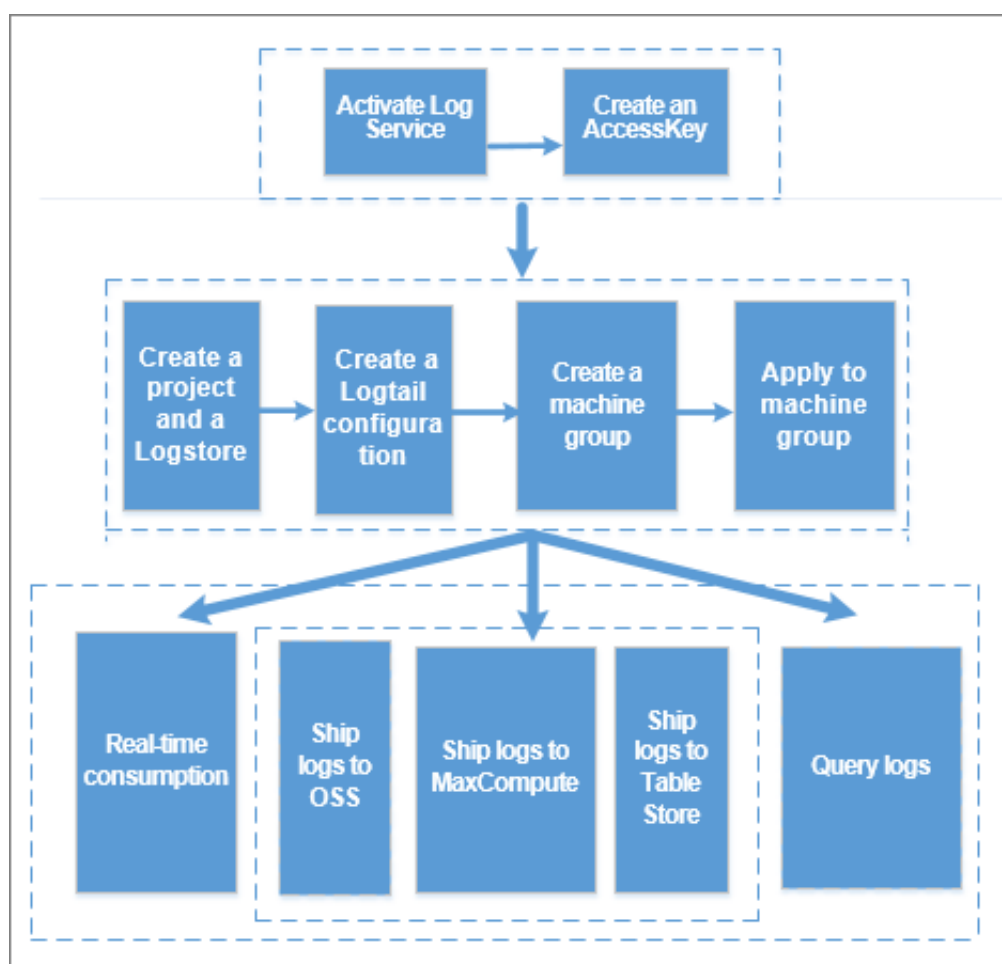
1 5-minute quick start

Log Service is a platform provided by Alibaba Cloud for collecting, storing, and querying massive logs. You can use Log Service to centrally collect all the logs from the service cluster. It also supports real-time consumption and query.

This document demonstrates the basic workflow of configuring Logtail to collect Alibaba Cloud Elastic Compute Service (ECS) logs in the Windows environment. This case is related to the basic functions of Log Service, such as collecting logs and querying logs in real time, and is an entry-level user guide of Log Service.

Log service operation process

Figure 1-1: Procedures



Step 1. Getting started

1. Activate Log Service

Use a registered Alibaba Cloud account to log on to the Log Service product page and click Get it Free.

2. Create an AccessKey (optional)



Note:

If you want to write data using SDK, create a primary account or sub-account AccessKey. Log collection does not require the creation of AccessKey.

In the Log Service console, hover your mouse over your avatar in the upper-right corner and click accesskeys in the displayed drop-down list. In the dialog box, click Continue to manage AccessKey to go to the Access Key Management page. Then, create an AccessKey. Make sure the status is set to Enabled. Then, create an AccessKey. Make sure the status is set to Enabled .

Figure 1-2: Enable AK

Access Key Management (5)					Refresh	Create Access Key
ⓘ Access Key ID and Access Key Secret are the API keys for you to access Aliyun. It has full access privilege of the account. Please keep it safe.						
Access Key ID	Access Key Secret	Status	Create Time	Action		
LTAD*****	Show	Enabled	2018-02-26 15:49:00	Disable	Delete	

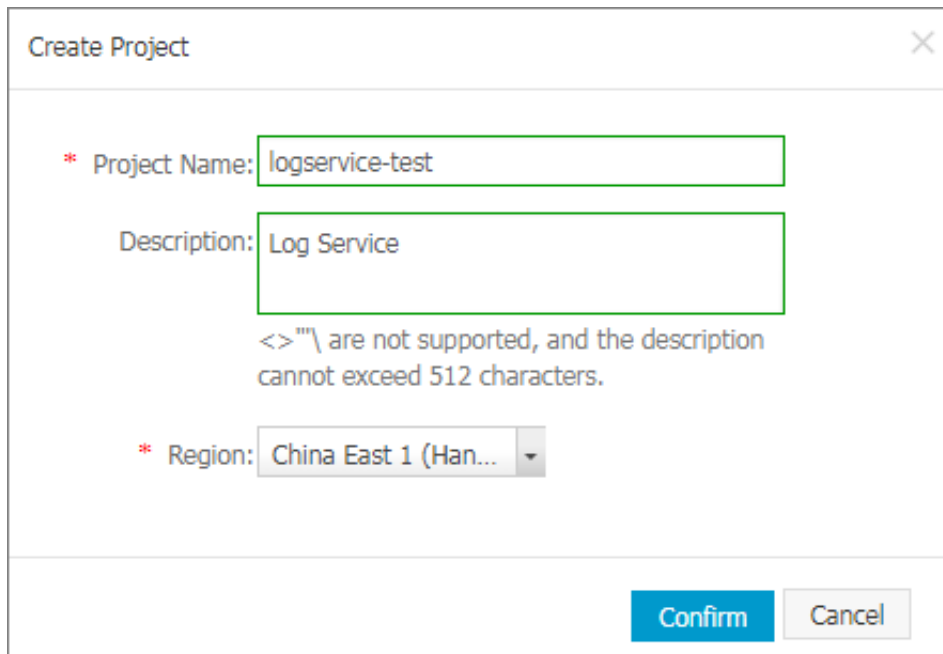
3. Create a project

If you have logged on to the Log Service console for the first time, the system prompts you to create a project. You can also click Create Project in the upper-right corner to create a project.

When creating a project, you must specify the Project Name and Region based on your actual needs. Among the regions, cn-shanghai-internal-prod-1 and cn-hangzhou-

internal-prod-1 are used for internal Log Service, while the other regions are in the public cloud.

Figure 1-3: Create a project



The image shows a 'Create Project' dialog box with a close button (X) in the top right corner. It contains three main input fields, each preceded by a red asterisk (*):

- Project Name:** A text input field containing 'logservice-test'.
- Description:** A text input field containing 'Log Service'. Below this field is a warning message: '<>""\ are not supported, and the description cannot exceed 512 characters.'
- Region:** A dropdown menu showing 'China East 1 (Han...' with a downward arrow.

At the bottom right of the dialog box are two buttons: a blue 'Confirm' button and a grey 'Cancel' button.

4. Create a Logstore

After creating a project, you is be prompted to create a Logstore. You can also go to the project and click Create in the upper-right corner. When creating a Logstore, you must specify how you are going to use these logs.

Figure 1-4: Creating a Logstore

Create Logstore

* Logstore Name: test

Logstore Attributes

* WebTracking: ☐ WebTracking supports the collection of various types of access logs in web browsers or mobile phone apps (iOS/Android). By default, it is disabled. ([Help Link](#))

* Data Retention Time: 30 Data retention time for LogHub and LogSearch is unified. The data lifecycle is determined by the LogHub setting (the unit is in days).

* Number of Shards: 2 [What is shard?](#)

* Billing: [Refer to pricing](#)

Confirm Cancel

Step 2. Install Logtail client on ECS instance

1. Download the installation package

Download the Logtail installation package to an ECS instance. Click [here](#) to download the Windows installation package.

2. Install Logtail

Extract the installation package to the current directory and then enter the `logtail_in staller` directory. Run cmd as an administrator, and run the

installation command `.\logtail_installer.exe install cn_hangzhou`.

**Note:**

You must run different installation commands according to the network environment and the region of Log Service. This document uses the ECS classic network in China East 1 (Hangzhou) as an example. For other areas, see [Install Logtail in Windows](#).

For the installation commands of other regions, see [Install Logtail in Windows](#) and [Install Logtail in Linux](#).

Step 3. Configure data import wizard

In the Log Service console, click the project name to go to the Logstore List page. Click 1 at the right of the Logstore to enter the Logtail configuration. You can also click Manage at the right of the Logstore to create a configuration in the Logtail configuration list.

Logtail configuration process includes the following steps: Select Data Source, Configure Data Source, Search, Analysis, and Visualization, Shipper & ETL. The last two steps are optional.

1. Select data source

Log Service supports the log collection of many cloud products, self-built softwares, and custom data. This document uses collecting text logs as an example. For more information, see [Collect text logs](#).

Click Text in Other Sources, and then click Next.

2. Configure data source

- Specify the Configuration name and Log path.

Follow the page prompts to enter the configuration name, log path, and log file name. The log file name can be a full name, and supports wildcard matching at the same time.

- Specify the log collection mode.

Log Service currently supports parsing logs in simple mode, delimiter mode, JSON mode, full mode, or Alibaba Cloud custom mode. This document uses the delimiter

mode as an example. For more information about the collection modes, see [Collect text logs](#) and [Configure and parse text logs](#).

Figure 1-5: Configure the data source

The screenshot shows a configuration window for a data source. It includes the following fields and options:

- Configuration Name:** A text field containing "logservice_test".
- Log Path:** A text field containing "C:\Program Files\" followed by a file pattern field containing "/*.*/*" and a file extension field containing "*.Log". Below this, a descriptive text explains that all files under the specified folder will be monitored, and provides examples for Linux and Windows paths.
- Docker File:** A toggle switch that is currently turned off. Below it, a text description explains that if the file is in a Docker container, Logtail will monitor the container's lifecycle and collect logs based on a specified label.
- Mode:** A dropdown menu currently set to "Delimiter Mode". Below it is a link: "How to set the Delimiter type configuration".
- Log Sample:** A large text area for entering log samples. Below it is a link: "Common Samples>>".
- Delimiter:** A dropdown menu currently set to "Tabs".

- Enter the log sample.

You must enter the log sample if Delimiter Mode or Full Mode is selected as the log collection mode. Log Service supports parsing the log sample according to selected configuration when configuring Logtail. If the log sample failed to be parsed, you must modify the delimiter configurations or regular expressions. Enter the log sample to be parsed in the Log Sample field.

- Specify the delimiter.

You can specify the delimiter as a tab, a vertical line, or a space. You can also customize the delimiter. Select the delimiter according to your log format. Otherwise, logs fail to be parsed.

- Specify the key in the log extraction results.

After you enter the log sample and select the delimiter, Log Service extracts log fields according to your selected delimiter, and defines them as Value. You must specify the corresponding Key for the Value.

Figure 1-6: Log content extraction results

* Extraction Results:	Key	Value
	ip	1.1.1.1
	time	[10/Apr/2017:21:28:23 +0800
	method	GET
	useragent	/test HTTP/1.1" 0.282 511 200 55 "" "Httpful/0.2.1.0 (eURL/7.15.5 PHP/

- Configure the advanced options as needed.

Generally, keep the default configurations of the advanced options. For how to configure the advanced options, see the related descriptions in [Collect text logs](#).

- Apply to the machine group.

If you have not created a machine group before, create a machine group according to the page prompts. Then, apply the Logtail configuration to the machine group.



Note:

To create Armory to associate with the machine group, jump to the specified internal link as instructed on the page.

After completing these steps, Log Service begins to collect logs from the Alibaba Cloud ECS instance immediately. You can consume the collected logs in real time in the console and by using API/SDK.

To query, analyze, ship, or consume the logs, click Next.



Note:

- It can take up to 3 minutes for the Logtail configuration to take effect.

- To collect IIS access logs, see [Use Logstash to collect IIS logs](#).
- For the Logtail collection errors, see [Query diagnosed errors](#).

Search, analysis, and visualization

After the collection is configured, your ECS logs are collected in real time. To query and analyze the collected logs, configure the indexes in the data import wizard as follows.

You can click Search on the Logstore List page to go to the query page. Click Enable in the upper-right corner and configure the indexes on the displayed Search & Analysis page.

- Full text index attributes

You can enable the Full Text Index Attributes. Confirm whether to enable Case Sensitive, and confirm the Token contents.

- Key/value index attributes

Click the plus icon at the right of Key to add a line. Configure the Key, Type, Alias, Case Sensitive, and Token, and select whether to enable analytics.



Note:

1. Full text or key/value indexes attributes at least one must be enabled. When both types are enabled, key/value index attributes prevail.
2. When the index type is long or double, the Case Sensitive and Token attributes are unavailable.
3. For how to configure indexes, see [Overview](#).

4. To use Nginx template or MNS template, configure the attributes on the Search & Analysis page after clicking Enable on the query page.

Figure 1-7: query analysis

custom

* Full Text Index Attributes:

Case Sensitive

Token

false

, ""'=()[]{}?@&<>/: \n\t

* Key/Value Index Attributes:

Key +	Type	alias	Case Sensitive	Token
requests	long	requests		
reading	long	reading		
connection	long	connection		
_response	double	_response		
waiting	long	waiting		
_method	text	_method	false	

1. Full text index and Key/Value index cannot be disabled at the same time.

2. When the index type is long or double, the Case Sensitive and Token attributes are not av

3. For how to set index attributes, refer to the document ([Help Link](#))

After configuring the query and analysis, click Next if you want to configure the log shipping. To experience the query and analysis, go back to the Logstore List page and click Search to go to the query page. You can enter the keyword, topic, or query & analysis statement, and select the time range to query logs. Log Service provides

intuitive histograms to preview the query results. You can click the histogram to query logs in a more detailed time range. For more information, see [Overview](#).

Log Service also supports querying and analyzing logs in many ways such as quick query and statistical graphs. For more information, see [Other functions](#).

For example, to query all the logs within the last 15 minutes, you can set an empty query condition and select 15 min as the time range.

4. Shipping

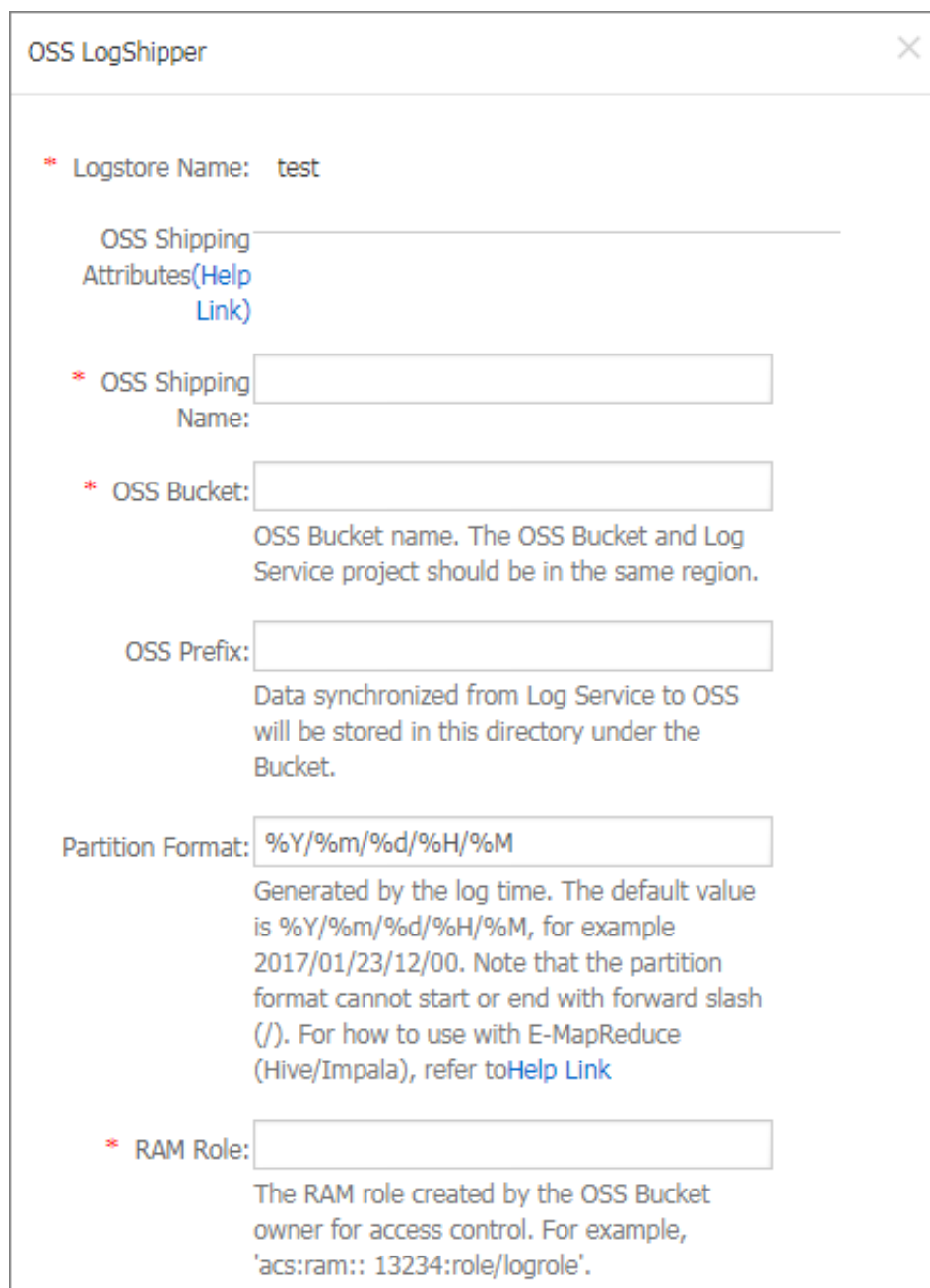
Log Service not only supports collecting data with multiple sources and formats in batch, managing and maintaining the data, but also supports shipping log data to cloud products such as Object Storage Service (OSS) for calculation and analysis.

To ship logs to OSS, click Enable.

This document uses OSS storage as an example. See [Ship logs to OSS](#) to complete the authentication.

Click Enable, the OSS LogShipper dialog box appears. For descriptions about the configurations, see [Ship logs to OSS](#). After the configuration is complete, click Confirm to complete the shipping.

Figure 1-8: Configure shipping



The screenshot shows the 'OSS LogShipper' configuration dialog box. It contains several fields and sections:

- Logstore Name:** test
- OSS Shipping Attributes:** A section with a [Help Link](#).
- OSS Shipping Name:** An empty text input field.
- OSS Bucket:** An empty text input field. Below it, a note states: 'OSS Bucket name. The OSS Bucket and Log Service project should be in the same region.'
- OSS Prefix:** An empty text input field. Below it, a note states: 'Data synchronized from Log Service to OSS will be stored in this directory under the Bucket.'
- Partition Format:** A text input field containing '%Y/%m/%d/%H/%M'. Below it, a note states: 'Generated by the log time. The default value is %Y/%m/%d/%H/%M, for example 2017/01/23/12/00. Note that the partition format cannot start or end with forward slash (/). For how to use with E-MapReduce (Hive/Impala), refer to [Help Link](#)'.
- RAM Role:** An empty text input field. Below it, a note states: 'The RAM role created by the OSS Bucket owner for access control. For example, 'acs:ram:: 13234:role/logrole'.'

Besides the basic functions such as accessing, querying, and analyzing logs, Log Service also provides many ways to consume logs. For more information, see [User Guide](#).

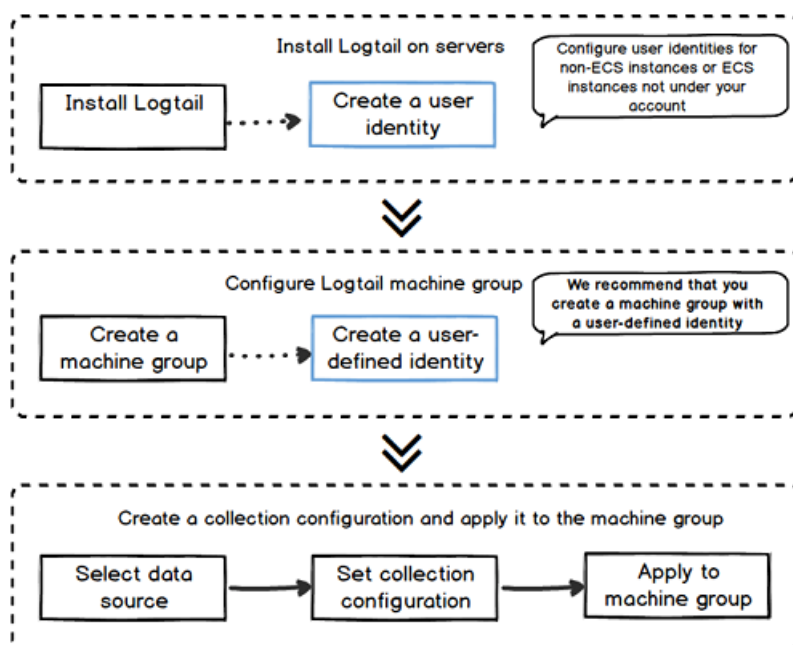
2 ECS ログの収集

本ドキュメントでは、ECS (Elastic Compute Service) の Log Service コンソールより、ログを収集する方法についてをます。。

設定手順

1. ECS インスタンスに Logtail をインストールします。
2. Logtail マシングループを構成します。
3. Configure the collection, and apply the configuration to the machine group.

Figure 2-1: Configuration procedure



前提要件

- ・ ECS および Log Service が有効であること。
- ・ Log Service でプロジェクトおよびログストアを作成していること。詳細については、[Preparation](#)をご参照ください。



Note:

Make sure that the Log Service Project and the ECS instance are located in the same region if the ECS instance runs in the classic network or virtual private cloud (VPC).

- ・ Your Log Service cannot automatically obtain owner information of the ECS if the Log Service and the ECS instance do not belong to the same account. You need to [Configure AliUids for ECS servers under other Alibaba Cloud accounts or on-premises IDCs](#).

Step 1: Logtail をインストール

1. インストールコマンドを実行します。

ECS インスタンスのリージョンに合わせて、Logtail インストールスクリプトを選択します。詳細については、[Install Logtail in Linux](#) または [Install Logtail in Windows](#) をご参照ください。

たとえば、中国 (杭州) リージョンの Linux インスタンスに Logtail をインストール場合は、次のコマンドを実行します。

```
wget http://logtail-release.oss-cn-hangzhou-internal.aliyuncs.com/linux64/logtail-linux64.tar.gz
chmod 755 logtail.sh ; sh logtail.sh install cn_hangzhou
```

2. Logtail の運用ステータスを確認します。

Logtail の運用ステータスを確認し、echo に `ilogtail is running` と記述されていれば、無事にインストールされたことになります。Logtail の運用ステータスを確認するには、次のコマンドを実行します。

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

Figure 2-2: Logtail のインストール



```
--2018-05-24 11:10:58-- http://logtail-release.oss-cn-hangzhou-internal.aliyuncs.com/linux64/logtail-linux64.tar.gz
正在解压缩 logtail-release.oss-cn-hangzhou-internal.aliyuncs.com (logtail-release.oss-cn-hangzhou-internal.aliyuncs.com)... 100.118.28.49, 100.118.28.50, 100.118.28.45, ...
正在连接 logtail-release.oss-cn-hangzhou-internal.aliyuncs.com (logtail-release.oss-cn-hangzhou-internal.aliyuncs.com)100.118.28.49:80... 已连接。
已发出 HTTP 请求，正在等待响应... 200 OK
长度: 8628916 (8.2M) [application/x-gzip]
正在保存至: "logtail-linux64.tar.gz"
100%[=====] 8,628,916 46.2MB/s 用时 0.2s

2018-05-24 11:10:58 (46.2 MB/s) - 已保存 "logtail-linux64.tar.gz" [8628916/8628916]

logtail-linux64.tar.gz download success
install logtail files success
注意: 正在将请求转发到 "systemctl enable ilogtailed.service".
Created symlink from /etc/systemd/system/multi-user.target.wants/ilogtailed.service to /etc/systemd/system/ilogtailed.service.
chkconfig add ilogtailed success
install logtail success
ilogtail is running
start logtail success
{
  "UID": "2281f10c-1111-4c78-b88d-70c7c2a8d111",
  "hostname": "100.118.28.49",
  "instance_id": "100.118.28.49-100.118.28.49-100.118.28.49",
  "ip": "10.10.10.10",
  "logtail_version": "0.10.0",
  "os": "Linux: 3.10.0-112.el7.x86_64 #1 SMP Tue Jul 4 15:04:05 UTC 2017; x86_64",
  "update_time": "2018-05-24 11:10:58"
}
```

Step 2 Configure the machine group.

1. Log Service コンソールのホームページで、プロジェクト名をクリックし、Logstore List ページに移動します。
2. 左のナビゲーションペインより、Logtail Machine GroupをクリックしてMachine Groupsページに移動します。

3. Click **Create Machine Group** in the upper-right corner, and enter the intranet IP or custom identifier of your ECS instance in the dialog box that appears, and then click **Confirm**.

**Note:**

- Log Service only supports ECS instances that run in the same region as the current Project.
- Windows and Linux ECS instances cannot exist in the same machine group at the same time.

Figure 2-3: Configuring the machine group

Create Machine Group

* Group Name: demo-group

Machine Group Identification: IPs

Machine Group Topic:

* IPs: 10.1.1.1

1. Only machines in the same region as the current project are supported.
2. Enter the ECS intranet IP addresses; each IP address should occupy one row.
3. Windows machines and Linux machines cannot be in the same machine group. ([Help Link](#))
4. Logtail requires a pair of valid primary account AK, please [log in console](#) to ensure the effectiveness, otherwise it will cause the heartbeat failure and other issues

Confirm Cancel

Step 3: Configure collection.

1. Enter the data access wizard.

On the Logstore List page, click the icon in the Data Import Wizard column.

2. Select the data type.

At the Select Data Source step, click **Text**, and click **Next**.

3. Set the data source.

For more information about setting the data source, see [Collect text logs](#). To set the data source in Simple Mode, follow these steps:

Enter the path of the ECS logs, and click Next.

Figure 2-4: Basic mode

1.Select Data Source 2.Configure Data Source 3.Search, Analysis, and Visualization 4.Shipper & ETL

Configuration Name: demo-config

Log Path: /var/log / **/ message

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

Docker File: ☐

If the file is in the docker container, you can directly configure the internal path and container label, Logtail will automatically monitor the create and destroy of the container, and collect the log of the specified container according to specified label

Mode: Simple Mode

Reminder : In Simple Mode, each line is treated as one log. The system will not extract the fields in the logs, and the parse time is used as the log time.

Advanced Options: Open

Previous Next

4. 設定をマシングループに適用します。

step 2で作成したマシングループを選択し、Apply to Machine Groupをクリックします。

Figure 2-5: マシングループに適用

1.Select Data Source 2.Configure Data Source 3.Search, Analysis, and Visualization 4.Shipper & ETL

Apply to Machine Group

+ Create Machine Group

☒ demo-group ☐ k8s-group-c12blasdfg423345y2

Apply to Machine Group

Now you have configured the machine group to collect ECS logs by using Logtail. To create indexes for collected logs and configure log shipping, continue with the follow-up steps.

ログを表示

設定が完了したら、ECS インスタンスにログインし直るか、or run 「echo "test message" >> / var / log / message », to generate logs in the local / var / log / message file. Then, Logtail collects these logs and stores them to Log Service.

Logstore List ページで、Search または Preview をクリックし、Logtail によって収集されたログを表示します。

Figure 2-6: ログの表示

Logstore List

Endpoint List

Create

Searching by logstore name

Search

Logstore Name	Data Import Wizard	Monitor	Log Collection Mode	Log Consumption Mode			Action
				LogHub	LogShipper	LogSearch	
demo-logstore			Logtail Config (Manage) Diagnose More Data	Preview	OSS	Search	Modify Delete
k8s-stdout			Logtail Config (Manage) Diagnose More Data	Preview	OSS	Search	Modify Delete

Total: 2 item(s),Per Page: 10 item(s)

<<

<

1

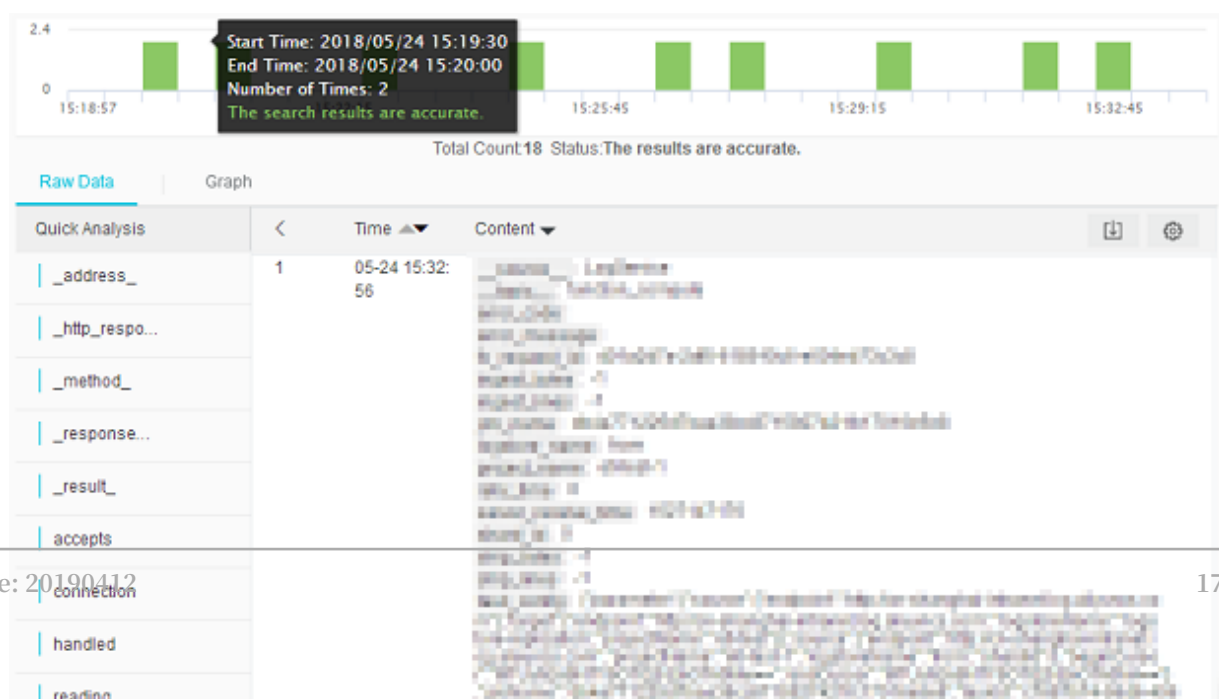
>

>>

Figure 2-7: Previewing logs

Back to Logstore List	
Shard: 0	15 min Preview
Log preview is only used to check whether log data is uploaded successfully. If you want to search logs through keywords, enable log index.	
Time/IP	Content
2018-05-24 15:19:30	<pre>get_logstore_name: demo-logstore, shard_id: 0, log_id: 1, log_offset: 0, log_size: 1024, log_type: 1, log_status: 1, log_content: [{}], log_time: 2018-05-24 15:19:30, log_ip: 10.10.10.10, log_host: 10.10.10.10, log_port: 80, log_protocol: http, log_method: GET, log_path: /, log_query: {}, log_body: {}, log_meta: {}, log_extra: {}]</pre>
2018-05-24 15:20:00	<pre>get_logstore_name: demo-logstore, shard_id: 0, log_id: 2, log_offset: 0, log_size: 1024, log_type: 1, log_status: 1, log_content: [{}], log_time: 2018-05-24 15:20:00, log_ip: 10.10.10.10, log_host: 10.10.10.10, log_port: 80, log_protocol: http, log_method: GET, log_path: /, log_query: {}, log_body: {}, log_meta: {}, log_extra: {}]</pre>

Figure 2-8: Retrieving logs



3 Collect Kubernetes logs

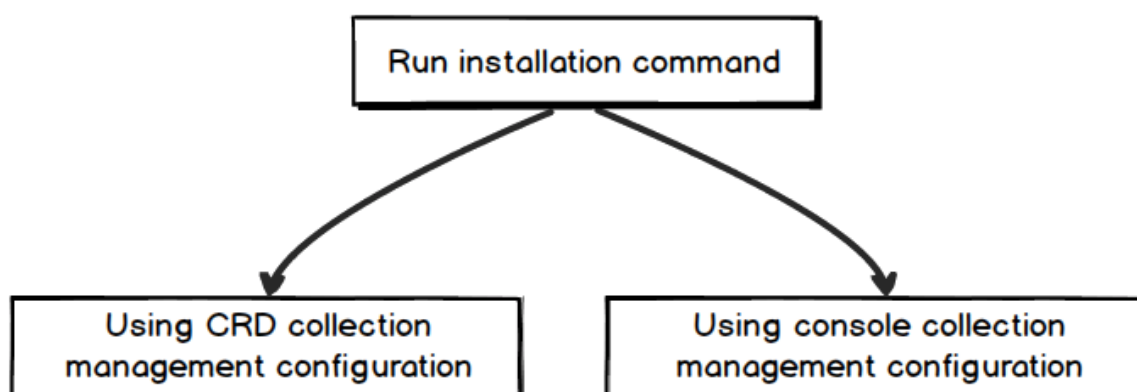
Log Service enables Logtail to collect Kubernetes cluster logs, and uses the CustomResourceDefinition (CRD) API to manage collection configurations. This document describes how to install and use Logtail to collect Kubernetes cluster logs.

Collection procedure

1. Install the alibaba-log-controller Helm package.
2. Configure the collection.

You can configure the collection in the Log Service console or by using the CRD API as required. To configure the collection in the console, follow these steps:

Figure 3-1: Procedure



Step 1 Install the package.

1. Log on to the Master node of the Alibaba Cloud Container Service for Kubernetes.

For how to log in, see [Access Kubernetes clusters by using SSH key pairs](#).

2. Replace the parameters and run the following command.

`${ your_k8s_c luster_id }` to your Kubernetes cluster ID in the following installation command, and run this command:

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

Installation example

Run the installation command to obtain the following echo:

```
[ root @ iZbp ***** biaZ ~]# wget http://logtail-release.oss-cn-hangzhou.aliyuncs.com/linux64/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 / ClusterRoleBinding
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 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 / ServiceAccount
NAME SECRETS AGE
alibaba - log - controller 1 0s

==> v1beta1 / CustomResourceDefinition
NAME AGE
aliyunlogconfigs.log.alibabacloud.com 0s

==> v1beta1 / ClusterRole
alibaba - log - controller 0s
```

```
[ SUCCESS ] install helm package : alibaba - log - controller success .
```

You can use `helm status alibaba - log - controller` to check the current Pod status. The Running status indicates a successful installation.

Then, Log Service creates the project that is named starting with `k8s-log`. You can search for this project by using the `k8s-log` keyword in the Log Service console.

Step 2: Configure the collection.

To create Logstore and collect standard output (stdout) from all K8s containers, follow these steps:

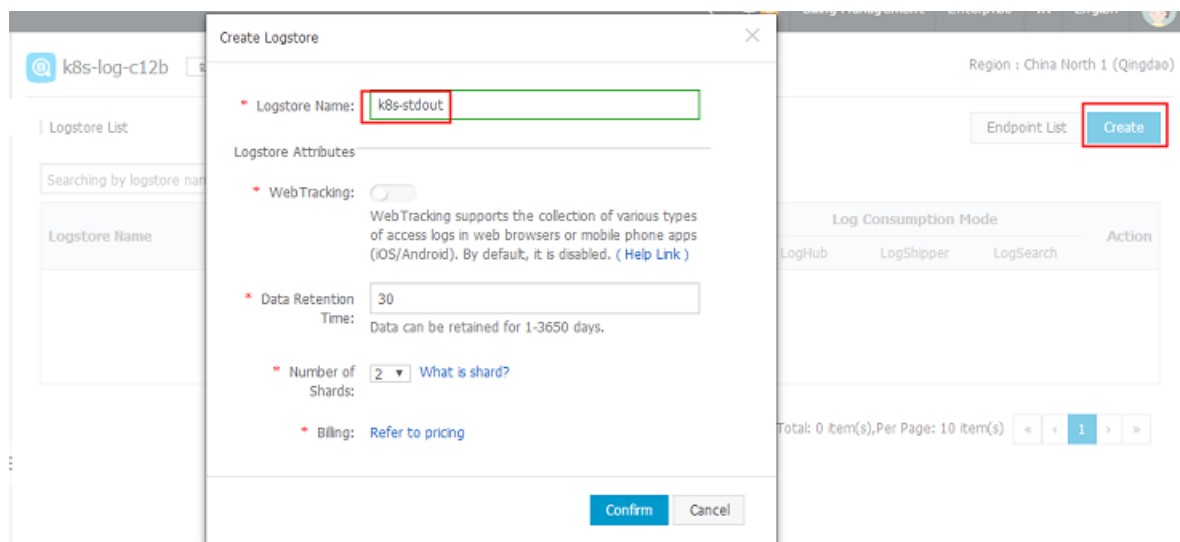
1. Go to the Logstore List page.

Click the project created in Step 1 to go to the Logstore List page.

2. Create Logstore.

Click **Create** in the upper-right corner, and in the dialog box that appears, create Logstore.

Figure 3-2: Creating Logstore

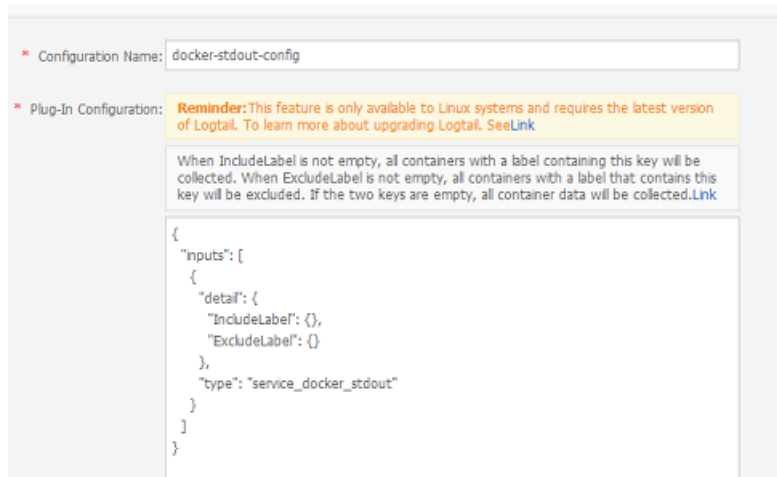


3. Configure the collection.

- a. Go to the Data Import Wizard page.
- b. Select Docker Stdout from Third-Party Software.

Click Apply to Machine Group on the configuration pages. Then, you can collect all stdout files from all containers.

Figure 3-3: Docker stdout

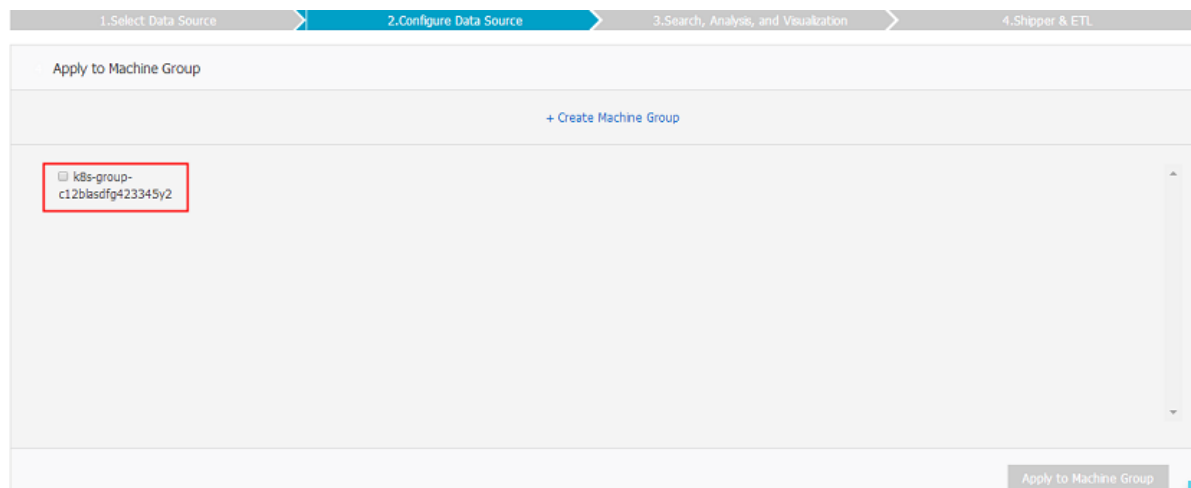


```
{
  "inputs": [
    {
      "detail": {
        "IncludeLabel": {},
        "ExcludeLabel": {}
      },
      "type": "service_docker_stdout"
    }
  ]
}
```

4. Apply the configuration to the machine group.

On the Apply to Machine Group page, select a machine group, and click Next.

Figure 3-4: Applying the configuration to the machine group



Now you have configured the collection. To configure indexes and log shipping, continue with the follow-up configurations. You can also exit the current page to complete the configuration.

Console Configuration

For more information about Console configuration, 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, set the log path on the configuration page to `/ logtail_ho st / home / logs / app_log /`.

CRD Configuration

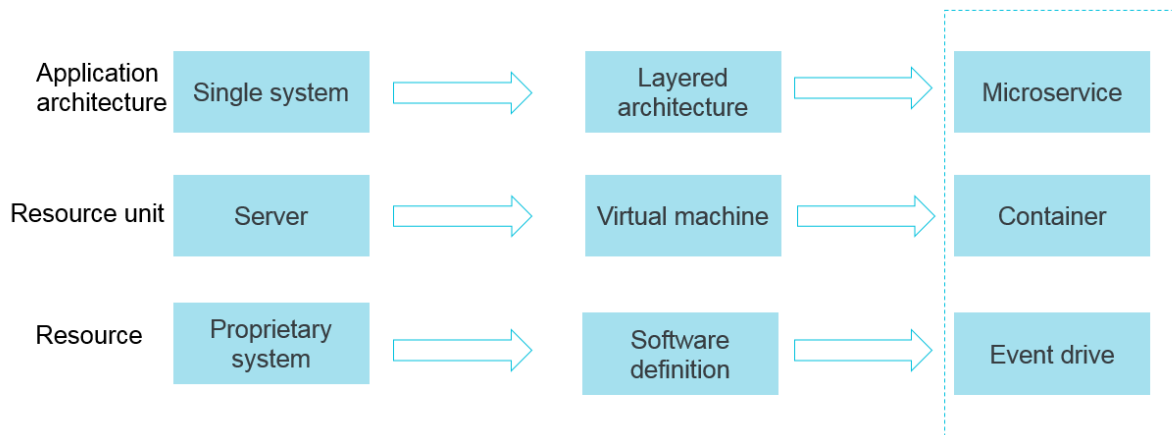
For more information about CRD(CustomResourceDefinition) configuration, see [Configure Kubernetes log collection on CRD](#).

4 Access - Log4j/Logback/Producer Lib

In recent years, the advent of stateless programming, containers, and serverless programming greatly increased the efficiency of software delivery and deployment. In the evolution of the architecture, you can see the following two changes:

- The application architecture is changing from a single system to microservices. Then, the business logic changes to the call and request between microservices.
- In terms of resources, traditional physical servers are fading out and changing to the invisible virtual resources.

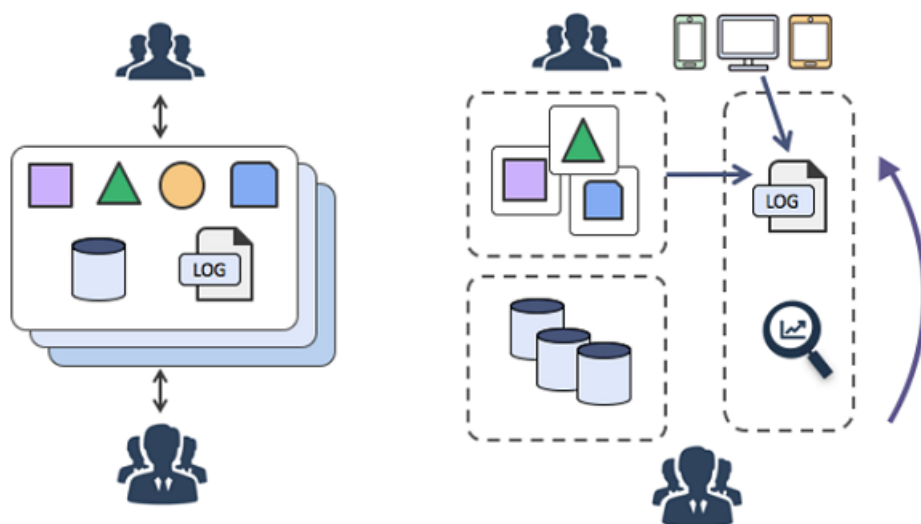
Figure 4-1: Architectural Evolution



The preceding two changes show that behind the elastic and standardized architecture, the Operation & Maintenance (O&M) and diagnosis requirements are becoming more and more complex. Ten years ago, you could log on to a server and

fetch logs quickly. However, the attach process mode no longer exists. Currently, we are facing with a standardized black box.

Figure 4-2: Changing Trends



To respond to these changes, a series of DevOps-oriented diagnosis and analysis tools have emerged. These include centralized monitors, centralized log systems, and various SaaS deployment, monitoring, and other services.

Centralizing logs solves the preceding issues. To do this, after applications produce logs, the logs are transmitted to a central node server in real time (or quasi-real time). Often, Syslog, Kafka, ELK, and HBase are used to perform centralized storage.

Advantages of centralisation

- **Ease of use:** Using Grep to query stateless application logs is troublesome. In the centralized storage, the previous long process is replaced by running a search command.
- **Separated storage and computing:** When customizing machine hardware, you do not have to consider the storage space for logs.
- **Lower costs:** Centralized log storage can perform load shifting to reserve more resources.

- **Security:** In case of hacker intrusion or a disaster, critical data is retained as the evidence.

Figure 4-3: Advantages of centralisation



Collector (Java series)

Log Service provides more than 30 data collection methods and comprehensive access solutions for servers, mobile terminals, embedded devices, and various development languages. Java developers need the familiar log frameworks: Log4j, Log4j2, and Logback Appender.

Java applications currently have two mainstream log collection solutions:

- Java programs flush logs to disks and use Logtail for real-time collection.
- Java programs directly configure the Appender provided by Log Service. When the program is running, logs are sent to Log Service in real time.

Differences between the two:

	Flush logs to disks + Use Logtail to collect logs	Use Appender for direct transmission
Timeliness	Logs are flushed to files and collected by using Logtail	Logs are directly sent to Log Service
Throughput	Big	Big
Resumable upload	Supported. Depends on the Logtail configuration	Supported. Depends on the memory size

	Flush logs to disks + Use Logtail to collect logs	Use Appender for direct transmission
Sensitive to application location	Required when configuring the collection machine group	Not required. Logs are initiatively sent
Local log	Supported	Supported
Disable collection	Delete Logtail configuration	Modify Appender configuration and restart the application

By using Appender, you can use Config to complete real-time log collection easily without changing any code. The Java-series Appender provided by Log Service has the following advantages:

- Configuration modifications take effect without modifying the program.
- Asynchrony + breakpoint transmission: I/O does not affect main threads and can tolerate certain network and service faults.
- High-concurrency design: Meets the writing requirements for massive logs.
- Supports context query: Supports precisely restoring the context of a log (N logs before and after the log) in the original process in Log Service.

Overview and usage of Appender

The provided Appenders are as follows. The underlying layers all use aliyun-log-producer-java to write data.

- [aliyun-log-log4j-appender](#)
- [aliyun-log-log4j2-appender](#)
- [aliyun-log-logback-appender](#)

Differences between the four:

Appender name	Description
aliyun-log-log4j-appender	Developed for Log4j 1.x. If your application uses the Log4j 1.x log framework, we recommend that you use this Appender.

Appender name	Description
aliyun-log-log4j2-appender	Developed for Log4j 2.x. If your application uses the Log4j 2.x log framework, we recommend that you use this Appender.
aliyun-log-logback-appender	Appender developed for logback, if your application is using logback Developed for Logback. If your application uses the Logback log framework, we recommend that you use this Appender.
aliyun-log-producer-java	The LogHub class library used for high-concurrency log writing, which is programmed for Java applications. All the provided Appender underlying layers use this Appender to write data. This highly flexible Appender allows you to specify the fields and formats of the data written to LogHub. If the provided Appender cannot meet your business needs, you can develop a log collection program based on this Appender as per your needs.

Step 1 Access Appender

Access the Appender by following the steps in [aliyun-log-log4j-appender](#).

The contents of the configuration file `log4j.properties` are as follows:

```
log4j . rootLogger = WARN , loghub
log4j . appender . loghub = com . aliyun . openservice . log .
log4j . LoghubAppender
# Log Service project name ( required parameter )
log4j . appender . loghub . projectName =[ your project ]
# Log Service LogStore name ( required parameter )
log4j . appender . loghub . logstore =[ your logstore ]
# Log Service HTTP address ( required parameter )
log4j . appender . loghub . endpoint =[ your project endpoint ]
#( Mandatory ) User identity
log4j . appender . loghub . accessKeyId =[ your accesskey id ]
log4j . appender . loghub . accessKey =[ your accesskey ]
```

Step 2 Query and analysis

After configuring the Appender as described in the previous step, the logs produced by Java applications are automatically sent to Log Service. You can use [LogSearch/](#)

[Analytics](#) to query and analyze these logs in real time. See the sample log format as follows. Log formats used in this example:

- Logs that record your logon behavior:

```
level : INFO
location : com . aliyun . log4jappen dertest . Log4jAppen
derBizDemo . login ( Log4jAppen derBizDemo . java : 38 )
message : User login successful ly . requestID = id4
userID = user8
thread : main
time : 2018 - 01 - 26T15 : 31 + 0000
```

- Logs that record your purchase behavior:

```
level : INFO
location : com . aliyun . log4jappen dertest . Log4jAppen
derBizDemo . order ( Log4jAppen derBizDemo . java : 46 )
message : Place an order successful ly . requestID = id44
userID = user8 itemID = item3 amount = 9
thread : main
time : 2018 - 01 - 26T15 : 31 + 0000
```

Step 3 Enable query and analysis

You must enable the query and analysis function before querying and analyzing data. Follow these steps to enable the function:

1. Log on to the Log Service console.
2. Click the project name on the Project List page.
3. Click Search at the right of the Logstore.
4. Click Enable in the upper-right > Modify..

5. If you have enabled the index before, click **Index Attributes > Modify**. The **Search & Analysis** page appears.

Figure 4-4: Specify a Query field

* Field Search

custom Nginx template MNS template

Key	Enable Search				Token	Enable Analytics	Delete
	Type	alias	Case Sensitive				
level	text		☐		☒	×	
location	text		☐		☒	×	
message	text		☐	, " ' = 0 0 0 ? @ & < > / : \ n t r	☒	×	
thread	text		☐		☒	×	

+

Step 4 Analyze logs

1. Count the top three locations where errors occurred most commonly in the last hour.

```
level : ERROR | select location , count (*) as count
GROUP BY location ORDER BY count DESC LIMIT 3
```

2. Count the number of generated logs for each log level in the last 15 minutes.

```
| select level , count (*) as count GROUP BY level
ORDER BY count DESC
```

3. Query the log context.

For any log, you can precisely reconstruct the log context information for the original log file. For more information, see [Context query](#).

4. Count the top three users who have logged on most frequently in the last hour.

```
Login | select maid ( message , ' userid = (? < userID >[ a -
zA - Z \ d ]+)', 1 ) AS userID , count (*) as count GROUP
BY userID ORDER BY count DESC LIMIT 3
```

5. Compile payment total statistics for the past 15 minutes for each user.

```
order | SELECT regexp_ext ract ( message , ' userID =(? <
userID >[ a - zA - Z \ d ]+)', 1 ) AS userID , sum ( cast (
regexp_ext ract ( message , ' amount =(? < amount >[ a - zA - Z \
d ]+)', 1 ) AS double )) AS amount GROUP BY userID
```

5 Analysis - Nginx access logs

Many webmasters use Nginx as the server to build websites. When analyzing the website traffic data, they must perform a statistical analysis on Nginx access logs to obtain data such as the page views and the access time periods of the website. In the traditional methods such as CNZZ, a js is inserted in the frontend page and will be triggered when a user accesses the website. However, this method can only record access requests. Stream computing and offline statistics & analysis can also be used to analyze Nginx access logs, which however requires to build an environment, and is subject to imbalance between timeliness and analytical flexibility.

Log Service supports querying and analyzing real-time logs, and saves the analytical results to Dashboard, which greatly decreases the analytical complexity of Nginx access logs and streamlines the statistics of website access data. This document introduces the detailed procedure of log analysis function by analyzing the Nginx access logs.

Scenarios

A webmaster builds a personal website by using Nginx as the server. The PV, UV, popular pages, hot methods, bad requests, client types, and referer tabulation of the website are obtained by analyzing Nginx access logs to assess the website access status.

Log format

We recommend that you use the following `log_format` configuration for better meeting the analytic scenarios:

```
log_format main '$ remote_addr r - $ remote_user r [$  
time_local ] "$ request " $ http_host '  
'$ status $ request_length $ body_bytes  
_sent "$ http_referer " '  
'"$ http_user_agent " $ request_time $  
upstream_response_time ';
```

The meaning of each field is as follows.

Field	Meaning
remote_addr	Client address
remote_user	The client username.

Field	Meaning
time_local	The server time.
request	The request content, including method name, address, and HTTP protocol.
http_host	The HTTP address used by the user request.
Status	The returned HTTP status code.
request_length	The request size.
body_bytes_sent	The returned size.
http_referer	The referer.
http_user_agent	The client name.
Request_time	The overall request latency.
upstream_response_time	The processing latency of upstream services.

Procedure

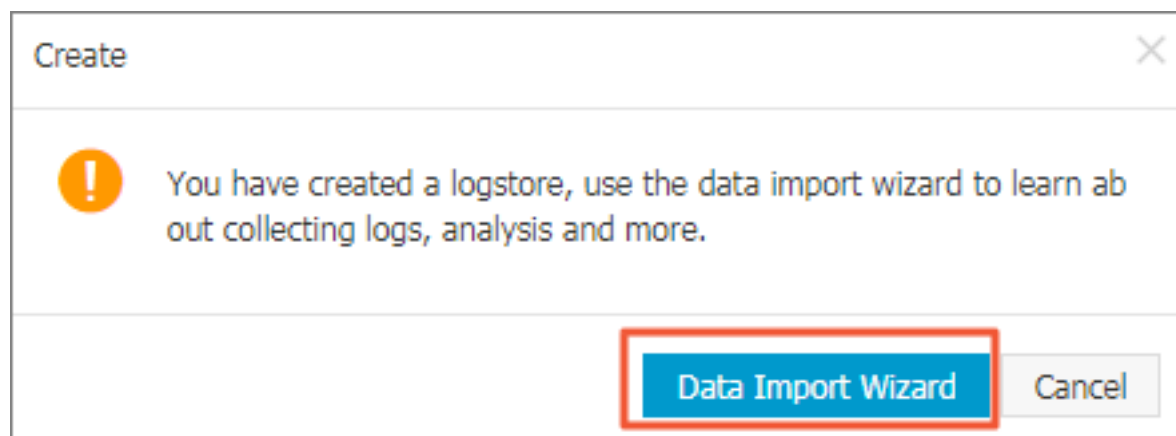
1. Open the data import wizard

Log Service provides the data import wizard to access data sources fast. To collect Nginx access logs to Log Service, use the following two methods to enter the data import wizard.

- Create a Project

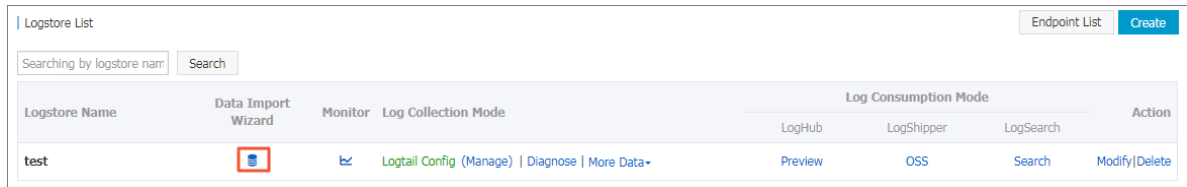
Click Wizard after creating a Logstore in an existing project or a newly created project.

Figure 5-1: Data Access wizard



- For an existing Logstore, click the Data Import Wizard icon 1 on the Logstore List page.

Figure 5-2: Logstore List



Logstore List						Endpoint List	Create
Searching by logstore name		Search					
Logstore Name	Data Import Wizard	Monitor	Log Collection Mode	Log Consumption Mode			Action
				LogHub	LogShipper	LogSearch	
test			Logtail Config (Manage) Diagnose More Data	Preview	OSS	Search	Modify Delete

2. Select a data source

Log Service provides many types of data sources, such as cloud service, third-party software, API, and SDK. To analyze the Nginx access logs, select NGINX ACCESSLOG > Third-Party Software.

3. Configure the data source

1. Enter the Configuration Name and Log Path according to your actual situation. Then, enter the recommended `log_format` information in the NGINX Log Format field.

Figure 5-3: Configure a data source

The screenshot shows a configuration form with the following fields and options:

- Configuration Name:** `test_nginx_log`
- Log Path:** `/nginx` and `log.log` (separated by `/**/`)
- Docker File:** A toggle switch is turned off. Below it, text explains that if the file is in a Docker container, Logtail will monitor container creation and destruction to collect logs based on a specified label.
- Mode:** A dropdown menu is set to `NGINX mode`.
- NGINX Log Format:** A text area containing the following log format definition:

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

Below the text area, a note states: "The standard NGINX configuration file log configuration section, usually begin with `log_format`".

Log Service automatically extracts the corresponding keys.



Note:

\$ request is extracted as two keys: request_method and request_uri.

Figure 5-4: Nginx key

NGINX Key:	Key
	remote_addr
	remote_user
	time_local
	request_method
	request_uri
	http_host
	status
	request_length
	body_bytes_sent
	http_referer
	http_user_agent
	request_time
	upstream_response_time

2. Apply to the machine groups.

If you have not created a machine group, you must create one first. For how to create a machine group, see [Create a machine group with an IP address as its identifier](#).



Note:

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

4. Search, analysis, and visualization

Make sure the heartbeat statuses of the machine groups that apply the Logtail configuration are normal and you can click Preview on the right to obtain the collected data.

Figure 5-5: Preview

Preview	
Time/IP	Content
2018-03-15 127.0.0.1	body_bytes_sent: 161 hostname: ███ http_referer: www.host9.com http_user_agent: Mozilla/5.0 (Linux; U; Android 6.0.1; zh-cn; OPPO R9s Plus Build/MMB29M) AppleWebKit/537.36 (KHTML, like Gecko) Version/4.0 Chrome/53.0.2785.134 Mobile Safari/537.36 OppoBrowser/4.3.9 http_x_forwarded_for:- remote_addr: 42.84.0.1 remote_user: request _method: POST request_time: 0.139 request_uri: /url3 sourceValue: 10.10.10.5 status: 301 streamValue: 6.708 targetValue: slb1 time_local: 15/Mar/2018:16:16:43 upstream_response_time: 1.630
2018-03-15 127.0.0.1	body_bytes_sent: 184 hostname: sun.tt http_referer: www.host9.com http_user_agent: Mozilla/5.0 (iPhone 4; CPU iPhone OS 7_0 like Mac OS X) AppleWebKit/537.51.1 (KHTML, like Gecko) Version/7.0 MQQBrowser/7.5.1 Mobile/11A465 Safari/8536.25 MttCustomUA/2 QBWebViewType/1 http_x_forwarded_for:- remote_addr: 169.235.24.133 remote_user: request method: POST request_time: 0.568 request_uri: /url8 sourceValue: 10.10.10.3 status: 200 streamValue: 1.153 targetValue: slb2 time_local: 15/Mar/2018:16:16:42 upstream_response_time: 1.726
2018-03-15 127.0.0.1	body_bytes_sent: 233 hostname: mike http_referer: www.host2.com http_user_agent: Mozilla/5.0 (Linux; U; Android 7.1.1; zh-CN; ONEPLUS A5000 Build/NMF26X) AppleWebKit/537.36 (KHTML, like Gecko) Version/4.0 Chrome/40.0.2214.89 UCBrowser/11.6.4.950 Mobile Safari/537.36 http_x_forwarded_for: 101.52.192.0 remote_addr: 42.83.144.0 remote_user: request method: POST request_time: 0.886 request_uri: /url4 sourceValue: 10.10.10.3 status: 500 streamValue: 6.766 targetValue: slb1 time_local: 15/Mar/2018:16:16:44 upstream_response_time: 1.930

Log Service provides predefined keys for analysis and usage. You can select the actual keys (generated according to the previewed data) to map with the default keys.

Figure 5-6: Key value index Properties

* Key/Value Index Attributes:

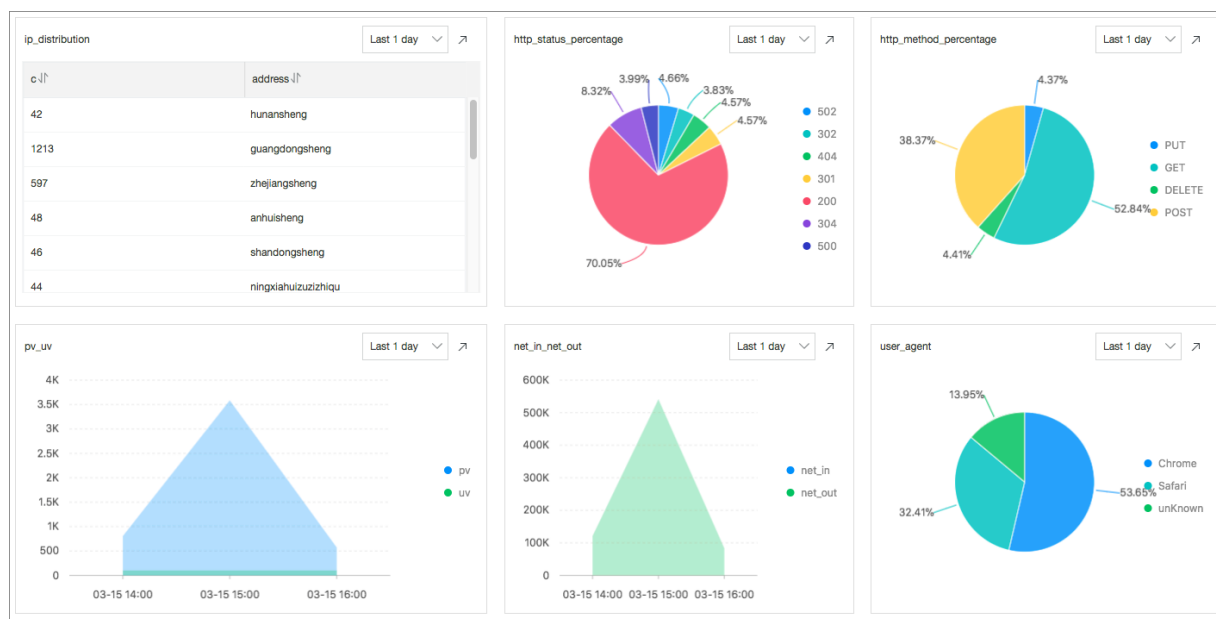
Actual Key	Type	Default Key	Case Sensitive	Token	Enable Analytics
Null	long	body_bytes_sent			☒
Null	long	bytes_sent			☒
Null	long	connection			☒
Null	long	connection_request			☒
Null	long	msec			☒
Null	long	status			☒
Null	text	time_iso8601	false	, ";=(){}?@&<>	☒
Null	text	time_local	false	, ";=(){}?@&<>	☒
Null	long	content_length			☒

Click Next. Log Service configures the index attributes for you and creates the `nginx` - dashboard dashboard for analysis and usage.

5. Analyze access logs

After the index feature is enabled, you can view the analysis of each indicator on the page where dashboards are generated by default. For how to use dashboards, see [Create and delete a dashboard](#).

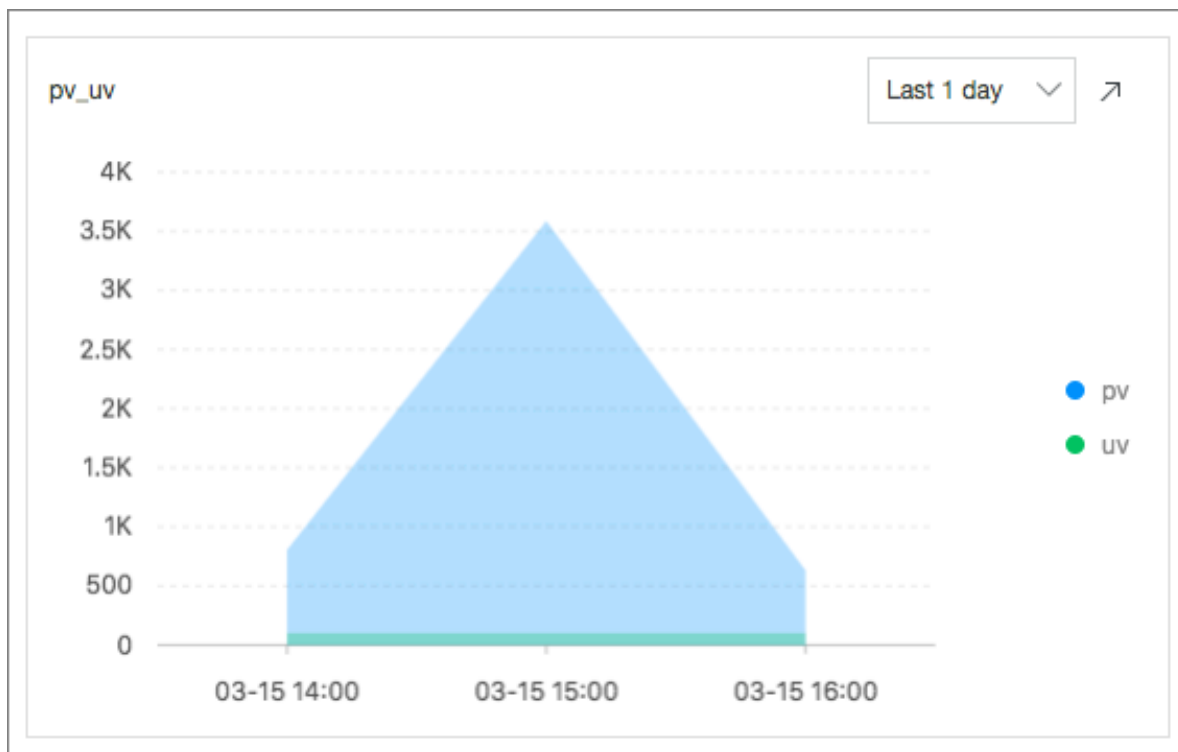
Figure 5-7: Dashboard



- PV/UV statistics (pv_uv)

Count the numbers of PVs and UVs in the last day.

Figure 5-8: PV/UV Statistics



Statistical statement:

```
* | select approx_distinct(remote_addr) as uv ,
      count(1) as pv ,
      date_format(date_trunc('hour', __time__), '%m-%d %H:%i') as time
group by date_format(date_trunc('hour',
__time__), '%m-%d %H:%i')
order by time
```

```
limit 1000
```

- Count the top 10 access pages (top_page)

Count the top 10 pages with the most PVs in the last day.

Figure 5-9: Statistical access

top_10_page		Last 1 day	
path	pv		
/url9	542		
/url1	529		
/url10	509		
/url8	502		
/url7	497		
/url3	496		
/url6	492		
/url4	488		
/url2	487		
/url5	480		

Statistical statement:

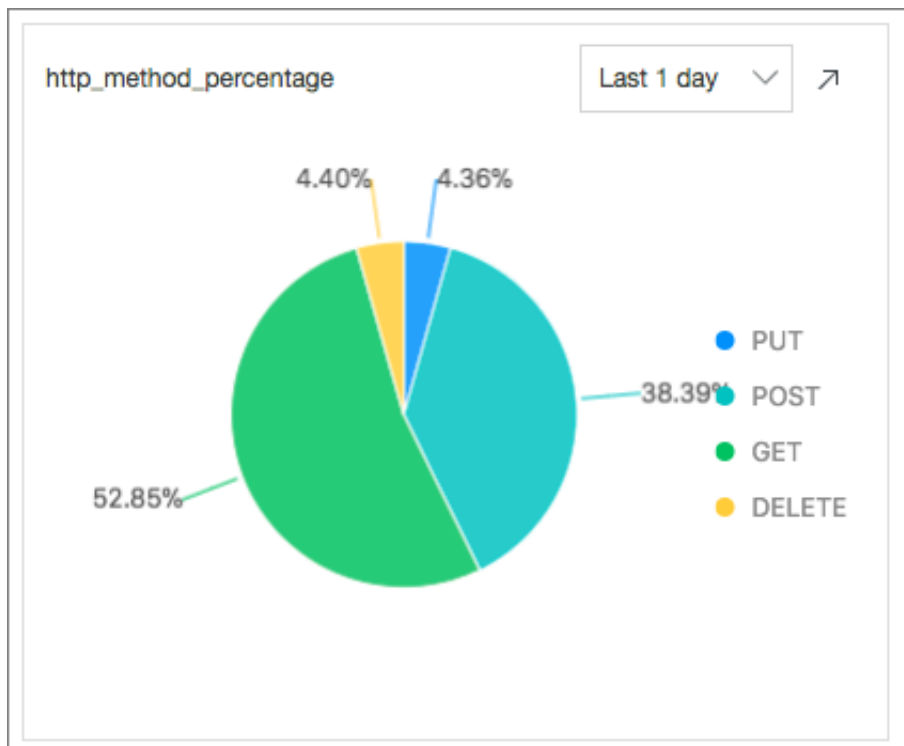
```
* | select split_part ( request_ur i , '?' , 1 ) as path ,
      count ( 1 ) as pv
      group by split_part ( request_ur i , '?' , 1 )
```

```
order by pv desc limit 10
```

- Count the ratios of request methods (http_method_percentage)

Count the ratio of each request method used in the last day.

Figure 5-10: Request Method share



Statistical statement:

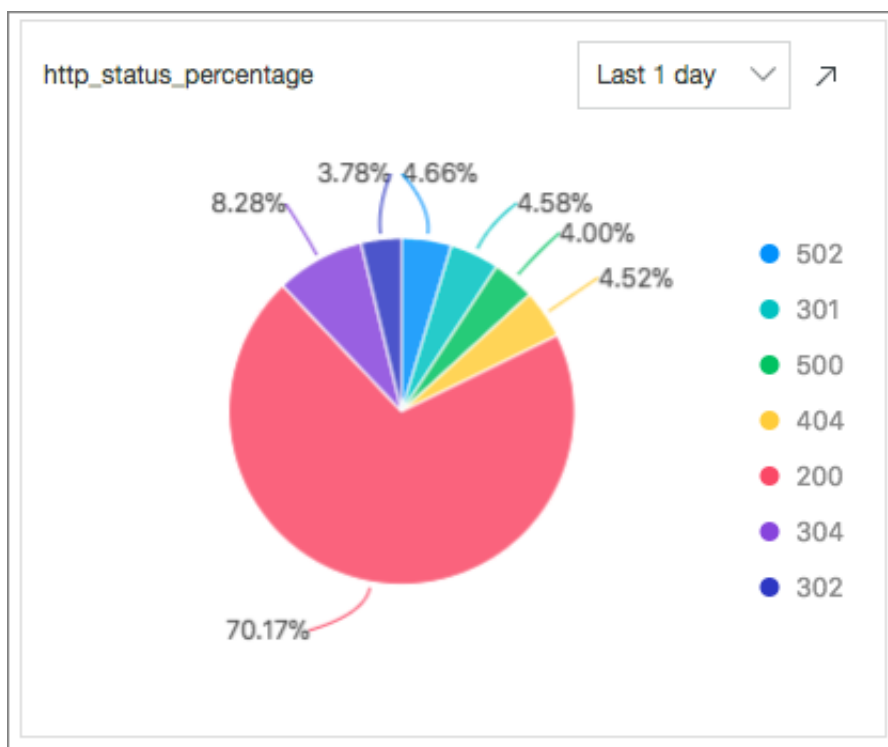
```
* | select count ( 1 ) as pv ,  
      request_method
```

```
group by request_method
```

- Count the ratios of request statuses (http_status_percentage)

Count the ratio of each request status (HTTP status code) in the last day.

Figure 5-11: Count the ratios of request statuses



Statistical statement:

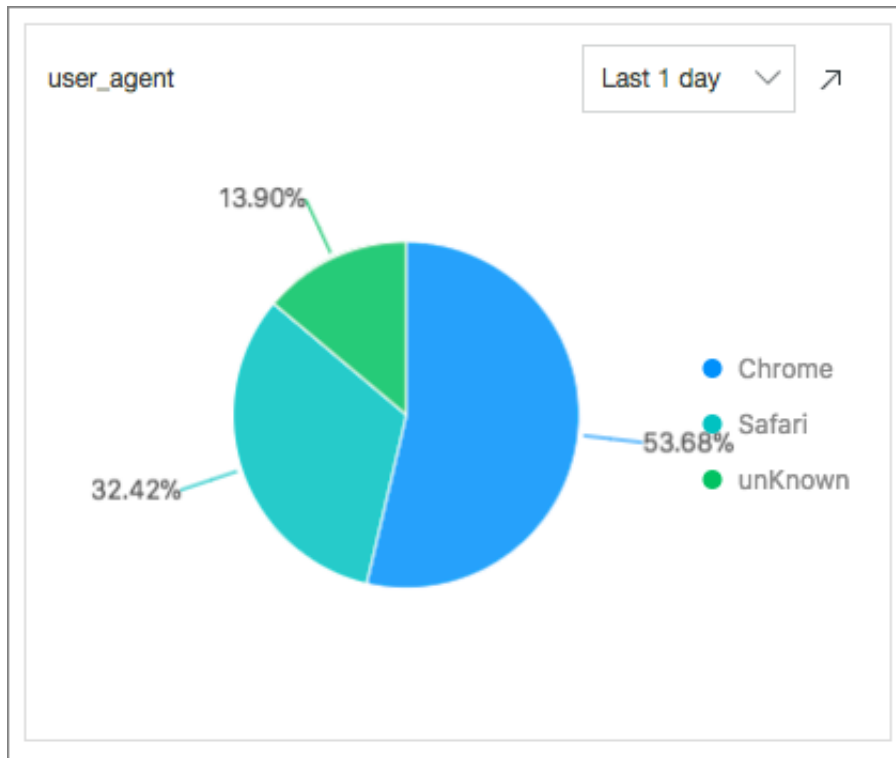
```
* | select count ( 1 ) as pv ,
      status
```

group by status

- Count the ratios of request UA (user_agent)

Count the ratio of each browser used in the last day.

Figure 5-12: Count the ratios of request UA



Statistical statement:

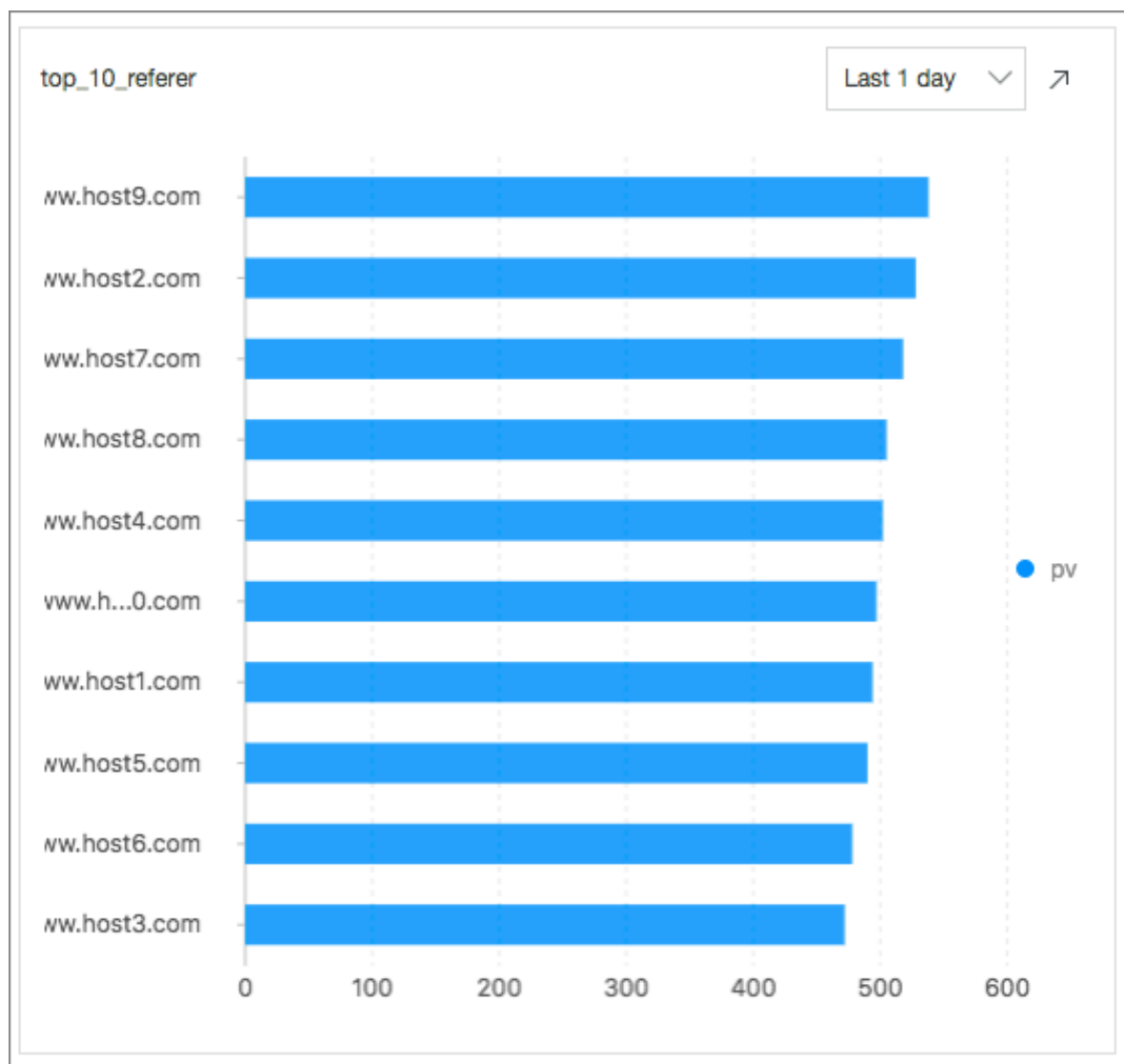
```
* | select count ( 1 ) as pv ,
    case when http_user_ agent like '% Chrome %' then '
Chrome '
    when http_user_ agent like '% Firefox %' then '
Firefox '
    when http_user_ agent like '% Safari %' then ' Safari
    ,
    else ' unKnown ' end as http_user_ agent
group by http_user_ agent
order by pv desc
```

limit 10

- Count the top 10 referers (top_10_referer)

Count the top 10 referers in the last day.

Figure 5-13: Count the top 10 referers



Statistical statement:

```
* | select count ( 1 ) as pv ,
      http_refer er
      group by http_refer er
      order by pv desc limit 10
```

6 Access diagnostics and Optimization

In addition to some default access indicators, webmasters often have to diagnose some access requests to check the latency of request processing, what are the long

latencies, and on what pages long latencies occur. Then, you can enter the query page for fast analysis.

- Count the average latency and the maximum latency

With the average latency and the maximum latency every five minutes, you can get a picture of the latency issue.

Statistical statement:

```
* | select from_unixt ime ( __time__ - __time__ % 300 ) as  
time ,  
      avg ( request_ti me ) as avg_latenc y ,  
      max ( request_ti me ) as max_latenc y  
group by __time__ - __time__ % 300
```

- Count the request page with the maximum latency

After knowing the maximum latency, you need to identify the corresponding request page to optimize page response.

Statistical statement:

```
* | select from_unixt ime ( __time__ - __time__ % 60 ) ,  
      max_by ( request_ur i , request_ti me )
```

```
group by __time__ - __time__ % 60
```

- Count the distribution of request latencies

Count the distribution of all the request latencies on the website. Place the latencies in ten buckets, and check the number of requests in each latency interval

.

Statistics statement:

```
* | select numeric_histogram ( 10 , request_time )
```

- Count the ten longest latencies

In addition to the maximum latency, the second to the tenth longest latencies and their values are also counted.

Statistics statement:

```
* | select max ( request_time , 10 )
```

- Tune the page with the maximum latency

Assume that the maximum access latency occurs on the `/ url2` page. To tune the `/ url2` page, count the PVs, UVs, numbers of various methods, statuses, and browsers, the average latency, and the maximum latency of the `/ url2` page.

Statistical statement:

```
request_uri : "/ url2 " | select count ( 1 ) as pv ,
      approx_distinct ( remote_addr ) as uv ,
      histogram ( method ) as method_pv ,
      histogram ( status ) as status_pv ,
      histogram ( user_agent ) as user_agent_pv ,
      avg ( request_time ) as avg_latency ,
      max ( request_time ) as max_latency
```

After obtaining the preceding data, you can make targeted and detailed assessments on the access status of this website.

6 Analyze IIS access logs

IIS is an extensible web server used to build and host websites. You can use access logs collected by IIS to obtain data such as page views, unique visitors, client IP addresses, bad requests, and network flow, to monitor and analyze access to your website.

Prerequisites

- You must have activated Log Service.
- You must have created a Project and a Logstore. For detailed steps about creating a Project and a Logstore, see [Preparation](#).

Context

Log format

We recommend that you use the W3C Extended Log Format so that you can specify configurations according to your requirements. In the IIS Manager, click the Select Fields toggle. Then, select sc-bytes and cs-bytes in the Standard Fields list.

Figure 6-1: Select fields

The configuration is as follows:

```
logExtFile Flags =" Date , Time , ClientIP , UserName , SiteName , ComputerName , ServerIP , Method , UriStem , UriQuery , HttpStatus , Win32Status , BytesSent , BytesRecv , TimeTaken , ServerPort , UserAgent , Cookie , Referer , ProtocolVersion , Host , HttpSubStatus "
```

- Field prefixes

Prefix	Description
s-	Server actions
c-	Client actions
cs-	Client-to-server actions
sc-	Server-to-client actions

· Field description

Field	Description
date	The date that the activity occurs
time	The time that the activity occurs
s-sitename	The Internet service name and instance number of the site visited by the client
s-computername	The name of the server on which the log entry is generated
s-ip	The IP address of the server on which the log entry is generated
cs-method	HTTP request methods such as GET and POST
cs-uri-stem	The target of the action
cs-uri-query	URI query The information following the question mark (?) in the HTTP request statement
s-port	The port number of the server that is connected with the client
cs-username	The name of the authenticated user that accessed the server. Authenticated users are referenced as <code>domain \ user name</code> . Anonymous users are indicated by a hyphen (-).
c-ip	The IP address of the client that makes the request
cs-version	The protocol version such as HTTP 1.0 or HTTP 1.1
user-agent	The browser that the client uses
Cookie	The content of the cookie sent or received. A hyphen (-) is used when there is no cookie.
referrer	The site that the user last visited. This site provides a link to the current site.
cs-host	The host header name
sc-status	The HTTP or FTP status code
sc-substatus	The status code of HTTP sub-protocol
sc-win32-status	The Windows status code.
sc-bytes	The number of bytes the server sends
cs-bytes	The number of bytes the server receives

Field	Description
time-taken	The length of time that the action took, which is indicated in milliseconds

Procedure

1. Start the Data Import Wizard steps.
 - a) On the homepage of the Log Service console, click the specified Project Name to enter the Logstore List page.
 - b) Click the icon in the Data Import Wizard column of the specified project.
2. In Step 1 Select Data Source, choose IIS ASSESSLOG under the Third-Party Software category.
3. Configure the data source.
 - a) Enter the Configuration Name and Log Path.
You can view the log path in the IIS Manager.

Figure 6-2: View the log path.

4. Select Log format.

Select the log format of your IIS access log.

- IIS : Microsoft IIS log file format
- NCSA : NCSA Common log file format
- W3C : W3C Extended log file format

5. Fill in the field IIS Logformat configuration.

- Microsoft IIS and NCSA Public formats have fixed configurations.
- To configure the IIS access log to W3C format, follow these steps:

a) Open the IIS configuration file.

- The default path for IIS5 configuration file: `C :\ WINNT \ system32 \ inetsrv \ MetaBase . bin`
- The default path for IIS6 configuration file: `C :\ WINDOWS \ system32 \ inetsrv \ MetaBase . xml`
- The default path for IIS7 configuration file: `C :\ Windows \ System32 \ inetsrv \ config \ applicationHost . config`

Figure 6-3: View the configuration file.

b) As shown in figure 3, copy the text inside the quotation marks in the field

`logFile logExtFile Flags .`

c) In the field IIS Logformat configuration in the console, paste the specified text inside the quotation marks.

Figure 6-4: Configure the Data Source

6. Confirm the key names.

IIS log service will automatically extract the key names.

Figure 6-5: IIS key names

7. Advanced Options (Optional)

Configuration item	Description
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.

Configuration item	Desceiption
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.
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 persistently 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.

Configuration item	Desceiption
Filter Configurati 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 whose level is WARNING or ERROR. · 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.

Confirm configurations and click Next.

8. Apply the configuration to the machine group.

Select the machine groups to which the specified configurations will apply. Click Apply to Machine Group at the bottom right corner of the page.

If you have not created any machine group, click Create Machine Group to create one.

9. Configure Search, Analysis, and Visualization (Optional).

When the heartbeat status of the machine group is normal, you can click **Preview** to view log data.

Figure 6-6: Preview logs

Confirm your Index Properties in the current page to view and analyze the collected log data. Click **Open** to view the Key/Value Index Attributes.

You can configure key name mapping. The key names are generated based on previewed data, and correspond to the default key names.

Figure 6-7: Key/Value Index Attributes

The system provides the default dashboard *LogstoreName-iis-dashboard* for you. After the preceding configurations are completed, you can view real-time data (including client IP distribution and the proportion of each HTTP status) on the dashboard.

Figure 6-8: Dashboard

- Use the following statement to obtain client IP distribution:

```
| select ip_to_geo (" c - ip ") as country , count ( 1 )
as c group by ip_to_geo (" c - ip ") limit 100
```

Figure 6-9: Client IP distribution

- Use the following statement to check recent page views and unique visitors:

```
*| select approx_distinct (" c - ip ") as uv , count ( 1 )
as pv , date_format ( date_trunc ( ' hour ', __time__ ),
'% m - % d % H : % i ') as time group by date_format (
```

```
date_trunc (' hour ', __time__ ), '% m -% d % H :% i ' ) order
by time limit 1000
```

Figure 6-10: Page views and unique visitors

- Use the following statement to obtain the proportion of each HTTP status:

```
*| select count ( 1 ) as pv , " sc - status " group by "
sc - status "
```

Figure 6-11: Proportion of each HTTP status

- Use the following statement to view the network flow:

```
*| select sum ( " sc - bytes " ) as net_out , sum ( " cs
- bytes " ) as net_in , date_forma t ( date_trunc ( ' hour
', time ), '% m -% d % H :% i ' ) as time group by
date_forma t ( date_trunc ( ' hour ', time ), '% m -% d % H :%
i ' ) order by time limit 10000
```

Figure 6-12: Traffic inflows and outflows

- Use the following statement to get the proportion of each Hrequest methodTTP request method:

```
*| select count ( 1 ) as pv , " cs - method " group by "
cs - method "
```

Figure 6-13: Proportion of each HTTP request method

- Use the following statement to obtain the proportion of each browser type:

```
*| select count ( 1 ) as pv , case when " user - agent
" like '% Chrome %' then ' Chrome ' when " user - agent "
like '% Firefox %' then ' Firefox ' when " user - agent "
like '% Safari %' then ' Safari ' else ' unknown ' end
as " user - agent " group by case when " user - agent
" like '% Chrome %' then ' Chrome ' when " user - agent "
like '% Firefox %' then ' Firefox ' when " user - agent "
```

```
like '% Safari %' then ' Safari ' else ' unKnown ' end
order by pv desc limit 10
```

Figure 6-14: Proportion of each browser type

- Use the following statement to view the top 10 most visited site addresses:

```
*| select count ( 1 ) as pv , split_part ( " cs - uri - stem
", '?' , 1 ) as path group by split_part ( " cs - uri -
stem " , '?' , 1 ) order by pv desc limit 10
```

Figure 6-15: Top 10 most visited site addresses

7 Analysis - Apache access logs

Log Service supports one-stop configuration of collecting Apache logs and setting indexes through the data import wizard. You can analyze website accesses in real time through the default dashboard and query analysis statements.

Prerequisites

- You have activated Log Service.
- You have created a project and a Logstore.

Context

A webmasters uses Apache as the server to build a website. To assess accesses to the website, the webmaster has to analyze Apache access logs to obtain PV, UV, IP address region distribution, client types, source pages, and other information.

Log Service supports one-stop configuration of collecting Apache logs and setting indexes through the data import wizard, and creates the access analysis dashboard for Apache logs by default.

We recommend that you use the following custom configuration for Apache logs to fit the analysis scenarios:

```
LogFormat "%h %l %u %t \"%r\" %>s %b \"%{Referer}i  
\" \"%{User-Agent}i\" %D %f %k %p %q %R %T %I %  
O" customized
```



Note:

Please check if the log content corresponding to some fields has spaces according to you log content. For example, %t, %{User-Agent}i, %{Referer}i, and more. If fields with spaces exist, use\" to wrap the fields in the configuration information to avoid affecting log resolution.

The meaning of each field is as follows:

Field	Field name	Description
%h	remote_addr	The client IP address.
%l	remote_ident	The client log name, from identd.
%u	remote_user	The client username.

Field	Field name	Description
%t	time_local	The server time.
%r	request	The request content, including the method name, address, and HTTP protocol.
%>s	status	The returned HTTP status code.
%b	response_size_bytes	The returned size.
%{Rererer}i	http\u0008_referer	The referer.
%{User-Agent}i	http_user_agent	The client information.
%D	request_time_msec	The request time in milliseconds.
%f	filename	The request file name with the path.
%k	keep_alive	The number of keep-alive requests.
%p	remote_port	The server port number.
%q	request_query	A query string. If no query string exists, this field is an empty string.
%R	response_handler	The handler for the server response.
%T	request_time_sec	The request time in seconds.
%I	bytes_received	The number of bytes received by the server, requiring the mod_logio module to be enabled.
%O	bytes_sent	The number of bytes sent by the server, requiring the mod_logio module to be enabled.

Procedure

1. Log on to the [Log Service console](#) and click the project name.
2. On the Logstores page, click the data import wizard icon for a Logstore.

3. Select data source as APACHE Access Log.

4. Configure data source.

- a) Enter Configuration Name.
- b) Enter Log Path.
- c) Select Log format.

Select Log format according to the format stated in your Apache log configuration file. To facilitate query analysis of log data, we recommend that you use a customized Apache log format.

d) Enter Apache Logformat Configuration.

If you select Log format as common or combined, the corresponding configuration is automatically entered here. If you select the Log format as Customized, enter your customized configuration here. We recommend that you enter the configuration recommended at the beginning of this document.

[illegible]

e) **Confirm APACHE Key Name.**

Log Service automatically parses your Apache key name, which you can confirm on the page.



Note:

The %r field is extracted to three keys: request_method , request_uri , and request_protocol .

NAME	My Name
mytable_id1	
mytable_id2	
mytable_id3	
mytable_id4	
mytable_id5	
mytable_id6	
mytable_id7	
mytable_id8	
mytable_id9	
mytable_id10	
mytable_id11	
mytable_id12	
mytable_id13	
mytable_id14	
mytable_id15	
mytable_id16	
mytable_id17	
mytable_id18	
mytable_id19	
mytable_id20	
mytable_id21	
mytable_id22	
mytable_id23	
mytable_id24	
mytable_id25	
mytable_id26	
mytable_id27	
mytable_id28	
mytable_id29	
mytable_id30	
mytable_id31	
mytable_id32	
mytable_id33	
mytable_id34	
mytable_id35	
mytable_id36	
mytable_id37	
mytable_id38	
mytable_id39	
mytable_id40	
mytable_id41	
mytable_id42	
mytable_id43	
mytable_id44	
mytable_id45	
mytable_id46	
mytable_id47	
mytable_id48	
mytable_id49	
mytable_id50	
mytable_id51	
mytable_id52	
mytable_id53	
mytable_id54	
mytable_id55	
mytable_id56	
mytable_id57	
mytable_id58	
mytable_id59	
mytable_id60	
mytable_id61	
mytable_id62	
mytable_id63	
mytable_id64	
mytable_id65	
mytable_id66	
mytable_id67	
mytable_id68	
mytable_id69	
mytable_id70	
mytable_id71	
mytable_id72	
mytable_id73	
mytable_id74	
mytable_id75	
mytable_id76	
mytable_id77	
mytable_id78	
mytable_id79	
mytable_id80	
mytable_id81	
mytable_id82	
mytable_id83	
mytable_id84	
mytable_id85	
mytable_id86	
mytable_id87	
mytable_id88	
mytable_id89	
mytable_id90	
mytable_id91	
mytable_id92	
mytable_id93	
mytable_id94	
mytable_id95	
mytable_id96	
mytable_id97	
mytable_id98	
mytable_id99	
mytable_id100	

f) Optional: Configure advanced options and click Next.

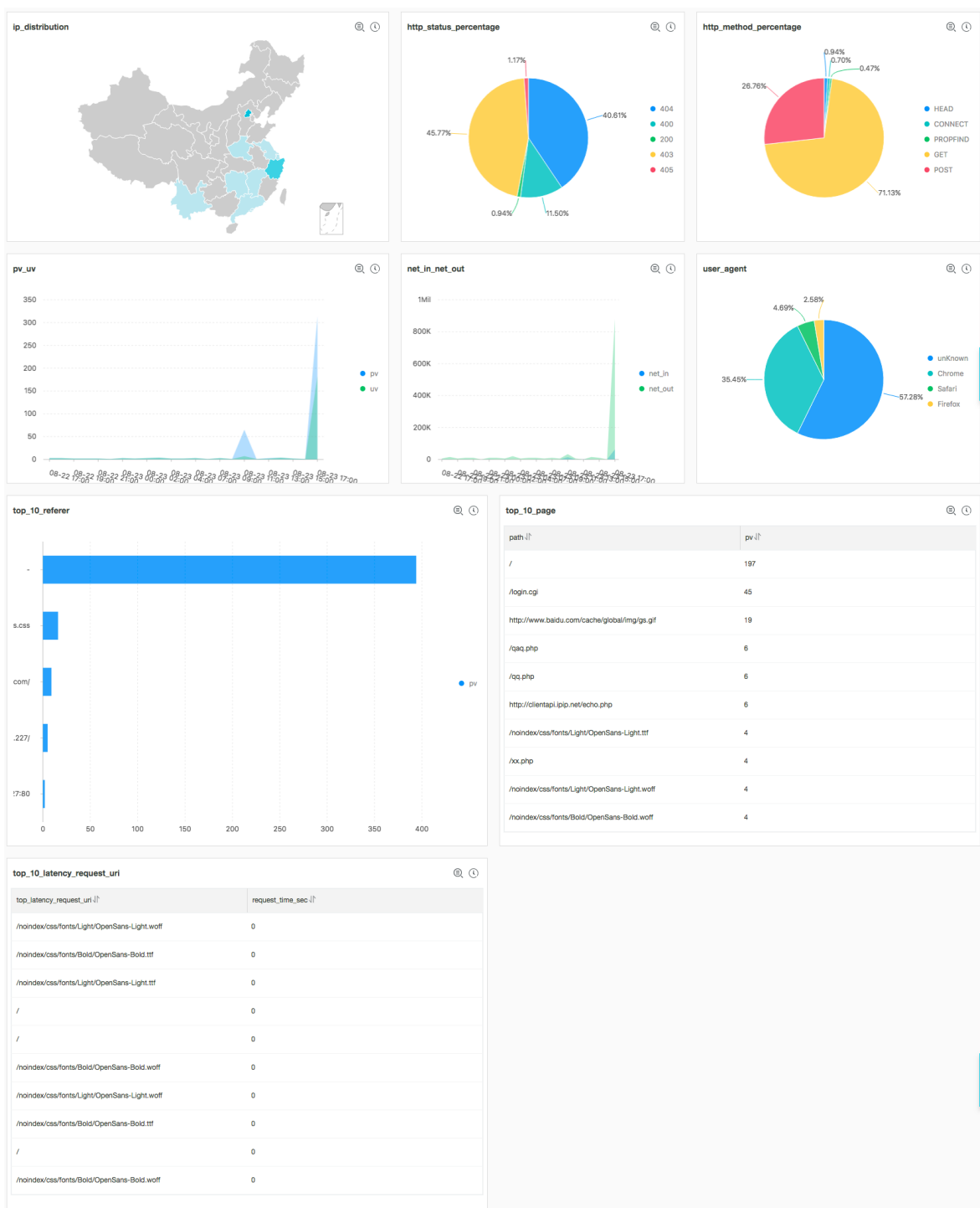
Config Maps	Details
Local Cache	Select whether to enable Local Cache. If this function is enabled, logs can be cached in the local directory of the machine when Log Service is unavailable and continue to be sent to Log Service after the service recovery. By default, at most 1 GB logs can be cached.
Upload 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.

Config Maps	Details
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.
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 persistently 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 Configuration	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 whose level is WARNING or ERROR. · 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 : .*^(?!.*(healthcheck)). *, indicates to filter logs with healthcheck in the url. Such as logs in which key is url and value is / inner / healthcheck / jiankong . html will not be collected. For similar examples, see regex-exclude-word and regex-exclude-pattern.

5. Apply the configuration to the machine group.

Select machine groups to which the specified configurations will apply. Click Apply to Machine Group at the bottom right corner of the page.

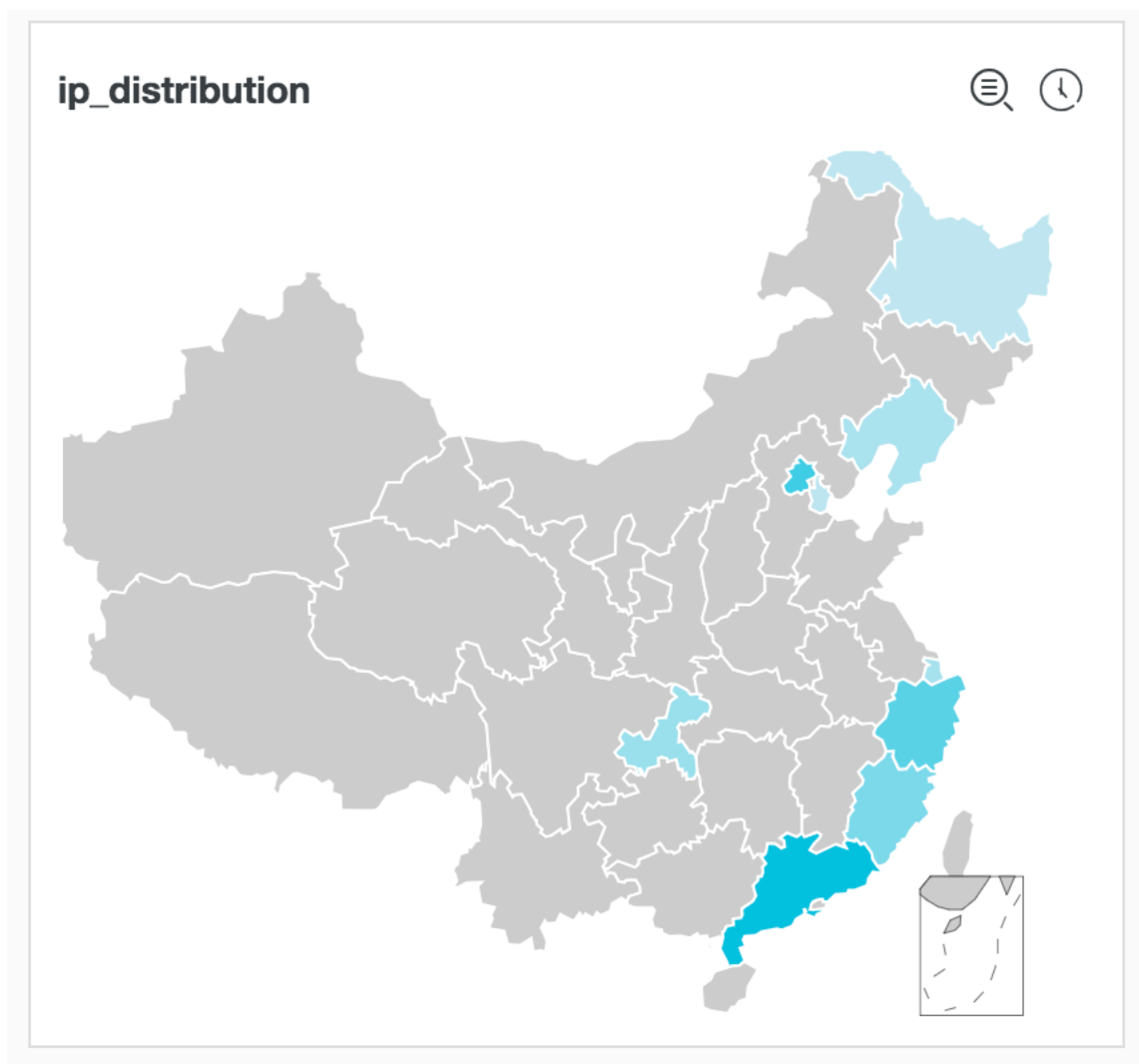
If you have not created any machine group, click Create Machine Group to create one.



- **Display source IP region distribution (ip_distribution):** Displays region distribution of the source IP addresses. The statistical statement is as follows:

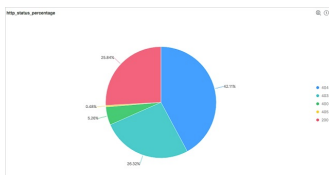
```
* | select ip_to_province ( remote_address ) as address ,
count ( 1 ) as c
```

```
group by address limit 100
```



- Count the ratios of request statuses (`http_status_percentage`) : Counts the ratio of each HTTP status code on the last day. The statistical statement is as follows:

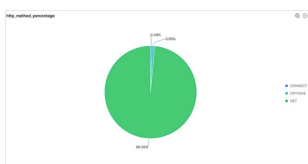
```
status > 0 | select status ,
                  count ( 1 ) as pv
              group by status
```



- Count the ratios of request methods (`http_method_percentage`) : Counts the ratio of each request method used on the last day. The statistical statement is as follows:

```
* | select request_method ,
           count ( 1 ) as pv
```

```
group by request_method
```



- **PV/UV statistics (pv_uv)** : Counts the numbers of PVs and UVs on the last day.

The statistical statement is as follows:

```
* | select date_format ( date_trunc ( ' hour ', __time__ ),
    '% m-% d % H :% i ' ) as time ,
    count ( 1 ) as pv ,
    approx_distinct ( remote_address ) as uv
    group by time
    order by time
    limit 1000
```



- **Count inbound and outbound traffic (net_in_net_out)**: Counts inbound and outbound traffic. The statistical statement is as follows:

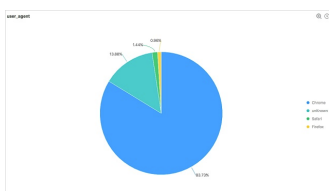
```
* | select date_format ( date_trunc ( ' hour ', __time__ ),
    '% m-% d % H :% i ' ) as time ,
    sum ( bytes_sent ) as net_out ,
    sum ( bytes_received ) as net_in
    group by time
    order by time
    limit 10000
```



- **Count the ratios of request UA (http_user_agent_percentage)**: Counts the ratio of each browser used on the last day. The statistical statement is as follows:

```
* | select case when http_user_agent like '% Chrome
% ' then ' Chrome '
    when http_user_agent like '% Firefox %' then
    ' Firefox '
    when http_user_agent like '% Safari %' then
    ' Safari '
    else ' unknown ' end as http_user_agent ,
    count ( 1 ) as pv
    group by http_user_agent
    order by pv desc
```

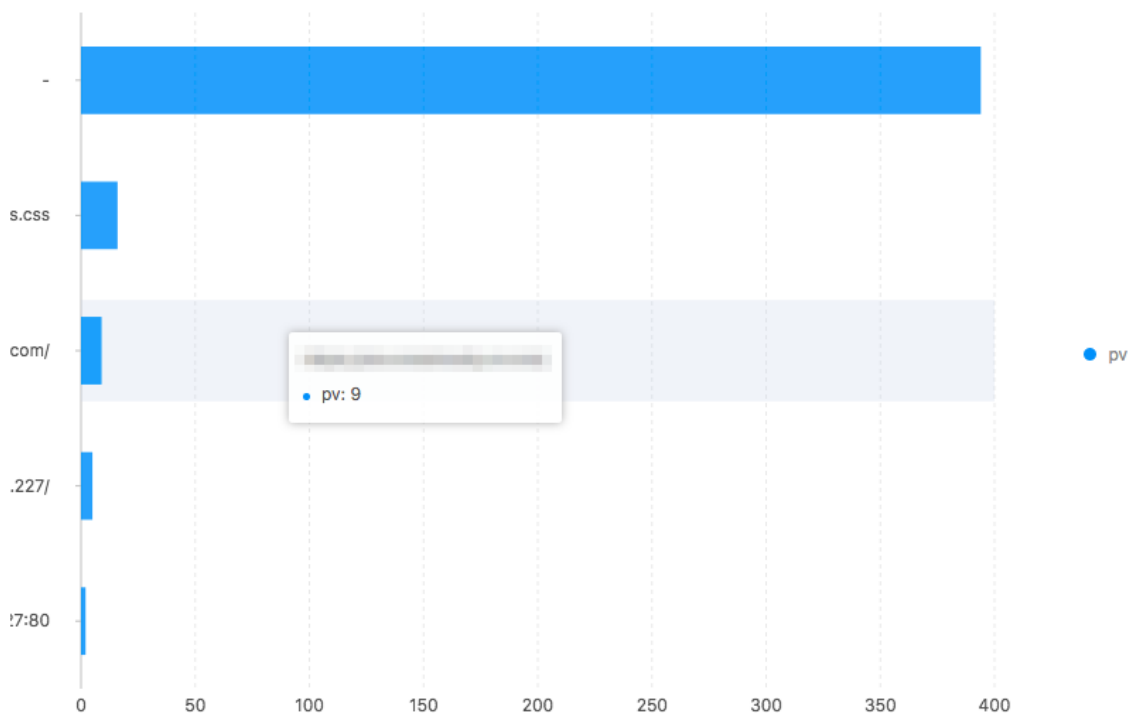
limit 10



- **Count the top 10 referers (top_10_referer):** Counts the top 10 referers on the last day. The statistical statement is as follows:

```
* | select http_referer ,
count ( 1 ) as pv
group by http_referer
order by pv desc limit 10
```

top_10_referer



- **Count the top 10 access pages (top_page):** Counts the top 10 pages with the most PVs on the last day. The statistical statement is as follows:

```
* | select split_part ( request_uri , '?' , 1 ) as path ,
count ( 1 ) as pv
group by path
```

```
order by pv desc limit 10
```

Top 10 pages

path	pv
/	55
/noindex/css/fonts/Bold/OpenSans-Bold.woff	16
/noindex/css/fonts/Light/OpenSans-Light.woff	16
/noindex/css/fonts/Light/OpenSans-Light.ttf	16
/noindex/css/fonts/Bold/OpenSans-Bold.ttf	16
/noindex/css/fonts/Bold/OpenSans-Bold.woff	15
/noindex/css/fonts/Light/OpenSans-Light.woff	15
/noindex/css/fonts/Light/OpenSans-Light.ttf	15
/noindex/css/fonts/Bold/OpenSans-Bold.woff	15
/noindex/css/fonts/Bold/OpenSans-Bold.ttf	15

- Count the top 10 uri addresses with longest response latency (top_10_latency_request_uri): Counts the top 10 uri addresses with the longest response latency for requests on the last day. The statistical statement is as follows:

```
* | select request_uri as top_latency_request_uri ,
      request_time_sec
      order by request_time_sec desc limit 10
10
```

top_10_latency_request_uri



top_latency_request_uri ↓↑	request_time_sec ↓↑
/noindex/css/fonts/Light/OpenSans-Light.woff	0
/noindex/css/fonts/Bold/OpenSans-Bold.ttf	0
/noindex/css/fonts/Light/OpenSans-Light.ttf	0
/	0
/	0
/noindex/css/fonts/Bold/OpenSans-Bold.woff	0
/noindex/css/fonts/Light/OpenSans-Light.woff	0
/noindex/css/fonts/Bold/OpenSans-Bold.ttf	0
/	0
/noindex/css/fonts/Bold/OpenSans-Bold.woff	0