Log Service

FAQ

FAQ

Hot Topics

Basic questions

Question list

1. What is Log Service?

Log Service (abbreviated to LOG) is a platform service used to collect, store, and subscribe to logs. Various types of logs can be collected in real time, managed in a centralized way, and consumed by using Log Service.

2. What functions does Log Service provide?

- Provide multiple ways of writing logs (use APIs, SDKs, and Logtail to access Log Service).
- Use Logtail to define the log collection and parsing methods.
- Use machine groups to manage log collection on thousands of machines.
- Consume and subscribe to logs in real time.
- Provide the simple configuration in the console, which allows you to perform all the operations on the Web.
- Seamlessly interconnect with multiple cloud products of Alibaba Cloud in the backend.

3. What are the basic concepts of Log Service?

- Core concepts: Project (the basic unit used to manage logs), Logstore, shard, topic (used for secondary classification of Logstores), log (number of logs), and log group.
- Concepts about log collection: Logtail configuration (used to define how to collect logs) and machine group (used to manage machines in groups).

4. What are the components of Log Service?

Log Service consists of a log collection client, a server, and other systems. Currently, the client is a log collection agent (Logtail) that is compatible with Windows and Linux. The server is responsible for reading, writing, and configuring Log Service APIs. Other systems include Alibaba Cloud products such as Object Storage Service, that is, Log Service supports synchronizing logs to cloud products such as OSS.

5. How is a log defined in Log Service?

A log contains three parts: time (required), log content (composed of key-value pairs), and metadata (the source IP address of logs).

Log management

Question list

1. How does Log Service store and manage logs?
2. Are logs lost after I delete a Logstore?
3. What is the log storage period of Log Service? Can I modify this period?
4. How to save logs in OSS, and reduce the cost of Log Service?

1. How does Log Service store and manage logs?

The Logstore is the basic unit for storing and querying logs in Log Service. Generally, a Logstore is used to store a specific type of logs. Currently, you can add, delete, modify, and query the Logstores in the console or by using APIs. You can write logs to a created Logstore by using APIs/SDKs. To collect logs from Alibaba Cloud Elastic Compute Service (ECS) instances, you can use Logtail, a log collection service provided by Log Service.

2. Are logs lost after I delete a Logstore?

Deleting a Logstore results in the loss of log data, so proceed with caution.

3. What is the log storage period of Log Service? Can I modify this period?

The following three functions of Log Service are related to the logs storage period:

- LogHub and LogSearch/Analytics: Configure the log storage period as needed.
- LogShipper: After shipping logs to Object Storage Service (OSS), configure the log storage period in OSS.

4. How to save logs in OSS, and reduce the cost of Log Service?

The index analysis of Log Service provides powerful functions. However, usage of this feature can generate certain cost. If you want to save logs in OSS, without need for customized log query and analysis, you can reduce the cost in the following way.

Precautions

Indexes are disabled by default. If you have not enabled the index analysis function, modify Logstore data storage time to reduce the cost.

Disabling the index analysis function can make log keyword query, log statistical analysis, Dashboard, and alarm features unavailable, so proceed with caution.

Cost saving method

Modify Logstore data storage time.

See Manage a Logstore to set Logstore data storage time to one day. Log Service also charges the amount of stored data. You can decrease the storage time to reduce the consumption.

Disable the index analysis function.

Activate Shipping logs to OSS function to ship the logs to OSS for saving in real

time.

Enter the **Logstore list** page, and click **Query**.

Delete the indexes to disable the index analysis function.

After you complete the preceding steps, Log Service only charges the usage of LogHub function. For more information, see Billing method.

Log Collection

Basic questions for log collection

Question list

1. What to do if logs failed to be collected?
2. What to do if the collected logs are gibberish?
3. What types of logs does Log Service collect?
4. In what ways does Log Service collect logs and which way is applicable to my case?
5. How does Log Service collect logs from ECS instances?
6. Does Log Service collect historical logs?
7. How is the log collection capability of Log Service and does it have any limits?
8. What to pay attention to when using Logtail to collect logs on NAS?

1. What to do if logs failed to be collected?

1. Check if the matching rule is passed (generally because the log sample and the logs actually collected are inconsistent).
2. Check if the log file is updated in real time. Log Service does not collect old logs.
3. The log time must contain date information such as the year.
4. A latency exists (about 1–2 minutes are needed for Log Service to read data). Wait patiently.
5. Check if the Logtail heartbeat status of a machine in a machine group is **FAIL**.
6. Log Service does not support collecting data that is not UTF-8 encoded.
7. Check the log time. Generally, logs are discarded for timeout because of the time zone

problem.

If the issue persists, contact our after-sales technical support.

2. What to do if the collected logs are gibberish?

Log Service supports collecting UTF-8 encoded logs. Gibberish may appear if the collected logs are not UTF-8 encoded.

To use SDKs to collect logs, transcode the character sets when writing codes. To use Logtail to collect logs, check the encoding of the log file monitored by Logtail.

If the issue persists, contact our after-sales technical support.

3. What types of logs does Log Service collect?

Log Service supports collecting text logs and syslog data with timestamp, and the log time must be within the last seven days and no more than 15 minutes later than the current time.

4. In what ways does Log Service collect logs and which way is applicable to my case?

Log Service allows you to write logs directly by using APIs and provides you with Logtail, which can be installed on Windows machines and Linux machines, to collect logs that are updated in real time from disk files.

1. If logs generated by applications are not flushed into the disk, you can directly use APIs to write these logs to Log Service.
2. If logs are written to the disk in real time, you can use Logtail to collect these logs.

5. How does Log Service collect logs from ECS instances?

You can use Logtail to collect logs that are flushed into the disk from Elastic Compute Service (ECS) instances as follows:

1. Create a Logstore in the Log Service console.
2. Complete the Logtail configuration.
3. Create a machine group.
4. Install the Logtail client by using the installation script.
5. Apply the Logtail configuration to the created machine group.

6. Does Log Service collect historical logs?

You can only write logs generated within the last seven days by using APIs. However, you cannot use Logtail to collect historical logs.

7. How is the log collection capability of Log Service and does it have any limits?

You can adjust the number of shards in a Logstore as needed. Logtail collects logs from an ECS instance at a maximum rate of 1 MB/s.

8. What to pay attention to when using Logtail to collect logs on NAS?

For example, to collect Nginx access logs, generally the Nginx configurations of Web servers are the same and traditionally logs are written to files with the same name on different machines (in this case, Logtail collects logs normally). After you use Network Attached Storage (NAS), Logtail will have missing logs or encounter a collection error if the Nginx logs on multiple machines are written to the same file on NAS (that is, to write logs concurrently to the same file). To avoid this problem, make sure that the logs on different Web servers are written to different files on NAS.

Basic questions for Logtail

Question list

1. What is Logtail?
2. Does Logtail collect static log files?
3. What platforms does Logtail support?
4. How to install and upgrade the Logtail client?
5. How to configure and use the Logtail client?
6. How does Logtail work?
7. Does Logtail support log rotation?
8. How does Logtail handle a network exception?
9. What is the log collection latency of Logtail?
10. How does Logtail process historical logs?
11. How long does a change in log collection configuration take effect for Log Service?
12. How to troubleshoot log collection problems of Logtail?

1. What is Logtail?

Logtail is a log collection client provided by Log Service and facilitates the log access. After being installed on your machine, Logtail monitors specified log files and automatically uploads the logs newly written to these files to your specified Logstore.

2. Does Logtail collect static log files?

Logtail monitors file changes based on change events in the file system and sends logs generated in real time to Log Service. Logtail does not collect the log file contents if the log file is not modified.

3. What platforms does Logtail support?

Currently, Logtail supports 64-bit Linux and 32-bit/64-bit Windows Server 2003 and later versions.

Linux:

- Aliyun Linux
- Ubuntu
- Debian
- CentOS
- OpenSUSE

Windows:

- Windows 7 (Client) 32 bit
- Windows 7 (Client) 64 bit
- Windows Server 2003 32 bit
- Windows Server 2003 64 bit
- Windows Server 2008 32 bit
- Windows Server 2008 64 bit
- Windows Server 2012 64 bit

4. How to install and upgrade the Logtail client?

Installation: Install the Logtail client by using the installation script.

Upgrade: Log Service regularly upgrades the Logtail client without interrupting the data collection process.

5. How to configure and use the Logtail client?

For more information, see Collect logs by configuring Logtail in the console.

6. How does Logtail work?

1. Configure the directory to be monitored, the log file name, and the corresponding parsing rule (regular expression) in the console.
2. If the log file is modified on your machine, Logtail receives an event from the file system and reads the new log.
3. Logtail parses the log format based on the regular expression and sends the log to Log Service.

7. Does Logtail support log rotation?

If the log file a.LOG reaches a given size or lasts for a given period of time after being created, a.LOG is renamed as a.LOG.1 (or another name). A new a.LOG file is created for writing new logs. This process is called rotation. Logtail automatically rotates logs based on event notifications from the file system.

8. How does Logtail handle a network exception?

If a network exception occurs or the write quota is exceeded, Logtail caches collected logs to the local disk and resends those logs later. The maximum disk cache capacity is 500 MB. Newly cached data overwrites the old one if the 500 MB limit is exceeded. Cached files that fail to be sent to Log Service within 24 hours are automatically deleted.

9. What is the log collection latency of Logtail?

Logtail collects logs based on events and generally sends collected logs to Log Service within three seconds.

10. How does Logtail process historical logs?

Logtail only collects real-time logs. If the difference between the log time and the system time at which Logtail processes the log is more than five minutes, the log is regarded as a historical log.

11. How long does a change in log collection configuration take effect for Log Service?

After you apply the modified configuration to the machine group in the console, Logtail loads the configuration and has it taken effect within three minutes.

12. How to troubleshoot log collection problems of Logtail?

1. Check if the Logtail heartbeat is normal. If not, reinstall Logtail.
2. Check if the log file in the log collection configuration is generated in real time.
3. Check if the regular expression in the log collection configuration matches the log content. If not, view the error in the Logtail running log (Linux:/usr/local/ilogtail/ilogtail.LOG).

For more information, see Troubleshoot log collection errors.

Logtail heartbeat error

If the Logtail heartbeat status is abnormal when you use Logtail to collect logs, you can troubleshoot the problem by using the automatic diagnosis tool of Logtail or manual diagnosis.

Automatic diagnosis

Log Service provides the Logtail automatic diagnosis tool to troubleshoot the heartbeat problem. For more information, see Logtail quick diagnosis tool.

> **Note:**
> If the diagnosis result is normal, see the echo message of the diagnosis tool or the result of the manual diagnosis to check if any exception occurs.

Manual diagnosis

The **FAIL** heartbeat status of Logtail is generally caused by the following reasons. Inspect the reasons one by one.

1. Network is disconnected

Run the following command to check the network connectivity. Make sure the network is normal.

Classic network

telnet logtail.cn-<region>-intranet.log.aliyuncs.com 80

Virtual Private Cloud (VPC)

telnet logtail.cn-<region>-vpc.log.aliyuncs.com 80

Internet

telnet logtail.cn-<region>.log.aliyuncs.com 80

2. Logtail is not installed

Run the following command to check the client status. If the Logtail client is not installed, install the Logtail in the same region and network type as your Log Service project. For more information, see Install Logtail on Linux or Install Logtail on Windows.

Linux

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

Windows

```
Control Panel -> Management Tool -> Service
Check the running status of the following two Windows services: LogtailDaemon and LogtailWorker
```

3. Parameter configured in the installation process is incorrect

Log Service has different regions. You must specify the correct service endpoint for the Logtail client during the installation. Check the configuration used by the installed Logtail client.

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

Make sure that:

- The endpoint that the Logtail client connects is in the same region as your project. For more information about the endpoint list, see Service endpoint.
- You have selected the correct domain name according to the network environment of your machine. For example, an internal domain name selected in the VPC environment cannot be connected. Telnet to the domain name configured in ilogtail_config.json, for example, telnet logtail.cn-hangzhou-intranet.log.aliyuncs.com 80.

4. IP address or user ID configured in Log Service is incorrect

Generally, Logtail obtains the IP address on the machine in the following way:

- If the IP address is bound with the hostname in the file /etc/hosts on the current machine, confirm the bound IP address. Run the **hostname** command to view the hostname.
- If the IP address is not bound with the hostname, Logtail obtains the IP address of the first

network adapter on the current machine.

To view the IP address on the server:

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

> **Note:**
> Logtail cannot work if in the file app_info.json the ip field is empty. Configure the IP address for the host and then restart Logtail.

If the IP address entered in the Log Service machine group is different from that obtained by the Logtail client, modify as follows:

- Modify the IP address that is entered incorrectly in the Log Service machine group. Wait for one minute and check the heartbeat status.
- If the network configurations on the machine are modified (such as modifying the /etc/hosts file), restart Logtail to obtain the new IP address.

Run the following command to restart Logtail, if necessary:

- Linux: sudo /etc/init.d/ilogtaild stop; sudo /etc/init.d/ilogtaild start
- Windows: **Control Panel** > **Management Tool** > **Service** > **Restart LogtailWorker**

5. The machine whose logs are being collected is not an ECS instance, or is an ECS instance that does not belong to the account of the Log Service project

You must authorize the machine where Logtail is installed to collect logs in any of the following situations. For more information, see Configure user ID for non-Alibaba Cloud ECS.

1. The machine is non-ECS machine.
2. The account used to purchase the ECS instance and the one used to create the Log Service project is not the same one.

If the problem persists, open a ticket and provide your project, Logstore, machine group, app_info.json, ilogtail_config.json, and the output of the automatic diagnosis tool in the ticket.

Logtail quick diagnosis tool

If an exception occurs during the log collection, you can use the Logtail automatic diagnosis tool to

check if an exception exists in the client and troubleshoot the problem quickly as instructed by the tool.

If the Logtail heartbeat status is **Fail**, see Logtail heartbeat error.

Preparations

Download the diagnosis tool script

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

wget http://logtail-corp.oss-cn-hangzhou-zmf.aliyuncs.com/linux64/checkingtool.sh

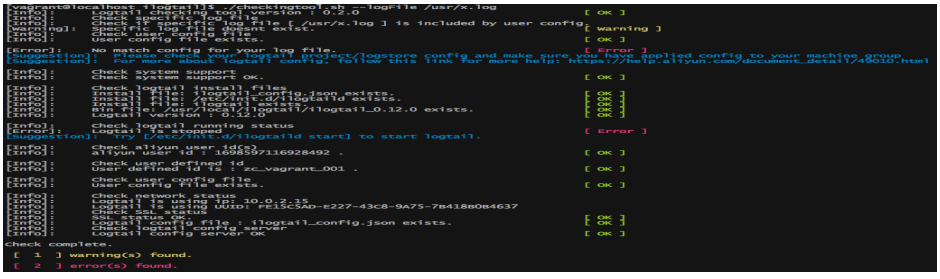
Common parameters for the diagnosis tool

- --help: View help document.
- --logFile [LogFileFullPath]: Check if Logtail collects logs from LogFileFullPath and check the basic running environment of Logtail such as installation file integrity, running status, Alibaba Cloud user ID, and network connectivity.
- --logFileOnly [LogFileFullPath]: Only check if Logtail collects logs from LogFileFullPath.
- --envOnly: Only check the running environment of Logtail.

Usage

Run the script ./checkingtool.sh --logFile [LogFileFullPath] to perform the check. If the script detects an exception, proceed as instructed by the script.

> **Note:** If the specific log file passes the check and the Logtail running environment is normal, we recommend that you log on to the Alibaba Cloud console to view the configuration exception logs of Log Service. For more information, see Query log collection errors.



Causes and solutions of common Logtail log collection errors

You can find the causes of Logtail log collection errors by running the Logtail quick diagnosis tool and then use the corresponding solution to solve the problem. The causes and solutions of common

Logtail log collection errors are as follows.

Cause	Solution
Installation file is missing	Reinstall Logtail.
Logtail is not running	Use the command /etc/init.d/ilogtaild start to start Logtail.
Multiple Logtail processes	Use the command /etc/init.d/ilogtaild stop to stop Logtail and then use the command /etc/init.d/ilogtaild start to start Logtail.
Port 443 is disabled	Open the port 443 in the firewall.
Cannot find the configuration server	Check if the installation is correct. If not, uninstall and then reinstall Logtail.
User configuration does not exist	Confirm that the Logtail configuration is created in the console, the machine group contains the client, and the configuration is applied to the machine group.
The specified log file is not matched	Check if the Logtail configuration is correct.
The specified log file is matched more than once	Logtail selects one configuration randomly when multiple matches exist. We recommend that you retain only one configuration that matches the specified log file.

Instructions

The Alibaba Cloud ID and dynamic machine group/user-defined identity configured in the client are output when the diagnosis tool is running. No alarm is triggered if they do not exist. If the client requires the configuration of Alibaba Cloud ID or dynamic machine group/user-defined identity, check if the output of the tool and your configuration are the same. If not, reconfigure them as described in Collect logs from non-Alibaba Cloud ECS instances or ECS instances not in your account and Configure a user-defined identity for a machine group.

The diagnosis tool must use curl to check the network connectivity. Make sure the machine has the curl tool installed.

Configure a regular expression

To configure Logtail to collect text logs, you must configure a regular expression based on your log sample if you select to use the **Full Mode** to parse logs.

Log Service supports automatically generating the regular expression. Paste your log sample in the **Log Sample** field to automatically generate the regular expression.

If the automatically generated regular expression do not fully cover your log sample, you can also write a regular expression manually by referring to the regular expression sample of Apache (for more information, see Apache logs).

Take the standard Nginx access logs as an example.

Configure parts of the regular expression first and use .* for the other parts.

(\d+.\d+.\d+.\d+)(.*)

Modify the regular expression based on the log sample.

(\d+.\d+.\d+.\d+)\s(\d+)(.*)

If you have any issues, contact our after-sales support.

Compare Log Service LogHub and Kafka

Kafka is a distributed messaging system with high throughput and horizontal scaling and is widely used for message publishing and subscription. It is available as an open-source software. You can build a Kafka cluster as needed.

Log Service is a log-specific platform service built based on Apsara Pangu, supports the real-time collection, storage, distribution, and real-time query of various types of logs, and provides services by using standard RESTful APIs.

The Log Service LogHub provides public channels of log collection and distribution. To not build and maintain the Kafka cluster on your own, you can use the Log Service LogHub.

Concept mapping between Log Service LogHub and Kafka

Concept	Kafka	LogHub

Storage object	Topic	Logstore
Horizontal partitioning	Partition	Shard
Data consumption location	Offset	Cursor

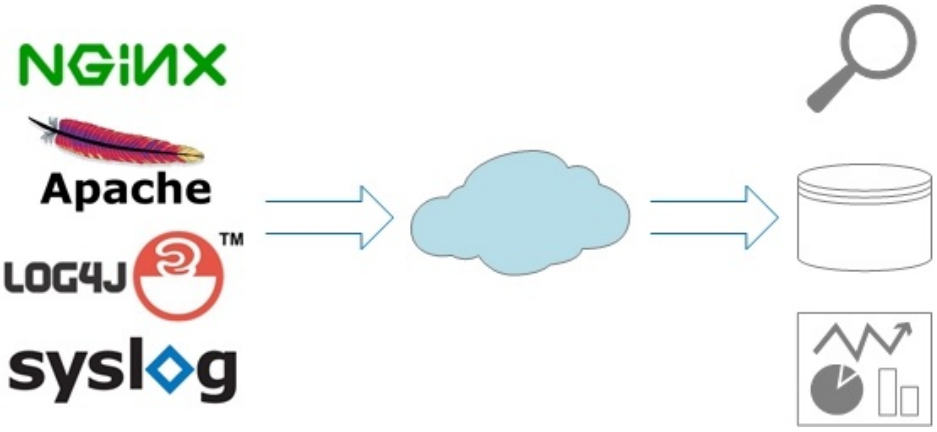
Function comparison between LogHub and Kafka

Function	Kafka	LogHub
Usage dependency	Self-built or shared Kafka cluster	Log Service
Communication protocol	Network interconnection by using TCP	HTTP (RESTful API), port 80
Access control	None	Signature authentication + access control based on an Alibaba Cloud account
Dynamic expansion	None	Auto scaling (merge/split) of shard quantities in a dynamic manner without affecting users
Multi-tenant QoS	None	Shard-based standard throttling
Number of data copies	Custom	Three copies by default and cannot be modified
Failover/replication	Completed by using tools	Completed in an automatic and perception-free manner
Expansion/upgrade	Completed by using tools with service impact	Perception-free
Write mode	Round robin/key hash	Currently, only round robin and key hash are supported
Current consumption location	Stored in the zookeeper of the Kafka cluster	Maintained in Log Service without user intervention
Storage period	Based on the configuration	Adjusted dynamically based on the requirement

Comparison among log collection tools: Logstash, fluentd, and Logtail

Assessments of log collection clients

Hundreds of millions of servers, mobile terminals, and network devices generate massive logs every day in the DT era. Centralized log processing solution effectively supports log consumption in the entire lifecycle. The first step is to collect logs from devices to the cloud.



Three log collection tools

Logstash

- Logstash is the “L” of the ELK stack, which is famous in the open source community. It plays an active role in the community and supports many plug-ins in the ecosystem.
- Logstash is implemented based on JRuby and can be run on JVM across platforms.
- Its modular design delivers high scalability and interoperability.

Fluentd

- Fluentd is a popular log collection tool in the open source community. td-agent, the commercial version of fluentd, is maintained by Treasure Data and is assessed in this document.
- Fluentd is implemented based on CRuby and re-implements the components essential for performance by using the C language. The overall performance is good.
- Fluentd features concise design and provides high reliability for the data transfer in the pipeline.
- Compared with Logstash, fluentd supports fewer plug-ins.

- Logtail

- Logtail is the producer of Alibaba Cloud Log Service and has been widely applied in massive big data scenarios of Alibaba Group for more than three years.

- Logtail is implemented by using the C++ language and delivers good performance after great efforts made to improve its stability, resource control capability, and management.
- Compared with the community support of Logstash and fluentd, Logtail is dedicated to log collection with lower functional variety.

Function comparison

Function	Logstash	Fluentd	Logtail
Log reading	Polling	Polling	Event triggered
File rotation	Supported	Supported	Supported
Failover (local checkpoint)	Supported	Supported	Supported
General log parsing	Grok parsing (based on a regular expression)	Parsing based on a regular expression	Parsing based on a regular expression
Specific log type	Supports mainstream formats such as delimiter, key-value, and JSON	Supports mainstream formats such as delimiter, key-value, and JSON	Supports mainstream formats such as delimiter, key-value, and JSON
Data compression before being sent	Supported by using plug-ins	Supported by using plug-ins	LZ4
Data filter	Supported	Supported	Supported
Buffer-based data transfer	Supported by using plug-ins	Supported by using plug-ins	Supported
Transfer exception handling	Supported by using plug-ins	Supported by using plug-ins	Supported
Running environment	JRuby implementation with JVM environment dependency	CRuby and C implementation with Ruby environment dependency	C++ implementation without special requirements
Thread support	Multiple threads	Multiple threads restricted by GIL	Multiple threads
Hot upgrade	Not supported	Not supported	Supported
Centralized configuration management	Not supported	Not supported	Supported
Self-detection of the running status	Not supported	Not supported	Supports CPU/memory threshold protection

Log file collection – Performance comparison

Log sample: Take the Nginx access logs as an example. The following log is a 365-byte Nginx access log with 14 structured fields.

The following test repeatedly writes the log to a file at different simulated pressures. The time field of each log is set to the system time when the log is written, and the other 13 fields are the same for all logs.

The log parsing process in the simulated scenario is the same as that in the actual scenario, except that the network traffic generated by the write operation is reduced due to a relatively high data compression rate.

Logstash

logstash-2.0.0 parses logs by using grok and writes parsed logs to Kafka (which has built-in plug-ins and enables Gzip compression).

Log parsing configuration:

```
grok {
patterns_dir=>"/home/admin/workspace/survey/logstash/patterns"
match=>{ "message"=>"%{IPORHOST:ip} %{USERNAME:rt} - \[%{HTTPDATE:time}\] \"%{WORD:method}
%{DATA:url}\" %{NUMBER:status} %{NUMBER:size} \"%{DATA:ref}\" \"%{DATA:agent}\" \"%{DATA:cookie_unb}\"
\"%{DATA:cookie_cookie2}\" \"%{DATA:monitor_traceid}\" %{WORD:cell} %{WORD:ups}
%{BASE10NUM:remote_port}" }
remove_field=>["message"]
}
```

Test results

Write TPS	Write traffic (KB/s)	CPU usage (%)	Memory usage (MB)
500	178.22	22.4	427
1000	356.45	46.6	431
5000	1782.23	221.1	440
10000	3564.45	483.7	450

Fluentd

td-agent-2.2.1 parses logs by using a regular expression and writes parsed logs to Kafka (which has

the third-party plug-in fluent-plugin-kafka and enables Gzip compression).

Log parsing configuration:

```
<source>
type tail
format /^(?<ip>\S+)\s(?<rt>\d+)\s-
\s\[(?<time>[^\]]*)\]\s"(?<url>[^\"]+)"\s(?<status>\d+)\s(?<size>\d+)\s"(?<ref>[^\"]+)"\s"(?<agent>[^\"]+)"\s"(?<
cookie_unb>\d+)"\s"(?<cookie_cookie2>\w+)"\s"(?
<monitor_traceid>\w+)"\s(?<cell>\w+)\s(?<ups>\w+)\s(?<remote_port>\d+).*$/
time_format %d/%b/%Y:%H:%M:%S %z
path /home/admin/workspace/temp/mock_log/access.log
pos_file /home/admin/workspace/temp/mock_log/nginx_access.pos
tag nginx.access
</source>
```

Test results

Write TPS	Write traffic (KB/s)	CPU usage (%)	Memory usage (MB)
500	178.22	13.5	61
1000	356.45	23.4	61
5000	1782.23	94.3	103

> **Note:** A single process of fluentd uses at most one CPU core due to GIL limits. You can use the multiprocess plug-in to support higher log throughput in multiple processes.

Logtail

Logtail 0.9.4 performs log structuring by using a regular expression and writes LZ4-compressed data to Alibaba Cloud Log Service in the HTTP protocol. The batch_size is set to 4000.

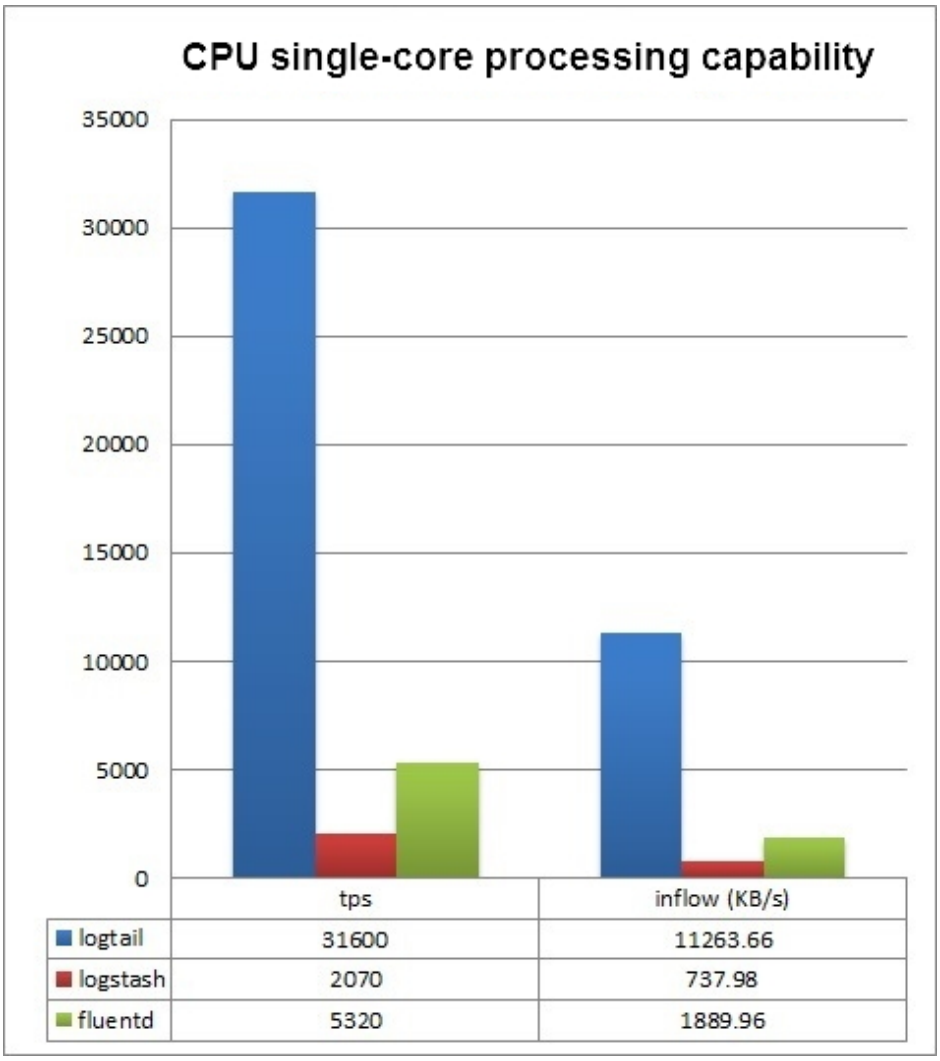
Log parsing configuration:

```
logRegex : (\S+)\s(\d+)\s-
\s\[([^]]+)]\s"([^"]+)"\s(\d+)\s(\d+)\s"([^"]+)"\s"([^"]+)"\s"(\d+)"\s"(\w+)"\s"(\w+)\s(\w+)\s(\w+)\s(\d+).*
keys : ip,rt,time,url,status,size,ref,agent,cookie_unb,cookie_cookie2,monitor_traceid,cell,ups,remote_port
timeformat : %d/%b/%Y:%H:%M:%S
```

Test results

Write TPS	Write traffic (KB/s)	CPU usage (%)	Memory usage (MB)
500	178.22	1.7	13
1000	356.45	3	15
5000	1782.23	15.3	23
10000	3564.45	31.6	25

Single-core processing capability comparison



Conclusion

Logstash, fluentd, and Logtail have their own features as follows:

- Logstash supports all the mainstream log types, diverse plug-ins, and flexible customization, but has relatively low performance and is prone to high memory usage because of JVM.
- Fluentd supports all the mainstream log types and many plug-ins, and delivers good performance.
- Logtail occupies the least CPU and memory resources of the machine, delivers good performance throughput, and provides full support for common log collection scenarios. However, it has no plug-in support and delivers lower flexibility and scalability than Logstash and fluentd.

Common errors of Log Service

Errors of Log Service

1. illegal param! [LogContent] is null
2. send data fail, error_code:WriteQuotaExceed error_message:Write quota exceed projectName:project_name
3. WARN Aliyun.SLS.Logtail.LogFileReader - logfile xxx.log has n old enough logs with first fail case xxx
4. Check if the IP address is correct. Currently, Log Service only supports ECS instances in your region
5. Different operating systems or invalid IP address. Currently, Log Service only supports ECS instances in your region
6. ShardWriteQuotaExceed

Solutions

1. illegal param! [LogContent] is null

Check the following configurations:

Check if you have entered the log sample.

Check if the regular expression of the first log line is correct.

If your issue persists, provide your log sample and regular expression for us to troubleshoot the issue.

If your issue is still not solved, contact our after-sales technical support.

2. send data fail, error_code:WriteQuotaExceed error_message:Write quota exceed projectName:project_name

If you find an error similar to the following one in ilogtail.log when using Log Service:

send data fail, error_code:WriteQuotaExceed error_message:Write quota exceed projectName:project_name

It means your write quota is insufficient, that is, the log volume you write exceeds the specified threshold or the write speed exceeds the limit.

Currently, the processing capability of each shard is:

Write: 5 MB/s, 500 times/s

Read: 10 MB/s, 100 times/s

If your data volume is beyond what a shard can process, you can split the shard. For more information, see **Split a shard** in Manage a shard.

The maximum number of write requests per minute at the project level is 300,000. If you are writing logs by using a program and some requests exceed the quota, we recommend that you write logs in batches or use Producer-lib to limit the maximum packet size and number of logs for each upload to 3 MB and 4096 respectively.

3. WARN Aliyun.SLS.Logtail.LogFileReader - logfile xxx.log has n old enough logs with first fail case xxx

You may have the following error in the ilogtail.log:

WARN Aliyun.SLS.Logtail.LogFileReader - logfile xxx.log has n old enough logs with first fail case xxx

This error is caused by old logs.

Log file collection rules for Logtail:

Process the historical data separately.

Reduce the cache time for data flushing into the disk or even perform real-time flushing.

Note the time zone when logs are modified.

Solutions:

If a new log file is monitored by Logtail, logs in the log file whose log time is not within the last minute are considered as old data and discarded.

For logs newly written to a log file, logs whose log time is not within the last five minutes are also considered as old data and discarded. This error is caused basically because your logs cached in the memory have timed out when they are written to the log file.

Logs whose log time is beyond the range of -7 days–360s are discarded by Log Service. This error is caused basically because the configured time zone of the log is incorrect.

If rule 1 is violated, the historical data is discarded. But if rule 2 and 3 are not violated, no error is reported in the subsequent logs.

If rule 2 is violated, an error is reported occasionally, and you can query some logs in the console.

If rule 3 is violated, some logs are not collected and then cannot be queried in the console.

If the issue persists, contact our after-sales technical support.

4. Check if the IP address is correct. Currently, Log Service only supports ECS instances in your region

An error occurs when you add machines to a machine group in the console, asking you to check if the IP address is correct and telling you that Log Service only supports ECS instances in your region. This error is caused because the intranet IP address of the server you entered when adding the machine group is incorrect.

Check the following configurations if this error occurs and enter the correct intranet IP address.

Make sure that your entered IP address of the ECS instance belongs to the logon cloud account.

Currently, you can only enter the IP address of the ECS instance that is in the same region as the project. For example, if the project is in China East 1 (Hangzhou), you must enter the IP addresses of the ECS instances in the region China East 1 (Hangzhou).

You must enter the intranet IP address of the ECS instance. To enter multiple IP addresses, make sure that each IP address is in a single line.

You can only enter the IP address of an ECS instance, while that of a Virtual Private Cloud (VPC) instance cannot be entered.

If the issue persists, contact our after-sales technical support.

5. Different operating systems or invalid IP address. Currently, Log Service only supports ECS instances in your region

An error occurs when you add machines to a machine group in the console, indicating that the ECS

instances are of different operating systems or the IP address is invalid. Currently, Log Service only supports ECS instances in your region. This error is caused because Windows ECS instances and Linux ECS instances cannot exist in the same machine group, that is, you must add Windows machines and Linux machines to different machine groups.

Complete the correct configurations and then add the machines to the machine group.

If the issue persists, contact our after-sales technical support.

6. ShardWriteQuotaExceed

This error occurs when the number of your shards is small and the write quota is exceeded. To increase the number of shards, see **Split a shard** in Manage a shard.

Currently, each shard has certain service capability:

Write: 5 MB/s, 500 times/s

Read: 10 MB/s, 100 times/s

Log query

FAQs about log query

Question list

1. How to query an IP address in logs?
2. How to query a keyword containing a space in logs?
3. How to query logs based on two query conditions?
4. What methods does Log Service provide to query collected logs?
5. What query capabilities does Log Service provide?
6. What are the limits of log query?

Solutions

1. How to query an IP address in logs?

Log Service supports querying IP addresses in logs in the full match way. You can directly query logs related to an IP address, such as the logs containing or excluding the specified IP address. However, partial match is not supported, that is, you cannot query a part of an IP address directly because a dot (.) is not the default token of Log Service. You can filter the IP addresses on your own if needed. For example, download data by using SDKs first and then use a regular expression or string.indexof in the codes to determine the IP addresses.

For example, the query condition in a Log Service project is not ip:121.42.0 not status:200 not 360jk not DNSPod-Monitor not status:302 not jiankongbao not 301 and status:403.

The 121.42.0 Classless Inter-Domain Routing (CIDR) block addresses are still in the query results. This is because Log Service considers 121.42.0.x as a word and only queries the correct results if the query condition is 121.42.0.x. Therefore, the IP address 121.42.0 is not filtered in the query results if the query condition is not 121.42.0.

2. How to query a keyword containing a space in logs?

Logs containing the keyword at the left or right of the space can also be queried if you directly enter the keyword containing a space as the query condition. Therefore, we recommend that you enclose the keyword containing a space in quotation marks (“”) and use the contents in the quotation marks (“”) as a keyword to query the logs. Then, only logs with the keyword containing a space can be queried.

For example, query the logs containing the keyword POS version in the following logs:

```
post():351]:&amp;nbsp;device_id:&amp;nbsp;BTAddr&amp;nbsp;:&amp;nbsp;B6:xF:xx:65:xx:A1&amp;nbsp;IMEI&amp;nbsp;:&amp;nbsp;35847xx22xx81x9&amp;nbsp;WifiAddr&amp;nbsp;:&amp;nbsp;4c:xx:0e:xx:4e:xx&amp;nbsp;|&amp;nbsp;user_id:&amp;nbsp;bb07263xxd2axx43xx9exxea26e39e5f&amp;nbsp;POS&amp;nbsp;version:903
```

Logs containing POS or version are also included in the query results if you directly use POS version as the query condition, which does not meet the query requirement. Use "POS version" as the query condition and then all the logs containing the keyword POS version can be queried.

3. How to query logs based on two query conditions?

Enter two statements at the same time if you have two query conditions.

For example, to query logs whose status is not OK or Unknown in a Logstore, directly use not OK not Unknown as the query condition to query the logs.

4. What methods does Log Service provide to query

collected logs?

Log Service provides three methods to query logs:

1. Query logs in the Log Service console. For more information, see Query logs.
2. Use SDKs to query logs. For more information, see SDK.
3. Use RESTful APIs to query logs. For more information, see API.

5. What query capabilities does Log Service provide?

- Supports filtering and querying logs by using a combined condition. For more information about the query syntax, see Query syntax.
- Supports querying one billion logs in one second for a single query. You can query logs based on specified query conditions, read the time-based distribution of the query results, and obtain the raw logs.
- Supports caching logs, allowing you to obtain more complete query results for a second query with the same query condition.

6. What are the limits of log query?

- Supports querying logs based on a combined condition composed of at most 30 words.
- Supports obtaining at most 100 lines of raw logs for a single query. You can download more logs by turning the page.
- Supports processing one billion lines of logs within one second for a single query.

Logs cannot be queried

If logs cannot be queried in Log Service, troubleshoot the problem as follows.

Troubleshoot the problem

1. Logs are not collected successfully

Logs cannot be queried if they are not successfully collected to Log Service. Check if logs exist on the preview page. If yes, it means logs are successfully collected to Log Service and we recommend that you troubleshoot the problem for other reasons. If not, logs are not collected possibly because of any of the following reasons:

The log source does not generate logs.

No log is shipped to Log Service if the log source does not generate logs. Check your log source.

The machine group does not have the heartbeat.

Check if the machine has the heartbeat on the **Machine Group Status** dialog box. If not, see Logtail heartbeat error.

The monitored file is not written in real time.

If the monitored file is written in real time, open the file /usr/local/ilogtail/ilogtail.LOG to view the error message. Common errors are as follows:

- parse delimiter log fail: An error occurred when collecting logs by using delimiters.
- parse regex log fail: An error occurred when collecting logs by using a regular expression.

2. Incorrect token settings

Check if the keyword is obtained after the log is segmented by the configured tokens. For example, if the tokens are ,;=()[]{}?@&<>/:' as default, a log that contains abc" defg,hij is segmented into two parts: abc" defg and hij. Therefore, you cannot query this log by using abc.

Fuzzy query is also supported. For more information about the query syntax, see Query syntax.

Note:

- To save your index costs, Log Service optimizes the index and configures the key/value index keys, without using the full text index. Assume that you configure the key/value index, add a space as the token (add the space in the middle of the token string), and have a log that contains a key named **message**. The log message: this is a test message can be queried by using the key:value format message:this, but cannot be queried by using this directly. This is because you have configured the key/value index keys without the full text index.
- Creating an index or changing an existing index only works for new data.

You can check if the configured tokens meet the requirements in the **Index Attributes**.

3. Other reasons

If logs are generated, you can modify the time range for the query first. In addition, the log preview function provides data in real time, but the query function has a latency of up to one minute, so you

can wait one minute after logs are generated and then query the logs.

If the problem persists, open a ticket.

Differences between log consumption and log query

Log Service provides two functions related to the read operation.

Log collection and consumption (LogHub): Provides public channels for log collection and distribution, sequential (first in, first out (FIFO)) read and write of full data, and functions similar to Kafka.

- Each Logstore has one or more shards. Data is written to a shard at random.
- You can read logs in batches from a specified shard according to the sequence that logs are written to the shard.
- You can set the start point (cursor) to pull logs from shards in batches based on the time when Log Service receives logs.
- By default, logs are retained in LogHub for two days, during which logs can be consumed.

Log query (index): Log Service supports querying massive logs based on LogHub. You can query logs by using keywords.

- Use the keyword to query logs that meet your requirement.
- Supports using the boolean combination of AND, NOT, and OR to query logs based on keywords.
- You can only query logs in all shards, but not a specified shard.

Differences

Function	Log query (LogSearch)	Log collection and consumption (LogHub)
Query logs by using keywords	Supported	Not supported
Read small amounts of data	Fast	Fast
Read full data	Slow (100 logs every 100ms, not recommended)	Fast (1 MB logs every 10ms, recommended)
Read logs by topic	Yes	No. Logs are read by shard
Read logs by shard	No. You can only query logs in all shards	Yes. You must specify a shard for reading logs
Cost	Relatively high	Low

Scenario	Scenarios that need to filter data such as monitoring and troubleshooting	Full processing scenarios such as stream computing and batch processing

Common errors for log query and analysis

This document describes the common errors for log query and analysis. For more information about the basic syntaxes, see Syntax description.

List of common errors

1. line 1:44: Column ‘my_key_field’ cannot be resolved;please add the column in the index attribute
2. Column ‘xxxx__line__ ‘ not in GROUP BY clause;please add the column in the index attribute
3. sql query must follow search query,please read syntex doc
4. please read syntex document,and make sure all related fields are indexed. error after select .error detail:line 1:10: identifiers must not start with a digit; surround the identifier with double quotes
5. please read syntex document,and make sure all related fields are indexed. error after select .error detail:line 1:9: extraneous input ‘’ expecting

1. line 1:44: Column ‘my_key_field’ cannot be resolved;please add the column in the index attribute

Error cause: The key my_key_field does not exist. Therefore, you cannot reference the key for query.

Solution: On the query page, add this field as a key/value index and enable the analytics in the index attributes.

2. Column ‘xxxx__line__ ‘ not in GROUP BY clause;please add the column in the index attribute

Error cause: You use the GROUP BY syntax for query, but a non-agg field which is not contained in GROUP BY is referenced in SELECT. For example, in select key1, avg(latency) group by key2, key1 is not contained in GROUP BY.

Solution: The correct syntax is select key1,avg(latency) group by key1,key2.

3. sql query must follow search query,please read syntex doc

Error cause: The filter condition is not specified. For example, select ip,count(*) group by ip.

Solution: The correct syntax is *|select ip,count(*) group by ip.

4. please read syntex document,and make sure all related fields are indexed. error after select .error detail:line 1:10: identifiers must not start with a digit; surround the identifier with double quotes

Error cause: The column name or variable name referenced in SQL starts with a number, which is not allowed.

Solution: Modify the name to start with a letter.

5. please read syntex document,and make sure all related fields are indexed. error after select .error detail:line 1:9: extraneous input ‘’ expecting

Error cause: The word spelling is incorrect.

Solution: Find the incorrect word spelling according to the location specified in the error and correct the spelling.

Knowledge Base