日志服务

常见问题

常见问题

基本问题

问题列表

1. 日志服务是什么？

日志服务（Log Service，简称LS）是对日志收集、存储、订阅平台化服务。服务提供各种类型日志的实时收集，中心化管理、消费功能。

2. 日志服务可以用来做什么事情？

- 多种方式（API、SDK及Logtail接入服务）的日志写入途径。
- 通过Logtail自由定义日志的收集以及解析方式。
- 利用机器组管理数以千计机器上的日志收集。
- 提供实时日志消费与订阅功能。
- 简单易用的控制台配置方式，所有操作都可以在Web端完成。
- 后台与阿里云多个云产品无缝对接。

3. 日志服务的基本概念有哪些？

- 核心概念为：Project（项目、管理日志基础单元）、Logstore（日志库）、Shard（分区）、Topic（主题、对于Logstore二级分类）、Log（日志条数）、LogGroup（日志组）。
- 日志收集概念：Logtail配置（定义如何收集日志配置）、机器分组（分组）。

4. 日志服务有哪几部分组成？

主要有日志收集客户端、服务端以及其他系统。客户端目前有Windows、Linux版本日志收集Agent（Logtail），服务端处理日志服务API读写、以及配置操作，其它系统包括OSS等阿里云产品，即支持向OSS等云产品同步日志数据。

5. 日志服务如何定义一条日志？

日志包含三部分：时间（必填），日志内容（Key：Value对组成），以及元数据（Source，日志来源IP）。

日志管理

问题列表

1. 日志服务如何存储、管理用户的日志？
2. 删除日志库，日志数据是否丢失？
3. 日志服务日志保存多长时间？可否修改这个保存时限？
4. 希望把日志最终存储到OSS，怎样节省在日志服务上的花费？

1. 日志服务如何存储、管理用户的日志？

日志库（Logstore）是日志服务中的日志存储和查询的基本单元，通常用于存储一类日志数据。目前，支持在控制台或者通过API完成对日志库的增删改查操作。日志库创建完成后，用户通过API或SDK向指定日志库写入日志数据。如果用户希望收集阿里云ECS服务器的数据，日志服务提供的Logtail日志收集服务同样非常方便地收集到日志数据。

2. 删除日志库，日志数据是否丢失？

删除日志库会导致日志数据丢失，请谨慎操作。

3. 日志服务日志保存多长时间？可否修改这个保存时限？

日志服务有三项功能都与日志保存时间有关，分别如下：

- LogHub（日志中枢）/LogSearch（日志索引与查询）：根据需求自行设置。
- LogShipper（日志投递）：日志投递至OSS、MaxCompute后，生命周期在以上产品中设置。

4. 希望把日志最终存储到OSS，怎样节省在日志服务上的花费？

日志服务的索引分析提供强大功能的同时会产生一定费用，如果您的需求是将日志保存到OSS上，且没有自定义日志查询、分析等需求，可以通过以下方式削减账单费用。

注意事项

索引默认关闭，如您并未开启索引和分析功能，请修改Logstore数据保存时间减少数据存储费用。

修改关闭索引分析功能，会使得日志关键词查询、日志统计分析、Dashboard、告警等功能不可用，请谨慎操作。

节省费用方式

修改Logstore数据保存时间。

参考**操作Logstore**，修改Logstore数据保存时间为1天。日志服务收取一定的Logstore数据存储费用，您可以选择缩减存储时间以降低消费。

关闭索引功能。

开启**OSS投递功能**，将Logstore数据准实时投递到OSS保存。

进入Logstore列表页，单击**查询**。

删除索引以关闭索引分析功能。

执行以上步骤后，日志服务仅收取您很低的使用LogHub功能费用，了解更多请参考计费方式。

日志采集

日志采集基本问题

问题列表

1. 日志采集失败，应如何解决？
2. 采集到的日志数据是乱码，应如何解决？
3. 日志服务可以采集哪些日志？
4. 日志服务有哪些渠道采集日志？应该如何选择这些渠道？
5. 日志服务如何采集ECS上的日志？
6. 日志服务可以采集历史日志吗？
7. 日志服务采集数据的能力如何？有何限制？
8. Logtail采集NAS上的日志需要注意什么？

1. 日志采集失败，应如何解决？

1. 请检查匹配规则是否已通过（比较常见的是设置时候的日志和实际日志存在不一致）。
2. log文件是否实时更新，如果以前的老日志会不被采集。
3. 时间要含年份等日期信息。
4. 有延迟（日志服务读取大约1-2分钟）请耐心等待。
5. 机器组里查看对应的机器心跳是否fail。
6. 不支持非UTF-8编码的数据。
7. 请核实一下日志内的时间，比较常见因为时区问题导致时间过久被丢弃。

如果问题仍未解决，请联系售后技术支持。

2. 采集到的日志数据是乱码，应如何解决？

日志服务插入的数据要求是UTF-8编码的，如果是其他的字符集可能出现乱码的情况。

如果用户的数据是通过SDK插入的，可以在代码写入的时候做字符集转码；如果用户的数据是通过Logtail写入的，可以检查一下Logtail监控的日志文件的编码。

如问题还未解决,请联系售后技术支持

3. 日志服务可以采集哪些日志？

日志服务支持带有时间戳的文本日志和syslog，日志的时间必须是最近7天以内的，并且不能超过当前时间15分钟。

4. 日志服务有哪些渠道采集日志？应该如何选择这些渠道？

日志服务支持用户直接使用API写入；同时提供Linux和Windows版本的Logtail，用于采集磁盘文件上实时更新的日志。

1. 如果应用程序生成的日志不落磁盘，则可直接使用API写入到日志服务。
2. 实时写入磁盘的日志，可以通过Logtail来采集。

5. 日志服务如何采集ECS上的日志？

可以使用Logtail来采集ECS上落在磁盘上的日志，过程如下：

1. 在日志服务控制台上，首先创建一个Logstore
2. 配置Logtail采集的配置
3. 创建机器分组
4. 通过安装脚本自助安装Logtail客户端
5. 将Logtail的配置应用到需要的机器分组即可

6. 日志服务可以采集历史日志吗？

用户可以通过API写入7天以内的数据，7天之前的数据写入会失败。但是，Logtail暂不支持采集历史数据。

7. 日志服务采集数据的能力如何？有何限制？

用户可根据需求调整日志库（Logstore）的分区（Shard）数量。在ECS环境，Logtail采集的速率被限制在1MB/秒。

8. Logtail采集NAS上的日志需要注意什么？

例如Nginx accesslog采集场景，Web Server的nginx配置一般来说都是相同的，传统的方式会写在不同机器上相同名称的文件（Logtail可以正常采集）。使用NAS后，如果多台机器的Nginx日志写入了NAS的相同文件（并发写相同文件场景），Logtail采集日志会缺失或出错。因此，请注意在使用NAS时，保证不同Web Server的日志写入NAS中的不同文件。

Logtail基本问题

问题列表

1. Logtail是什么？

Logtail是日志服务提供的一种便于日志接入的日志采集客户端。通过在您的机器上安装Logtail来监听指定的日志文件并自动把新写入到文件的日志上传到您所指定的日志库。

2. Logtail是否可以采集静态不变的日志文件？

Logtail基于文件系统的修改事件来监听文件的变化，并将实时产生的日志发送到日志服务。如果日志文件没有发生任何修改行为，日志文件内容将不会被Logtail采集。

3. Logtail支持哪些平台？

目前支持Linux 64位和Windows Server2003（含）以后 32/64 位系统。

Linux：

- Aliyun Linux
- Ubuntu
- Debian
- CentOS
- OpenSUSE

Windows：

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

4. 如何安装、升级Logtail客户端？

安装：目前需要用户通过安装脚本自助安装Logtail客户端。升级：Logtail客户端的升级由日志服务定期完成，升级过程数据采集不中断。

5. 如何配置使用Logtail客户端？

请参考：控制台配置Logtail采集日志说明。

6. Logtail如何工作？

1. 用户在控制台配置需要监控的目录、日志文件名以及相应的解析规则（正则表达式）等。
2. 用户机器上，日志文件发生修改，Logtail收到来自文件系统的事件并读取新产生的日志。
3. Logtail根据正则表达式解析日志格式并发往日志服务。

7. Logtail是否支持日志轮转？

对于日志文件a.LOG，当文件达到一定大小或创建超过一定时间后，a.LOG被mv为a.LOG.1（或其它名称），然后新建一个a.LOG继续写入日志。这个过程称为轮转。Logtail基于文件系统的事件通知，可以自动处理日志轮转的场景。

8. Logtail如何处理网络异常？

网络异常、写入Quota满时，Logtail会将采集到的日志内容写入本地磁盘缓存，并在稍后进行重试。磁盘缓存最大支持500MB，新缓存会覆盖旧缓存；超过24小时未发送成功的缓存文件将被自动删除。

9. Logtail日志采集延时如何？

Logtail基于事件进行日志采集，一般会在3秒内将日志发往日志服务。

10. Logtail如何处理历史日志？

Logtail只用于采集实时日志，如果日志内容的时间与Logtail处理该日志的系统时间相差5分钟以上，即被认为是历史日志。

11. 日志服务修改日志采集配置后多久可以生效？

用户在控制台应用配置到机器组后，Logtail最迟会在3分钟之内加载新配置并生效。

12. 如何调查Logtail采集日志问题？

完整步骤logtail日志采集异常排查。常见问题如下：

1. 查看Logtail心跳是否正常，如不正常，请尝试重新安装Logtail。
2. 确认日志采集配置中的日志文件是否为实时生成。
3. 查看日志采集配置的正则表达式是否与日志内容相匹配。如正则匹配错误，可以在Logtail运行日志查看到相关错误。错误日志路径Linux:/usr/local/ilogtail/ilogtail.LOG。

Logtail 机器无心跳

配置Logtail采集日志数据时，如果Logtail机器组心跳状态不正常，可使用Logtail自动诊断工具或人工诊断的方式排查问题。

自动诊断

日志服务提供Logtail自动诊断工具，排查步骤请参考 Logtail自动诊断工具。

注意： 如果诊断正常，请参考诊断工具的回显信息、参考**人工诊断**结果，查看是否出现异常。

人工诊断

Logtail心跳失败一般由以下原因造成，请逐个排查。

1. 网络未联通

请执行以下命令查看网络连通性，确保网络正常。

经典网络

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

VPC网络

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

公网

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

2. 未安装Logtail

请执行以下命令查看客户端状态，如未安装Logtail客户端，请参考Logtail安装，务必按照您日志服务Project所属Region以及网络类型进行安装。

Linux查看客户端状态:

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

Windows查看客户端状态：

```
控制面板 -> 管理工具 -> 服务
查看LogtailDaemon、LogtailWorker两个Windows Service运行状态。
```

3. 安装时所选参数错误

日志服务是地域化的，需要在安装时为客户端指定正确的服务端访问入口，请查看您已安装的客户端使用的配置：

- 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

  确认以下两点：

  客户端连接的网络入口所属Region是否与您Project所在Region一致。网络入口列表参考服务入口。

- 是否根据您的机器所属网络环境选择了正确的域名。如VPC环境如果选择了内部域名，是无法联通的。可以Telnet测试ilogtail_config.json中配置的域名，如：telnet logtail.cn-hangzhou-intranet.log.aliyuncs.com 80。

4. 服务端配置了错误的IP或用户标识

一般来说，Logtail在机器上获取IP的方式为：

- 如果本机在文件/etc/hosts中设置了主机名绑定，需要确认绑定的IP。执行命令hostname可以查看主

机名。

如果没有设置主机名绑定，会取本机的第一块网卡IP。

在服务器上查看IP地址：

- 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

- **注意：** 如果app_info.json文件中ip字段为空，logtail无法工作。此时需为主机设置ip地址并重启logtail

如果服务端机器组内填写的IP与客户端获取的IP不一致，则根据情况进行修改：

若服务端机器组填写了错误IP，请修改机器组内IP并保存，等待1分钟再查看。

若修改了机器上的网络配置（如/etc/hosts修改），请重新启动Logtail以获取新的IP。

如有需要，可以执行以下命令重启Logtail。

- Linux：sudo /etc/init.d/ilogtaild stop; sudo /etc/init.d/ilogtaild start
- Windows : **控制面板** -> **管理工具** -> **服务** -> **重启LogtailWorker**

5. 非ECS或者日志服务Project和ECS非同一账号

以下两种情况需要为logtail安装的机器授权收集日志，具体步骤参考 aliuid配置。

1. 非ECS机器
2. 创建日志服务Project和购买ECS的账号不是同一账号

如果您的问题仍未解决，请提工单联系我们（工单中请提供您的Project、Logstore、机器组、app_info.json、ilogtail_config.json以及自助诊断工具的输出内容）。

Logtail 快速诊断工具

当日志采集发生异常时，用户可通过Logtail自助检测工具查看客户端是否存在异常情况，根据工具提示快速定位并解决问题。

准备工作

下载检测工具脚本

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

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

注意事项

- 检查工具需要使用curl进行网络连通性检查，请确保机器安装curl工具。

使用方法

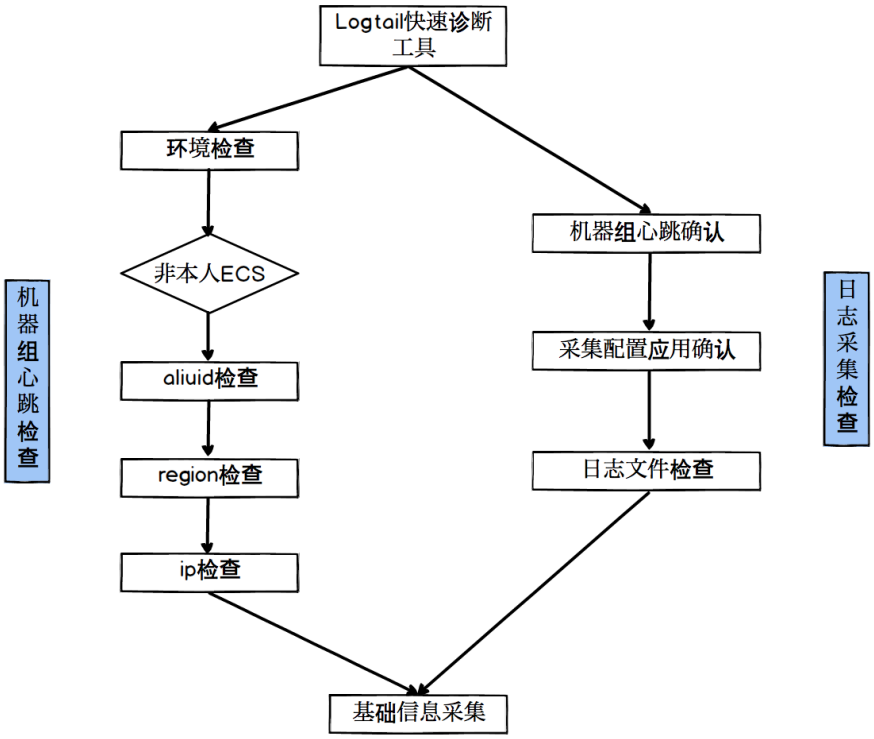
脚本会根据您的选择执行不同检查流程，运行命令如下

```
chmod 744 ./checkingtool.sh
sh checkingtool.sh
[Info]: Logtail checking tool version : 0.3.0
[Input]: please choose which item you want to check :
1. MachineGroup heartbeat fail.
2. MachineGroup heartbeat is ok, but log files have not been collected.
Item :
```

请根据提示选择检查项目：

1. 机器组心跳失败，输入1并回车确认
2. 机器组心跳成功，但日志文件没有被采集，输入2并回车确认

整体检查流程如下图所示：



机器组心跳检查流程

选择机器组心跳检查流程后会进行下述一系列的检查：

基础环境检查

基础环境检查包括：

1. Logtail是否安装
2. Logtail是否运行
3. SSL状态是否正常
4. 网络是否可以连通日志服务

```
[Info]:     Logtail checking tool version : 0.3.0
[Input]: please choose which item you want to check :
1. MachineGroup heartbeat fail.
2. MachineGroup heartbeat is ok, but log files have not been collected.
Item :1
[Info]: Check logtail install files
[Info]: Install file: ilogtail_config.json exists. [ OK ]
[Info]: Install file: /etc/init.d/ilogtaild exists. [ OK ]
[Info]: Install file: ilogtail exists. [ OK ]
[Info]: Bin file: /usr/local/ilogtail/ilogtail_0.14.2 exists. [ OK ]
[Info]: Logtail version : [ OK ]

[Info]: Check logtail running status
[Info]: Logtail is runnings. [ OK ]
[Info]: Check network status
[Info]: Logtail is using ip: 11.XX.XX.187
[Info]: Logtail is using UUID: 0DF18E97-0F2D-486F-B77F-XXXXXXXXXXXX
[Info]: Check SSL status
[Info]: SSL status OK. [ OK ]
[Info]: Check logtail config server
[Info]: config server address: http://config.sls.aliyun-inc.com
[Info]: Logtail config server OK [ OK ]
```

若其中检查出现Error信息，请参考提示进行处理。

确认是否非本人ECS

基础环境检查通过后，请确认您的服务器是否为非本人ECS。

若此服务器不是ECS或者ECS购买账号和日志服务账号不同，输入y，否则输入N。

```
[Input]:  Is your server non-Alibaba Cloud ECS or not belong to the same account with the current Project of Log Service ? (y/N)
```

当输入y后，检查工具会输出本地配置的aliuid信息，请确认其中是否包含了您的aliuid，若未包含请参考文档创建aliuid标识。

```
[Input]:  Is your server non-Alibaba Cloud ECS or not belong to the same account with the current Project of Log Service ? (y/N)y
[Info]: Check aliyun user id(s)
[Info]: aliyun user id : 126XXXXXXXXXX79 . [ OK ]
[Info]: aliyun user id : 165XXXXXXXXXX50 . [ OK ]
[Info]: aliyun user id : 189XXXXXXXXXX57 . [ OK ]

[Input]: Is your project owner account ID is the above IDs ? (y/N)
```

安装Region检查

请确认您的Project所在区域是否和Logtail安装时所选区域一致，若不一致请重新安装Logtail。

```
[Input]:  please make sure your project is in this region : { cn-hangzhou } (y/N) :
```

ip配置检查

请确认您机器组配置的ip和Logtail工作ip一致，若不一致请参考机器组管理修改。

若您配置的是自定义标识机器组，请确认本地配置的标识与服务端配置一致，若不一致请参考自定义标识机器组修改。

```
[Input]:  please make sure your machine group's ip is same with : { 11.XX.XX.187 } or your machine group's userdefined-id is in : { XX-XXXXX } (y/N) :
```

心跳正常但日志未采集检查流程

选择日志未采集检查流程后会进行下述一系列的检查：

ip配置确认

请确认您机器组配置的ip和Logtail工作ip一致且心跳正常，若不一致请订正机器组配置。

```
[Input]:  please make sure your machine group's ip is same with : { 11.XX.XX.187 } (y/N) :
```

采集配置应用确认

请确认您的采集配置已经成功应用到该机器组中，如何查看机器组应用配置参见机器组配置管理。

```
[Input]:  please make sure you have applied collection config to the machine group (y/N) :Y
```

日志文件检查

检查时请输入您需要检查的日志文件**全路径**，若未找到匹配项，请确认配置的路径信息可以匹配给定的日志文件。

若配置错误请重新修改采集配置并保存，1分钟后再次执行此脚本重新检查。

```
[Input]:  please input your log file's full path (eg. /var/log/nginx/access.log) :/disk2/logs/access.log
[Info]: Check specific log file
[Info]: Check if specific log file [ /disk2/logs/access.log ] is included by user config.
[Warning]: Specific log file doesnt exist. [ Warning ]
[Info]: Matched config found: [ OK ]
[Info]: [Project] -> sls-zc-xxxxxx
[Info]: [Logstore] -> release-xxxxxxx
[Info]: [LogPath] -> /disk2/logs
[Info]: [FilePattern] -> *.log
```

检查通过但采集依然异常

若所有的检查全部通过，但采集依然出现异常，请在脚本最后的选择中输入y并回车确认。

请您将检查脚本输出的信息作为附件，提交工单给我们的售后工程师。

```
[Input]:  please make sure all the check items above have passed. If the problem persists, please copy all the outputs and submit a ticket in the ticket system. : (y/N)y
```

快速检查

快速检查运行时无需确认，可用于二次封装自定义检查脚本。

- **注意：** 快速检查运行时会输出客户端配置的**阿里云ID**和**动态机器组/自定义标识**，不存在时并不会给出告警，如果客户端需要阿里云ID或动态机器组/自定义标识的配置，请查看工具的输出和您配置的是否一致，不一致时按照以下方法重新配置：阿里云UserId配置、动态机器组配置。

快速检查使用方法

请运行脚本./checkingtool.sh --logFile [LogFileFullPath]进行检查。 检测脚本发现异常时，请根据脚本提示进行处理。

- **注意：** 若指定日志文件检查通过且Logtail运行环境正常，建议进入阿里云控制台中查看该日志服务配置项的异常日志，参见日志收集错误查询。

Logtail采集异常的常见问题

运行Logtail快速诊断工具后，可以诊断出Logtail采集异常的原因，您可以根据具体原因查找对应的解决方案。常见Logtail采集问题原因及解决方案如下。

常见问题	解决方法

安装文件丢失	重装Logtail。
Logtail未运行	使用命令/etc/init.d/ilogtaild start开启。
多个Logtail进程	使用命令/etc/init.d/ilogtaild stop关闭，再用命令/etc/init.d/ilogtaild start开启。
443端口被禁用	防火墙打开443端口。
无法找到配置服务器	确认安装正确性，若安装错误，卸载后重新安装。
不存在用户配置	确认控制台已经创建好Logtail配置、机器组中包含该客户端且已经将配置应用到机器组。
没有匹配指定日志文件	确认Logtail配置正确性。
指定日志文件匹配多次	多个匹配时Logtail随机选择一个配置，建议去重。

检测工具常用参数

- 直接运行进行常规检查流程
- --help查看帮助文档
- --logFile [LogFileFullPath]检测Logtail是否收集路径为LogFileFullPath的日志，同时检查基本的Logtail运行环境(安装文件完整性、运行状态、阿里云userID、网络连通性等)
- --logFileOnly [LogFileFullPath]只检测Logtail是否收集路径为LogFileFullPath的日志
- --envOnly 只检测Logtail运行环境

日志采集功能与 Kafka对比

Kafka是分布式消息系统，由于其高吞吐和水平扩展，被广泛使用于消息的发布和订阅。以开源软件的方式提供，各用户可以根据需要搭建Kafka集群。

日志服务(Log Service)是基于飞天Pangu构建的针对日志平台化服务。服务提供各种类型日志的实时采集，存储，分发及实时查询能力。通过标准话的Restful API对外提供服务。

Log Service Loghub提供公共的日志采集、分发通道，用户如果不想自己搭建、运维kafka集群，可以使用Log Service LogHub功能。

Log Service Loghub & Kafka 概念映射

概念	Kafka	Loghub
存储对象	topic	logstore
水平分区	partition	shard
数据消费位置	offset	cursor

Loghub & Kafka 功能比较

功能	Kafka	LogHub
使用依赖	自建或共享Kafka集群	Log Service服务
通信协议	TCP 打通网络	Http (restful API)，80端口
访问控制	无	基于云账号的签名认证+访问控制
动态扩容	暂无	支持动态shard个数弹性伸缩(Merge/Split)，对用户无影响
多租户Qos	暂无	基于shard的标准化流控
数据拷贝数	用户自定义	暂不开放，默认3份拷贝
failover/replication	调用工具完成	自动，用户无感知
扩容/升级	调用工具完成，影响服务	用户无感知
写入模式	round robin/key hash	暂只支持round robin/key hash
当前消费位置	保存在kafka集群的zookeeper	服务端维护、无需关心
保存时间	配置确定	根据需求动态调整

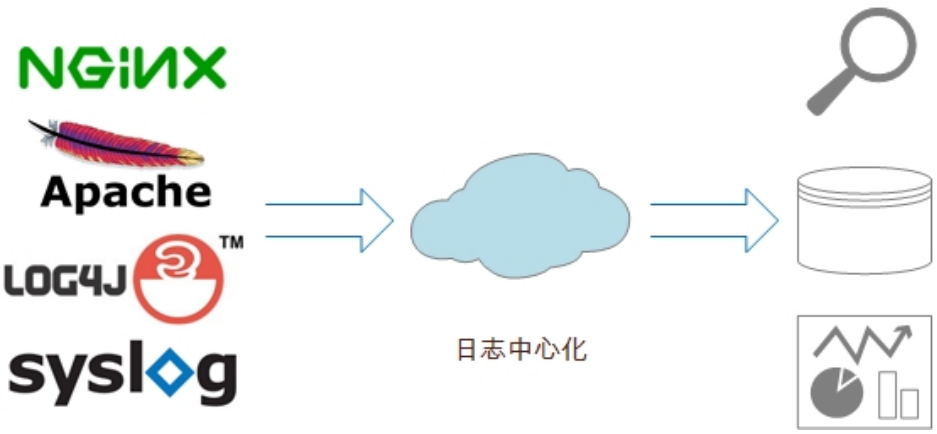
成本对比

参见成本对比（其他方案）下LogHub部分。

日志采集Agent对比

日志采集场景下客户端测评

DT时代，数以亿万计的服务器、移动终端、网络设备每天产生海量的日志。中心化的日志处理方案有效地解决了在完整生命周期内对日志的消费需求，而日志从设备采集上云是始于足下的第一步。



三款日志采集工具

Logstash

- 开源界鼎鼎大名ELK stack中的”L”，社区活跃，生态圈提供大量插件支持
- Logstash基于JRuby实现，可以跨平台运行在JVM上
- 模块化设计，有很强的扩展性和互操作性。

Fluentd

- 开源社区中流行的日志采集工具，td-agent是其商业化版本，由Treasure Data公司维护，是本文选用的评测版本。
- Fluentd基于CRuby实现，并对性能表现关键的一些组件用C语言重新实现，整体性能不错。
- Fluentd设计简洁，pipeline内数据传递可靠性高
- 相较于Logstash，其插件支持相对少一些。

- Logtail

- 阿里云日志服务的生产者，经过3年多阿里集团大数据场景考验
- 采用C++语言实现，对稳定性、资源控制、管理等下过很大的功夫，性能良好
- 相比于Logstash、Fluentd的社区支持，Logtail功能较为单一，专注日志采集功能。

功能对比

功能项	Logstash	Fluentd	Logtail
日志读取	轮询	轮询	事件触发
文件轮转	支持	支持	支持
Failover处理 (本地checkpoint)	支持	支持	支持
通用日志解析	支持grok（基于正则表达式）解析	支持正则表达式解析	支持正则表达式解析
特定日志类型	支持delimiter、key-	支持delimiter、key-	支持delimiter、key-

	value、json等主流格式	value、json等主流格式	value、json等主流格式
数据发送压缩	插件支持	插件支持	LZ4
数据过滤	支持	支持	支持
数据buffer发送	插件支持	插件支持	支持
发送异常处理	插件支持	插件支持	支持
运行环境	JRuby实现，依赖JVM环境	CRuby、C实现，依赖Ruby环境	C++实现，无特殊要求
线程支持	支持多线程	多线程受GIL限制	支持多线程
热升级	不支持	不支持	支持
中心化配置管理	不支持	不支持	支持
运行状态自检	不支持	不支持	支持cpu/内存阈值保护

日志文件采集场景 - 性能对比

日志样例:以Nginx的access log为样例，如下一条日志365字节，结构化成14个字段：

在接下来的测试中，将模拟不同的压力将该日志重复写入文件，每条日志的time字段取当前系统时间，其它13个字段相同。

相比于实际场景，模拟场景在日志解析上并无差异，有一点区别是：较高的数据压缩率会减少网络写出流量。

Logstash

logstash-2.0.0版本，通过grok解析日志并写出到kafka（内置插件，开启gzip压缩）。

日志解析配置：

```
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"]
}
```

测试结果：

写入TPS	写入流量 (KB/s)	CPU使用率 (%)	内存使用 (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版本，通过正则表达式解析日志并写入kafka（第三方插件fluent-plugin-kafka，开启gzip压缩）。

日志解析配置：

```
<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>
```

测试结果：

写入TPS	写入流量 (KB/s)	CPU使用率 (%)	内存使用 (MB)
500	178.22	13.5	61
1000	356.45	23.4	61
5000	1782.23	94.3	103

注：受GIL限制，Fluentd单进程最多使用1个cpu核心，可以使用插件multiprocess以多进程的形式支持更大的日志吞吐。

Logtail

logtail 0.9.4版本，设置正则表达式进行日志结构化，数据LZ4压缩后以HTTP协议写到阿里云日志服务，设置batch_size为4000条。

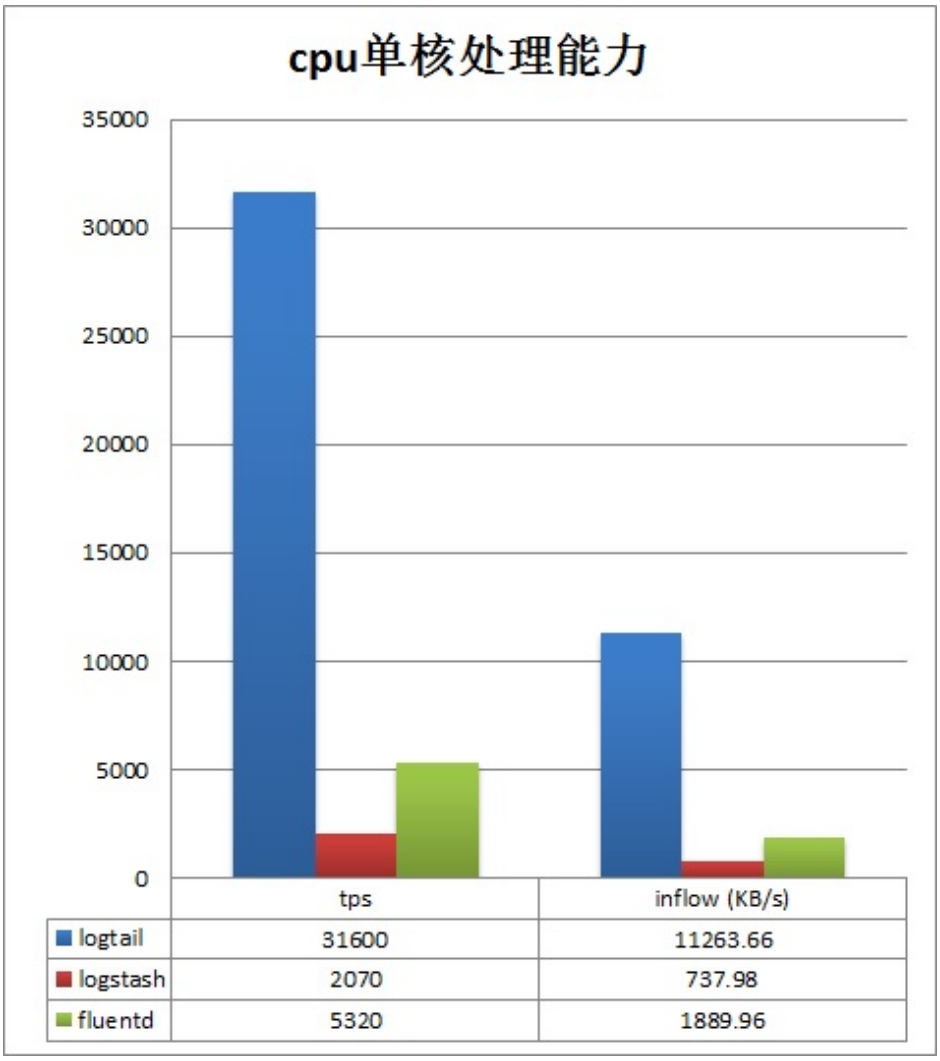
日志解析配置：

```
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
```

测试结果：

写入TPS	写入流量 (KB/s)	CPU使用率 (%)	内存使用 (MB)
500	178.22	1.7	13
1000	356.45	3	15
5000	1782.23	15.3	23
10000	3564.45	31.6	25

单核处理能力对比



总结

可以看到三款日志工具各有特点：

- Logstash支持所有主流日志类型，插件支持最丰富，可以灵活DIY，但性能较差，JVM容易导致内存使用量高。
- Fluentd支持所有主流日志类型，插件支持较多，性能表现较好。
- Logtail占用机器CPU/内存资源最少，性能吞吐量较好，针对常用日志场景支持全面，但缺少插件等机制，灵活性和可扩展性不如以上两个客户端。

日志查询

日志查询常见问题

问题列表

1. 如何在日志数据中搜索IP地址？
2. 如何在日志中搜索包含空格的关键字？
3. 如何完成双重条件检索？
4. 日志服务提供哪些渠道查询采集的日志？
5. 日志服务提供什么样的查询能力？
6. 日志查询有什么限制？

解决方案

1. 如何在日志数据中搜索IP地址？

在日志数据中搜索IP地址，支持全部匹配的方式检索。您可以直接在日志数据中直接搜索指定IP地址相关的日志信息，比如包含指定IP地址、过滤指定IP地址等。但是目前尚不支持部分匹配的方式检索，即不能直接搜索IP地址的一部分，因为小数点不是日志服务默认的分词项。如果需要的话，建议自行过滤，比如用SDK先下载数据，然后在代码里用正则或者用string.indexof等方法判断。

例如，在日志服务的Project中搜索条件如下：not ip:121.42.0 not status:200 not 360jk not DNSPod-Monitor not status:302 not jiankongbao not 301 and status:403。

搜索结果中仍会出现121.42.0网段地址。因为日志服务会认为121.42.0.x是一个词，所以只有搜121.42.0.x能搜到结果，而121.42.0的话不会搜到这个结果，同理加上not也就不会过滤该地址。

2. 如何在日志中搜索包含空格的关键字？

搜索包含空格的关键字时，如果直接搜索，则会得到包含空格左侧关键字或右侧关键字的所有日志。建议您在查询的包含空格的关键字时，把关键字用双引号包裹起来，将引号中的内容作为一个关键字进行搜索，搜索结果就是符合条件的日志内容。

例如，在以下日志中搜索包含关键字POS version的日志。

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

如果直接搜索POS version，则会得到包含POS或者version的所有日志，不符合搜索要求。如果搜索"POS version"，则会得到包含关键字POS version的所有日志。

3. 如何完成双重条件检索？

双重条件检索时，只需同时输入两个语句即可。

例如，需要在Logstore中搜索数据状态不是OK或者Unknown的日志。直接搜索not OK not Unknown即可得到符合条件的日志。

4. 日志服务提供哪些渠道查询采集的日志？

日志服务提供了三种方式查询日志：

1. 通过日志服务控制台查询。
2. 通过SDK查询。详见SDK。
3. 通过Restful API查询，详见API。

5. 日志服务提供什么样的查询能力？

- 提供组合条件过滤查询，查询语法参见日志查询语法。
- 能够提供单次查询10亿/S日志的能力。用户可以根据一定的条件筛选出需要的日志，读取命中日志在时间维度上的分布，或者拿到原始日志。
- 查询提供了cache的功能，第二次查询相同的条件获得更加完整的查询结果。

6. 日志查询有什么限制？

- 最多能够查询30个词组成的组合条件。
- 单次查询结果最多获取100行原始数据，可以通过翻页下载更多日志。
- 单次查询1秒内可以处理10亿行数据。

日志消费与查询区别

日志消费与日志查询区别

日志服务提供了两项功能都和“读”有关：

日志收集与消费（LogHub)：提供公共的日志收集、分发通道。全量数据顺序（FIFO）读写，提供类似Kafka的功能

- 每个LogStore有一个或多个Shard，数据写入时，随机落到某一个shard中
- 可以从指定shard中，按照日志写入shard的顺序批量读取日志
- 可以根据server端接收日志的时间，设置批量拉取shard日志的起始位置（cursor）

日志查询（Search/Analytics）：在LogHub基础上提供海量日志查询+分析功能，根据条件进行日志查询与统计

- 通过查询条件查找符合要求的数据
- 支持关键词 AND、NOT、OR的布尔组合和结果SQL统计
- 数据查询不区分shard

两者区别：

功能	日志查询(LogSearch)	日志收集与消费(LogHub)
关键词查找	支持	不支持
小量数据读取	快	快
全量数据读取	慢(100条日志100ms，不建议通过该方式读取数据)	快（1MB日志10ms，推荐方式）
读取是否区分topic	区分	不区分，只以shard作为标识
读取是否区分shard	不区分，查询所有shard	区分，单次读取需要指定shard
费用	较高	低
适用场景	监控、问题调查与分析等场景	流式计算、批量处理等全量处理场景

日志查询分析常见报错

本文档主要介绍日志服务常见的查询分析报错，基本语法请查看分析语法。

常见报错列表

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. key word(where) is not supported,please read query syntex # | *select apiName,count()* as count where apiName=”” group by apiName order by count desc limit 10
5. 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
6. 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

报错原因：my_key_field这个Key不存在，所以您在query中无法引用该Key。

解决方案：在查询页面，右上角查询分析属性里，添加该字段为键值索引，同时打开统计功能。

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

报错原因：您在查询中使用了GROUP BY语法，但是在Select中引用了一个非agg字段，该字段没有出现在GROUP BY中。例如select key1, avg(latency) group by key2 ，key1没有出现在GROUP BY中。

解决方案：正确语法是select key1,avg(latency) group by key1,key2。

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

报错原因：没有指定filter条件，例如select ip,count(*) group by ip。

解决方案：正确的写法为*|select ip,count(*) group by ip。

4. key word(where) is not supported,please read query syntex # | *select apiName,count()* as count where

apiName=”” group by apiName order by count desc limit 10

报错原因：Where条件出现在了SQL语法中，不符合规范。

解决方案：将Where放在filter条件中，正确语法为apiName:"" | select apiName,count(*) as count group by apiName order by count desc limit 10。

5. 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

报错原因：SQL中引用到的列名、变量名等以数字开头，不符合规范。

解决方案：建议更改该名称，以字母开头。

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

报错原因：有单词拼写错误。

解决方案：请根据报错中指出的错误位置，修改至正确。

日志投递

日志服务常见报错

日志服务报错内容：

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. 请检测IP是否正确，现在只支持本区域的云服务器
5. 不同操作系统或无效的IP，现在只支持本区域的云服务器
6. ShardWriteQuotaExceed

解决方法

1. illegal param! [LogContent] is null

请检查以下配置：

是否已经填写日志样例的内容。

检查首行正则是否填写正确。

如果以上排查过程均无法解决问题，用户可以把日志样例和正则表达式提供给我们，以便复现并解决。

如问题还未解决，请联系售后技术支持。

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

如果您在日志服务的使用过程中，发现ilogtail.log里出现类似报错：

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

这是因为写入的quota不足导致的。用户写入的日志量大于规定的阈值，或写入速度超出限制。

目前，每个分区处理能力为：写入：5MB/s，2000次/s读取：10MB/s，100次/s

如果您的数据量超出以上处理能力，您可以选择分裂分区，详情请参考分裂分区。Project级别写入日志每分钟最高请求次数为30W。如果您是通过程序写入日志，且会有部分请求超出quota，建议您批量写入日志或使用Producer-lib，批量限制为每次上传数据包的大小不超过3M、数量不超过4096条。

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

用户的ilogtail.log可能会出现以下报错：

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

这个报错是旧日志导致的。

Logtail收集日志文件规则：

历史数据单独处理。

减少数据落盘的缓存时间，甚至能做到实时落盘。

修改日志内容，注意时区问题。

用户分配排查后的解决方法：

新文件被Logtail监控时，日志文件里的1分钟前的日志会被认为是旧数据而丢弃。

文件里新写入的数据，如果是5分钟以前的数据也会被认为是就是而丢弃。出现这个报错的原因基本是因为用户的日志缓存在内存里，等到真正写入文件的时候已经超时了。

如果日志时间超过-7天~360s之间的范围，会被服务端丢弃。出现这个报错的原因基本是因为用户的日志设置的时区有问题。

如果违反了规则1，历史的数据会被丢弃。但是如果没有违反后续的规则的话，那后面的日志就不会报错。

如果违反了规则2，会偶尔报错，控制台能查到部分日志。

如果违反了规则3，有的日志会被没收集起来，控制台查不到日志。

如果问题还未能解决，请联系售后技术支持。

4. 请检测IP是否正确，现在只支持本区域的云服务器

在控制台上为机器组添加机器时，如果提示“请检测IP是否正确，现在只支持本区域的云服务器”，是因为您在添加机器组时未正确填写服务器的内网IP地址。

出现该提示时，请核对以下配置，并填写正确的内网IP地址。

请确保您填写的云主机IP为此登陆云账号所有。

目前只支持当前Project所在区域的云服务器，如当前Project是杭州节点，需要添加杭州节点的服务器IP。

必须填写云服务器实例的内网IP，多个IP需使用换行分割。

只能添加普通ECS服务器，VPC服务器是无法添加的。

如问题还未解决,请联系售后技术支持。

5. 不同操作系统或无效的IP，现在只支持本区域的云服务器

在控制台上为机器组添加机器时，如果提示“不同操作系统或无效的IP，现在只支持本区域的云服务器”，是由于同一机器组中不允许同时存在Windows与Linux云服务器，也就是您添加的服务器需要同为Linux或同为Windows系统。

请正确配置后再添加机器。

如问题还未解决，请联系售后技术支持。

6. ShardWriteQuotaExceed

ShardWriteQuotaExceed报错表示您的Shard分区比较少，写入超过了限制。您可以参考下文档，扩容分区数量。

目前，日志服务每个分区可提供一定的服务能力：

写入：5MB/s，2000次/s

读取：10MB/s，100次/s

日志消费