阿里云
日志服务

常见问题

文档版本：20181108

法律声明

阿里云提醒您在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。
2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的”现状“、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含”阿里云”、Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

格式	说明	样例
	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	 **禁止：** 重置操作将丢失用户配置数据。
	该类警示信息可能导致系统重大变更甚至故障，或者导致人身伤害等结果。	 **警告：** 重启操作将导致业务中断，恢复业务所需时间约10分钟。
	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	 **说明：** 您也可以通过按**Ctrl** + **A**选中全部文件。
>	多级菜单递进。	设置 > 网络 > 设置网络类型
粗体	表示按键、菜单、页面名称等UI元素。	单击 **确定**。
`courier` 字体	命令。	执行 `cd /d C:/windows` 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	`bae log list --instanceid` *`Instance_ID`*
[]或者[a|b]	表示可选项，至多选择一个。	`ipconfig` *`[-all|-t]`*
{}或者{a|b}	表示必选项，至多选择一个。	`swich` *`{stand | slave}`*

目录

1 基本问题

为您介绍日志服务的基本问题。

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。

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。

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）。

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。

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）。

2 日志管理

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。

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。

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。

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。

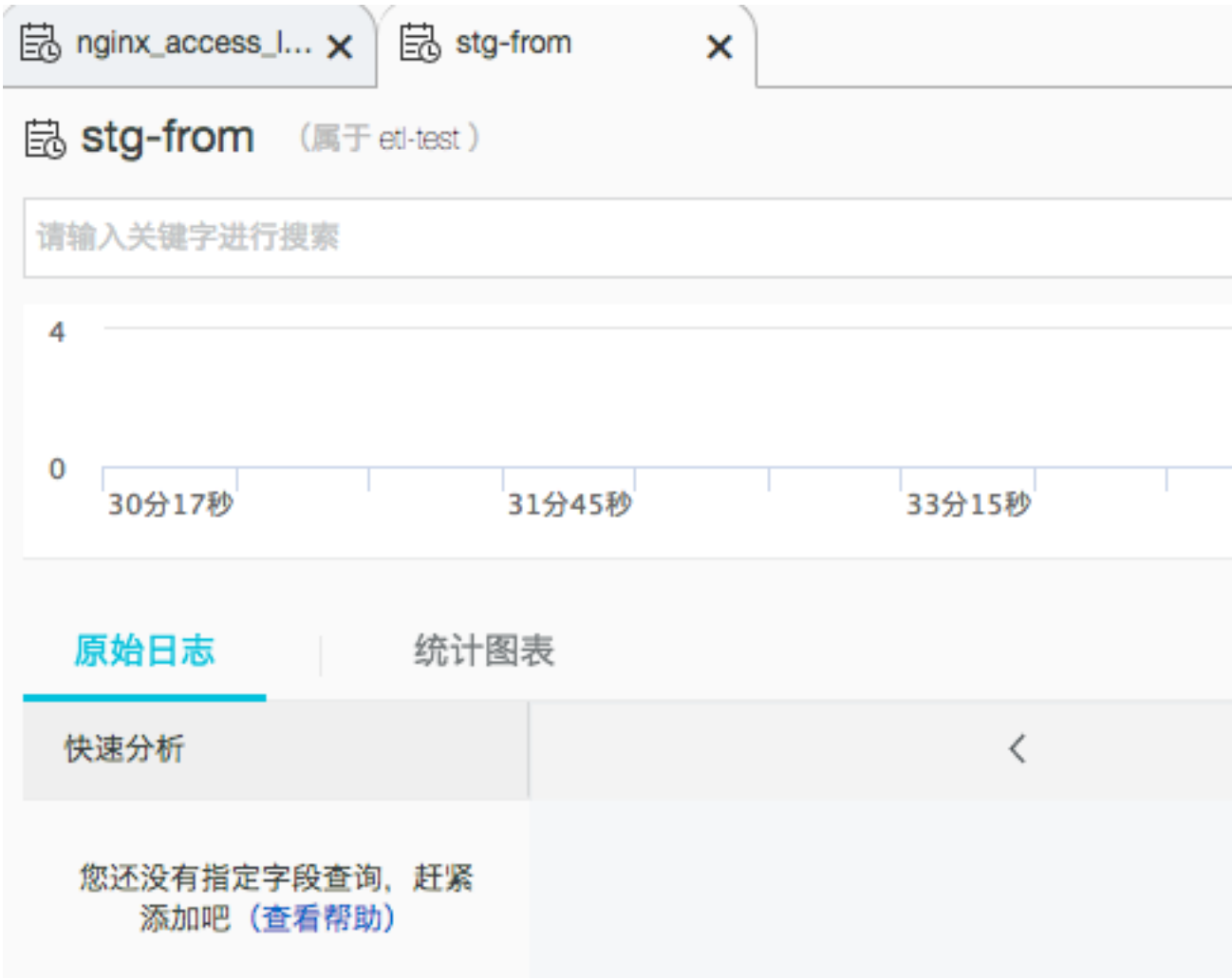
注意事项

- 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。
- 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。

  1. 开启*OSS*投递功能，将Logstore数据准实时投递到OSS保存。
  2. 在**Logstore**列表页，单击**查询**。

3. 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。

3 日志采集

3.1 日志服务常见报错

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。

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，500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，详情请参考操作*Shard*。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 Library*，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。

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。

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。

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。

ShardWriteQuotaExceed

ShardWriteQuotaExceed报错表示您的Shard分区比较少，写入超过了限制。您可以参考操作*Shard*，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，500次/s
- 读取：10MB/s，100次/s

3.2 日志采集基本问题

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。

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。

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。

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。

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. 通过安装脚本自助安装Logtail客户端。
2. 在日志服务控制台上，创建Project和Logstore。
3. 创建机器组。
4. 配置Logtail采集的配置。
5. 将Logtail的配置应用到需要的机器分组。

详细步骤请参考采集*ECS*日志。

日志服务可以采集历史日志吗？

Logtail默认只采集增量的日志文件，如果您需要导入历史文件，可使用Logtail自带的导入历史文件功能。

Logtail基于事件进行文件采集，事件通常由监听或定期轮询文件修改产生。除以上方式外，Logtail还支持从本地文件中加载事件，以此驱动日志采集。历史文件采集就是基于本地事件加载实现的功能。详细说明请参考导入历史日志文件。

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。

Logtail采集NAS上的日志需要注意什么？

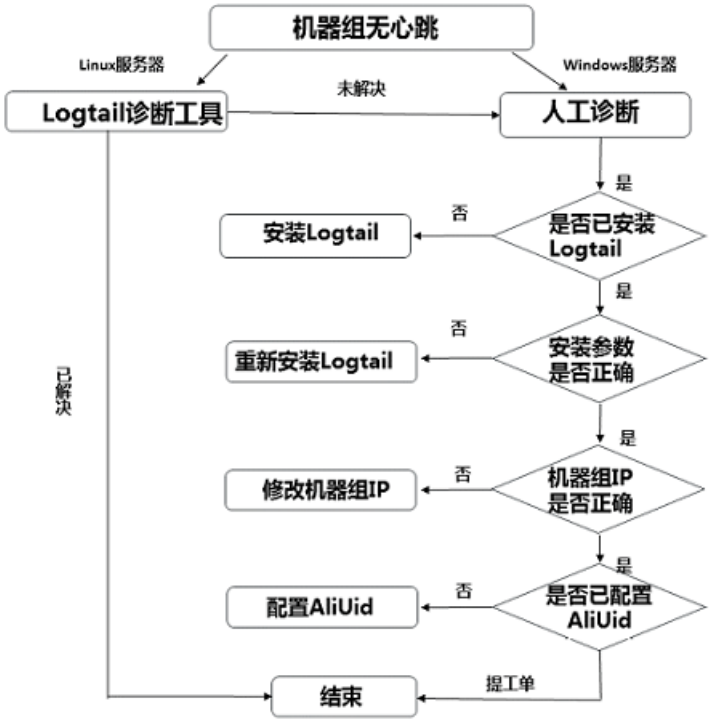
例如Nginx accesslog采集场景，Web Server的nginx配置一般来说都是相同的，传统的方式会写在不同机器上相同名称的文件（Logtail可以正常采集）。使用NAS后，如果多台机器的Nginx日志写入了NAS的相同文件（并发写相同文件场景），Logtail采集日志会缺失或出错。因此，请注意在使用NAS时，保证不同Web Server的日志写入NAS中的不同文件。

3.3 Logtail 机器无心跳

配置Logtail采集日志数据后，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。

如果使用Logtail采集日志，在服务器上安装Logtail之后，Logtail会定时向服务端发送心跳包。如果机器组状态页面显示机器无心跳，说明客户端和服务端未成功联通。日志服务提供*Logtail*自动诊断工具和人工诊断步骤，您可以根据需求选择排查方式。

- 自动诊断：日志服务提供Linux版Logtail自动诊断工具，排查步骤请参考*Logtail*自动诊断工具。
- 人工诊断：Logtail诊断工具未检查出问题、或服务器为Windows服务器，请参考本文档逐步排查。



1. 检查是否已安装**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*Linux*和*Windows*，务必按照您日志服务Project所属Region以及网络类型进行安装。

查看Logtail安装状态：

- Linux：

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

如果显示`ilogtail is running`，表示已安装Logtail，例如：

```
[root@****************~]# sudo /etc/init.d/ilogtaild status
ilogtail is running
```

- Windows：

1. 在控制面板中单击管理工具，并单击服务。
2. 查看LogtailDaemon、LogtailWorker两个Windows Service的运行状态。如果正在运行，表示已安装Logtail。

如果Logtail正在运行，请执行下一步检查。

2. 检查**Logtail**安装参数是否正确

安装Logtail时，需要为客户端指定正确的服务端访问入口，即根据日志服务Project所在地域选择*表1*，并根据网络类型选择不同的安装方式。如果安装参数或安装脚本错误，可能会导致Logtail机器无心跳。

Logtail配置文件*`ilogtail_config.json`*中记录了Logtail安装参数及所选的安装方式，该文件的路径为：

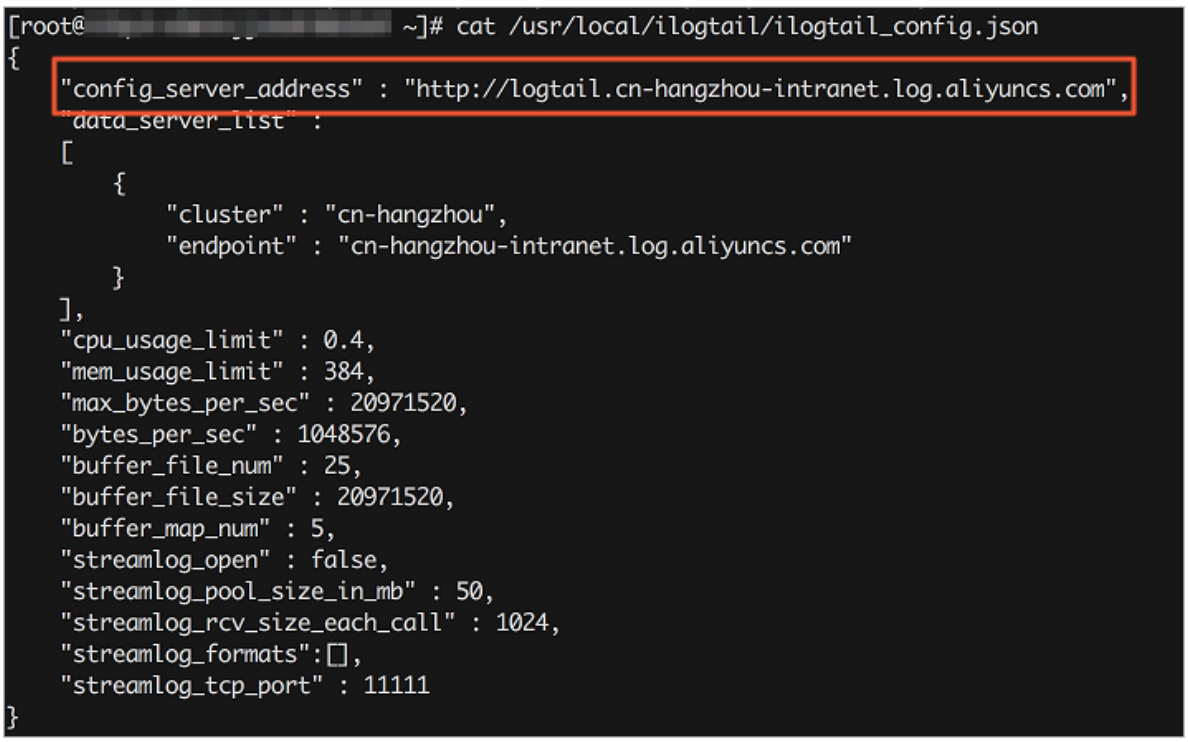
- Linux：*`/usr/local/ilogtail/ilogtail_config.json`*
- Windows x64：*`C:\Program Files (x86)\Alibaba\Logtail\ilogtail_config.json`*
- Windows x32：*`C:\Program Files\Alibaba\Logtail\ilogtail_config.json`*

1. 检查安装参数。

   检查文件*`ilogtail_config.json`*中客户端连接的网络入口所属Region是否与您Project所在Region一致。

   例如以下回显信息表明Logtail安装在华东一（杭州）地域的ECS中。

图 **3-1:** 检查安装参数



2. 检查安装方式。

   Telnet测试*ilogtail_config.json*中配置的域名，检查是否根据服务器所属网络环境选择了正确的安装方式。

   例如，*ilogtail_config.json*中记录Logtail配置的域名为`cn-hangzhou-intranet`，则可以执行`telnet logtail.cn-hangzhou-intranet.log.aliyuncs.com 80`检查连通性，如果已联通，说明安装方式正确。

   例如，查看Linux ECS的网络连接性：

   ```
   [root@********** ~]# telnet logtail.cn-hangzhou-intranet.log.
   aliyuncs.com 80
   Trying 100*0*7*5...
   Connected to logtail.cn-hangzhou-intranet.log.aliyuncs.com.
   Escape character is '^]'.
   ```

   如果telenet失败，说明安装时选择了错误的参数，以至于执行了错误的安装命令。请参考*Linux*和*Windows*选择正确的安装参数。

如果Logtail已正确安装，请执行下一步检查。

3. 检查机器组配置的IP地址是否正确

机器组中配置的IP地址必须和Logtail获取到的服务器地址一致，否则机器组无心跳、或无法采集到日志数据。

Logtail在机器上获取IP的方式：

- 如果没有设置主机名绑定，会取服务器的第一块网卡IP。
- 如果在文件*/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。

排查步骤

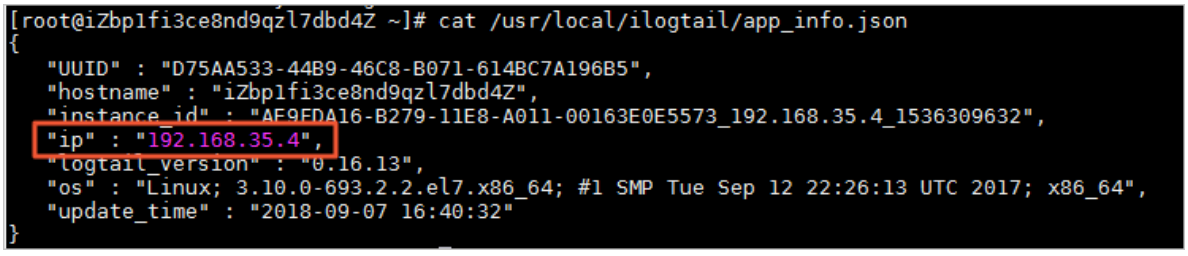
1. 查看Logtail获取的IP地址。

   文件*app_info.json*的ip字段中记录了Logtail获取的IP地址，该文件的路径为：

   - 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*

   > 说明：
   > - 如果app_info.json文件中ip字段为空，Logtail无法工作。此时需为服务器设置IP地址并重启Logtail。
   > - 文件app_info.json仅做记录，修改该文件并不会改变Logtail获取的IP地址。

   图 **3-2:** 查看**Logtail**获取的**IP**地址

2. 查看机器组中配置的地址。

   机器组中配置的IP地址，请在日志服务控制台**机器组列表**页面单击**查看状态**。

图 3-3: 查看机器组

如果服务端机器组内填写的IP与客户端获取的IP不一致，则需要修改。

- 若服务端机器组填写了错误IP，请修改机器组内IP地址并保存，等待1分钟再查看心跳状态。
- 若修改了机器上的网络配置（如修改*/etc/hosts*），请重新启动Logtail以获取新的IP，并根据`app_info.json`文件中的`ip`字段修改机器组内设置的IP地址。

重启**Logtail**的方式：

- Linux：

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

- Windows：在控制面板中单击管理工具 > 服务，找到并重启LogtailWorker。

机器组配置的IP地址的确为Logtail获取的IP地址，请执行下一步检查。

4. 检查非本账号下**ECS**是否已配置**AliUid**

如果您的ECS服务器和日志服务Project不在同一账号下、或服务器为其他云厂商服务器、自建IDC，则需要在服务器上配置*AliUid*，为安装Logtail的机器授权。

检查*/etc/ilogtail/users*目录下是否有账号ID同名文件。

如果没有，请创建同名文件。

- Linux：

```
touch /etc/ilogtail/users/1559122535028493
touch /etc/ilogtail/users/1329232535020452
```

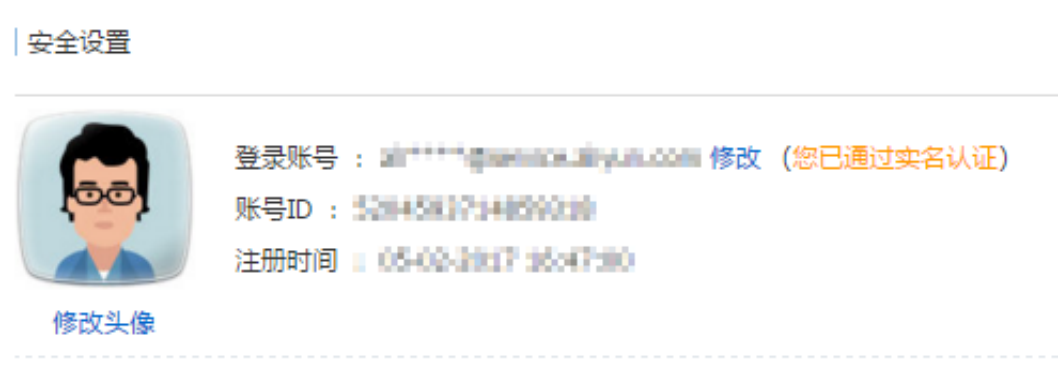
- Windows：

  C:\LogtailData\users\1559122535028493

说明：

账号ID请在阿里云控制台个人信息中查看。

图 3-4: 查看账号ID



如果您的问题仍未解决，请提工单到日志服务。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。

3.4 Logtail 快速诊断工具

当日志采集发生异常时，您可以通过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。

说明：

本工具目前仅支持Linux系统的服务器。

准备工作

1. 下载检测工具脚本。

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

```
                checkingtool.sh
```

> 说明：
>
> 如果无法正常下载，请通过以下备用地址重试。
>
> ```
> wget http://logtail-corp.oss-cn-hangzhou-zmf.aliyuncs.com/linux64/
> checkingtool.sh
>                -O checkingtool.sh
> ```

2. 安装curl工具。

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

运行诊断工具

1. 执行以下命令运行诊断工具：

   ```
   chmod 744 ./checkingtool.sh
   ./checkingtool.sh
   sh checkingtool.sh
   ```

   回显信息：

   ```
   [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 :
   ```

2. 请根据提示输入1或2，脚本会根据您的选择执行不同检查流程。

   其中：

   - 1表示执行机器组心跳检查，机器组心跳失败时请选择此项。
   - 2表示执行日志采集检查，机器组心跳成功，但日志文件没有被采集时，请选择此项。

   选择检查项目后，诊断工具会自动执行对应检查流程。

诊断流程

机器组心跳检查

选择机器组心跳检查流程后会进行下述一系列的检查：

1. 基础环境检查。
   - 是否安装Logtail。
   - 是否运行Logtail。
   - SSL状态是否正常。
   - 与日志服务之间是否有网络联通。

```
[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-
XXXXXXXXXXX
[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
```

若其中检查出现Error信息，请参考提示进行处理。

2. 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ECS、是否由本账号购买。

   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 : 126XXXXXXXXX79 .
           [ OK ]
[Info]:     aliyun user id : 165XXXXXXXXX50 .
           [ OK ]
[Info]:     aliyun user id : 189XXXXXXXXX57 .
           [ OK ]
[Input]:  Is your project owner account ID is the above IDs ? (y/N)
```

3. 检查Region。

   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) :
```

4. 检查IP配置。

   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*IP*地址机器组修改。

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) :
```

日志采集检查

选择日志未采集检查流程后会进行下述一系列的检查：

1. 确认IP配置。

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

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

2. 确认采集配置应用。

   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管理机器组。

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

3. 检查日志文件。

   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

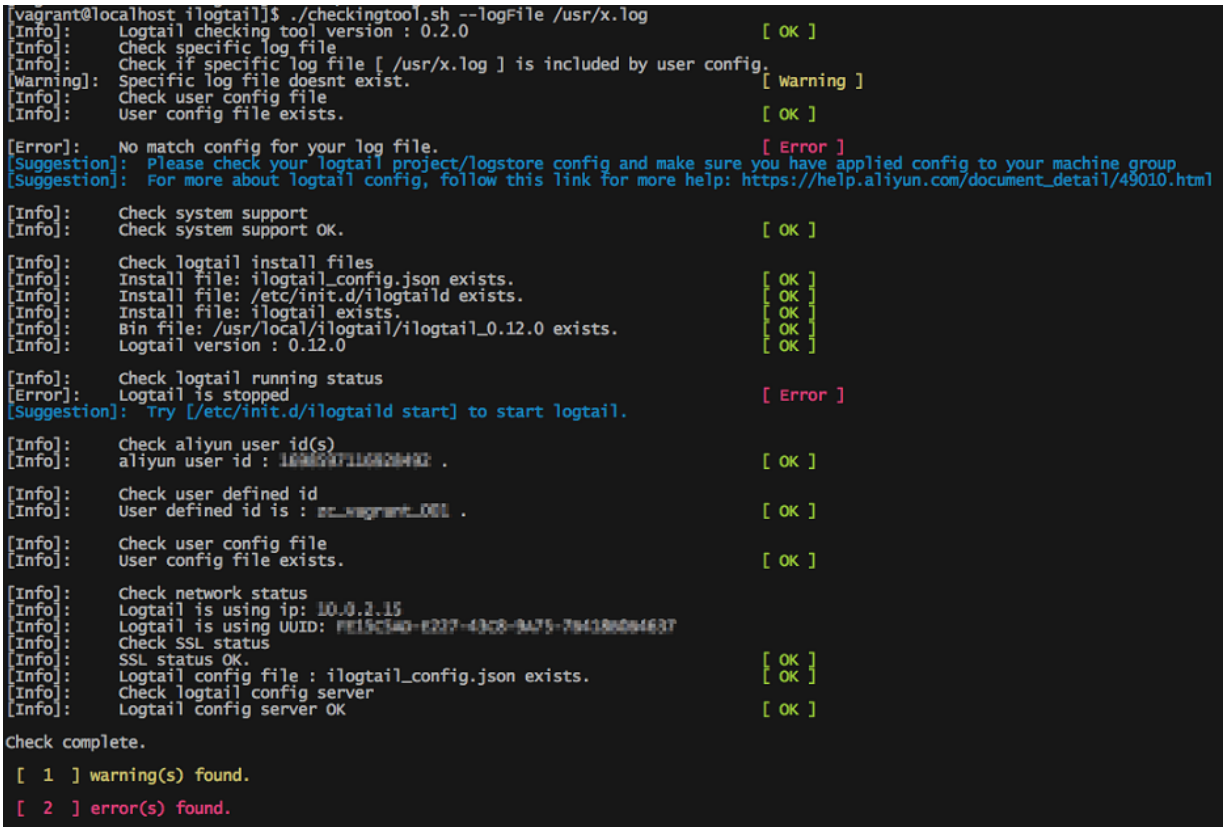
快速检查运行时无需确认，可用于二次封装自定义检查脚本。

> 说明：
> 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：创建*AliUid*标识、自定义标识机器组。

操作步骤

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。

> 说明：
> 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，参见诊断采集错误。



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Logtail。
多个Logtail进程	使用命令`/etc/init.d/ilogtaild stop`关闭Logtail，然后执行命令`/etc/init.d/ilogtaild start`开启。
443端口被禁用	防火墙开放443端口。
无法找到配置服务器	确认是否已正确安装 *Linux Logtail*，若安装错误，重新执行安装命令。
不存在用户配置	确认是否已执行以下操作： 1. 控制台已经创建好Logtail配置。 2. 机器组中包含该服务器。 3. 已经将配置应用到机器组。
没有匹配指定日志文件	确认是否正确配置了Logtail。

常见问题	解决方法
指定日志文件匹配多次	匹配多次时Logtail会随机选择一个配置，建议去重。

检测工具常用参数

常用参数	说明
`--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。

3.5 如何配置正则表达式

在配置Logtail采集文本日志时，如果选择正则模式解析日志，需要您根据自己的日志样例配置正则表达式。

日志服务提供自动生成正则表达式的功能，您可以粘贴日志样例到配置界面，以自动生成正则表达式。

如果自动生成的正则表达式不能完全涵盖您的日志样例，您可以也参考Python的正则样本，手动书写正则表达式。详细信息请参考*Python*日志。

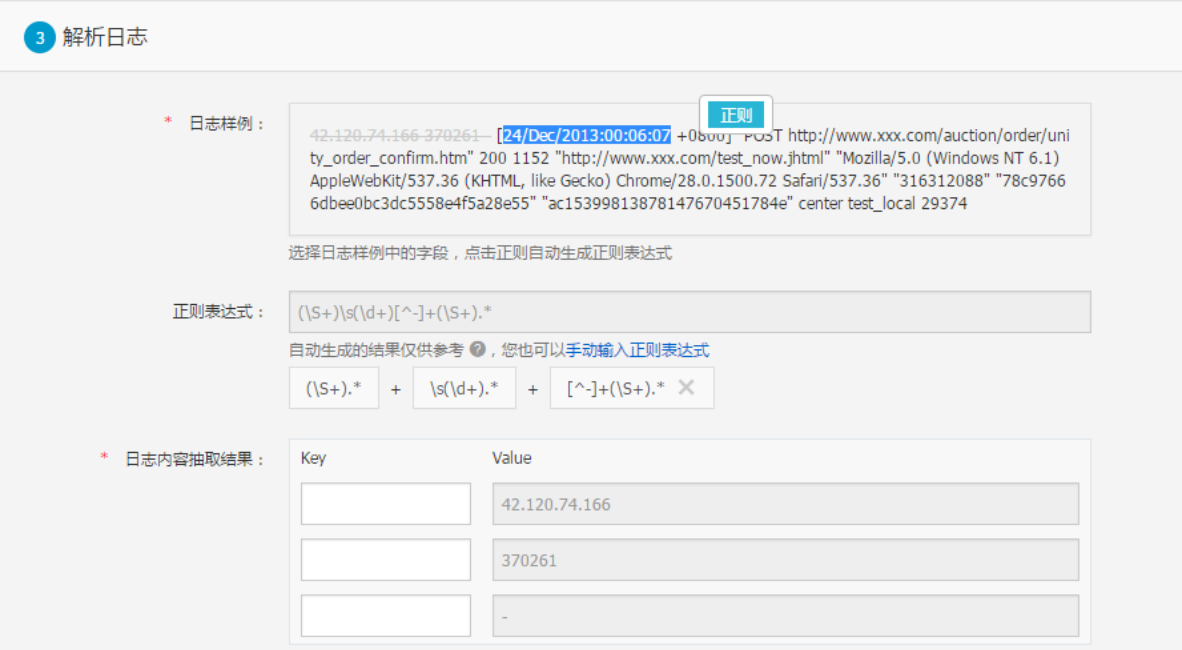
以标准的Nginx访问日志为例：

1. 先配置一部分，其他的部分先用.*来匹配。

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

2. 根据日志样例修改正则表达式。

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



如果您的问题仍未解决，请提工单联系售后支持。

3.6 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。

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

参见成本优势中LogHub部分。

4 日志查询

4.1 日志查询常见问题

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。

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]:&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;bb07263xxd2axx43xx9exxea26e39e
5f&amp;nbsp;POS&amp;nbsp;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。

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。

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。
3. 通过Restful API查询。

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。

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。

4.2 查询不到日志数据

在使用日志服务产品的日志查询功能时，如果查询不到日志数据，请按照以下原因进行排查。

1. 未成功采集日志数据

如果并未成功采集日志数据到日志服务，则无法查询到目标日志。请在预览界面查看是否有日志数据。

如果有日志数据，说明日志数据已成功采集到日志服务中，建议您排查其他原因。

如果没有日志数据，可能是以下原因造成，请进一步排查。

- 日志源没有生产日志数据。

  日志源没有日志产生的情况下，没有日志可以投递到日志服务。请检查您的日志源。

- Logtail无心跳。

  请在机器组状态页面中查看机器是否有心跳。没有心跳请参考*Logtail* 机器无心跳。

- 监控文件没有实时写入。

如果监控文件有实时写入，您可以打开*/usr/local/ilogtail/ilogtail.LOG*查看报错信息。常见错误如下：

- parse delimiter log fail：分割符收集日志出错。
- parse regex log fail：正则收集日志出错。

2. 分词设置错误

查看已设置的分词符，检验根据分词符对日志内容进行分割后，是否刚好得到关键字。例如分割符为默认的`,;=()[]{}?@&<>/:'`那么用户的日志里如果是有abc”defg,hij会被分割成abc”defg和hij两部分，用abc就搜不到这条日志。

同时支持模糊查询，具体查询语法，请参考查询语法。

> 说明：
>
> - 为了节约您的索引费用，日志服务进行了索引优化，配置了键值索引的Key，不进全文索引。例如，日志中有名为message的key，并且配置了键值索引，加了空格做分词（加空格做分词，请把空格加到分词字符串的中间）。”message: this is a test message“可以用 key:value 的格式 message:this 查到，但是直接查this查询不到，因为配置了键值索引的key，不进全文索引。
> - 创建索引或者对索引做任何更改，只对新进的数据有效，旧数据一律无效。
>
> 您可以点开索引属性，检查已设置的分词是否符合要求。

3. 其他原因

如果日志有产生，可以先在查询处修改查询的时间范围。另外由于日志预览的功能数据是实时的，但是查询的功能是有最多1分钟的延迟的，所以用户可以在日志产生后等1分钟再查。

如您的问题仍未解决，请提工单联系我们。

4.3 日志消费与查询区别

日志服务提供日志消费和查询的功能，都属于对日志的读操作，区别在于消费提供收集和分发通道，查询提供日志查询功能。

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

4.4 日志查询分析常见报错

日志查询分析的常见报错如下。

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. 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。

5. 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。

5 日志投递

5.1 日志投递MaxCompute后，如何检查数据完整性

在日志服务数据投递MaxCompute场景下，需要在MaxCompute表分区维度上检查数据完整性，即MaxCompute表中某个分区中数据是否已经完整。

使用保留字段__partition_time__作为表分区列，如何判断分区数据是否已完整

`__partition_time__`由日志的time字段计算得到，由日志真实时间按照时间格式字符串向下取整得出。其中，日志真实时间既不是投递数据的时间，也不是日志写入服务端时间。

举例：日志时间为2017-05-19 10:43:00，分区字段格式字符串配置为`yyyy_MM_dd_HH_mm`，每1小时投递一次。那么：无论该日志是什么时刻写入服务端，这条日志会存入MaxCompute的`2017_05_19_10_00`分区，计算细节参考*MaxCompute*投递字段说明。

如果不考虑写入了历史数据等问题，在日志实时写入的情况下，有以下两种方法判断分区数据是否已完整：

- 通过控制台或API/SDK判断（推荐）

  使用*API*、*SDK*或者控制台获取指定Project/Logstore投递任务列表。例如API返回任务列表如下。控制台会对该返回结果进行可视化展示。

```
{
  "count" : 10,
  "total" : 20,
  "statistics" : {
      "running" : 0,
      "success" : 20,
      "fail" : 0
  }
  "tasks" : [
      ...
      {
          "id" : "abcdefghijk",
          "taskStatus" : "success",
          "taskMessage" : "",
          "taskCreateTime" : 1448925013,
          "taskLastDataReceiveTime" : 1448915013,
          "taskFinishTime" : 1448926013
      },
      {
          "id" : "xfegeagege",
          "taskStatus" : "success",
          "taskMessage" : "",
          "taskCreateTime" : 1448926813,
          "taskLastDataReceiveTime" : 1448930000,
          "taskFinishTime" : 1448936910
```

```
        }
    ]
}
```

`taskLastDataReceiveTime`表示该批任务中最后一条日志到达服务端时的机器系统时间，对应控制台的接收日志数据时间，根据该参数判断时间为T以前的数据是否已经全部投递到MaxCompute表。

- 如果`taskLastDataReceiveTime < T + 300s`（300秒是为了容忍数据发送服务端发生错误重试）以前的每个投递任务状态都是success，说明T时刻的数据都已经入库。
- 如果任务列表中有ready/running状态任务，说明数据还不完整，需要等待任务执行结束。
- 如果任务列表中有failed状态任务，请查看原因并在解决后重试任务。您可以尝试修改投递配置，以解决投递问题。

- 通过MaxCompute分区粗略估计

  比如在MaxCompute中以半小时做一次分区，投递任务为每30分钟一次，当表中包含以下分区：

  ```
  2017_05_19_10_00
  2017_05_19_10_30
  ```

  当发现分区2017_05_19_11_00出现时，说明11:00之前分区数据已经完整。

  该方法不依赖API，判断方式简单但结果并不精确，仅用作粗略估计。

使用自定义日志字段作为表分区列，如何判断分区数据已完整

比如日志中有一个字段date，取值：20170518，20170519，在配置投递规则时将date列映射到表分区列。

这种情况下，需要考虑date字段与写入服务端时间差，结合使用保留字段方法，根据接收日志数据时间判断。

投递成功，但MaxCompute表数据有缺失，应如何解决

MaxCompute投递任务状态成功，但表中数据缺失，一般有以下原因：

- 表分区列映射的日志服务字段的名称不存在。此时投递过去的列值为null，而MaxCompute表不允许分区列值为null。
- 表分区列映射的日志服务字段的值包含/或其他特殊符号。MaxCompute将这些字符作为保留字，不允许在分区列中出现。

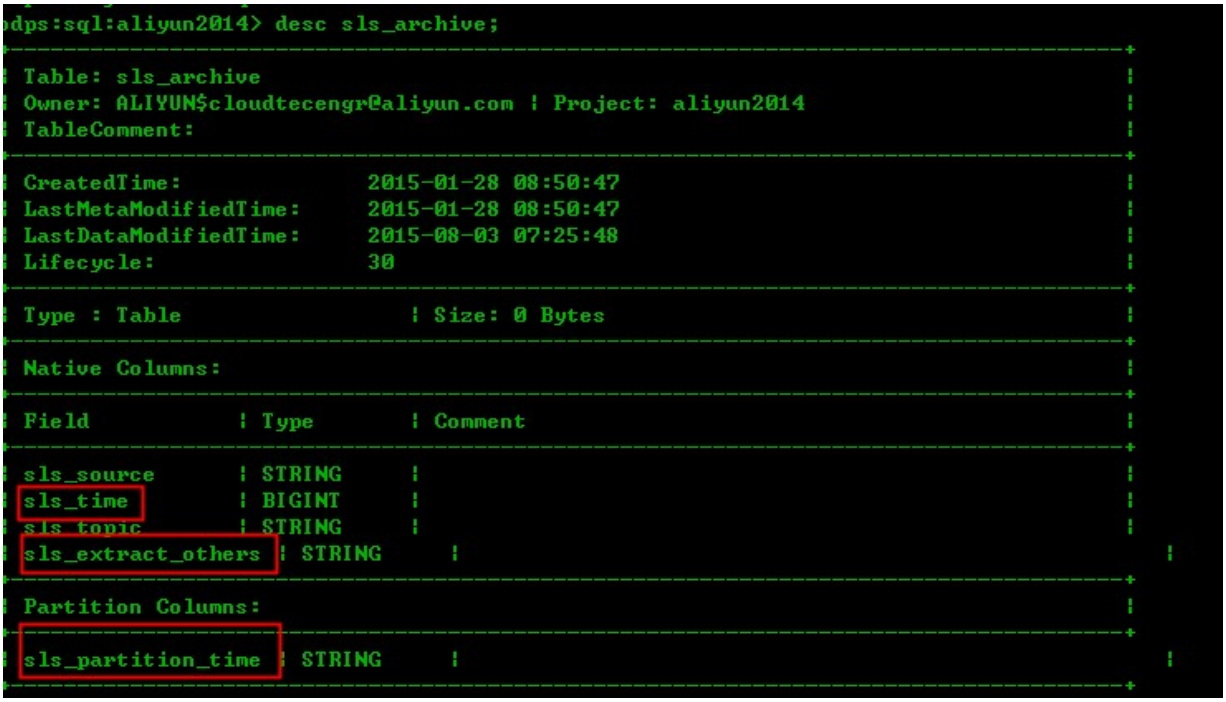
遇到这些情况时，投递策略为忽略异常的日志行，并继续任务，在该次任务中其它分区正确的日志行可以成功同步。

因此，在配置字段映射不当的情况下，可能出现任务成功但是表中缺少数据的情况，请修改分区列配置。建议使用保留字段__partition_time__作为分区。

更多细节请参考MaxCompute投递相关限制说明。

5.2 投递MaxCompute时的参数时间

日志服务投递日志至MaxCompute中有多个时间参数。



其中涉及到的时间参数及其含义如下：

时间参数	含义
sls_time	sls的服务器收到日志的时间
sls_extract_others	是用户真实的时间
sls-partition_time	MaxCompute归档的时间

说明：

sls_time一般都是略晚于sls_extract_others的时间。

如您的问题仍未解决，请联系售后技术支持。

6 停用日志服务

若您不再需要使用日志服务，可以通过删除所有数据来停用日志服务。

背景信息

若您不再需要使用日志服务，删除所有Project和Logstore以清除所有日志数据。

> 说明：
>
> - 日志数据清除的当天仍会产生存储等费用，次日会收到账单。清除数据第三天开始不会收到日志服务的账单，因为从清除数据第二天开始没有产生任何计费项。
> - 当您的Project被删除后，其中的所有日志数据及配置信息都会永久释放，不可恢复。所以，在删除Project 前请慎重确认，避免数据丢失。

操作步骤

1. 登录日志服务控制台。
2. 在**Project**列表页面，找到需要删除的项目。
3. 单击项目名称右侧的**删除**。
4. 在弹出的对话框中，单击**获取验证码**。
5. 输入您收到的校验码并单击**确认** 。