



日志服务 数据投递

文档版本: 20220706



# 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。

- 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。
- 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。
- 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	
⚠ 危险	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。	☆ 危险 重置操作将丢失用户配置数据。	
▲ 警告	该类警示信息可能会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 时间约十分钟。	
〔〕 注意	用于警示信息、补充说明等,是用户必须 了解的内容。	▶ 注意 权重设置为0,该服务器不会再接受新 请求。	
⑦ 说明	用于补充说明、最佳实践、窍门等,不是 用户必须了解的内容。	⑦ 说明 您也可以通过按Ctrl+A选中全部文件。	
>	多级菜单递进。	单击设置> 网络> 设置网络类型。	
粗体	表示按键、菜单、页面名称等UI元素。	在 <b>结果确认</b> 页面,单击 <b>确定</b> 。	
Courier字体	命令或代码。	执行    cd /d C:/window    命令,进入 Windows系统文件夹。	
斜体	表示参数、变量。	bae log listinstanceid	
[] 或者 [alb]	表示可选项,至多选择一个。	ipconfig [-all -t]	
{} 或者 {alb}	表示必选项,至多选择一个。	switch {act ive st and}	

# 目录

1.数据投递概述	06
2.投递日志到OSS(旧版)	08
2.1. 将日志服务数据投递到OSS	08
2.2. JSON格式	12
2.3. CSV格式	12
2.4. Parquet格式	14
2.5. 解压Snappy压缩文件	16
2.6. RAM授权	18
3.投递日志到OSS(新版)	22
3.1. OSS投递新旧版本对比	22
3.2. 准备权限	22
3.2.1. 授权概述	22
3.2.2. 授予RAM用户操作OSS投递的权限	23
3.2.3. 通过默认角色访问数据	25
3.2.4. 通过自定义角色访问数据(同账号场景)	26
3.2.5. 通过自定义角色访问数据(跨账号场景)	28
3.3. 创建OSS投递作业(新版)	31
3.4. JSON格式	34
3.5. CSV格式	35
3.6. Parquet格式	37
3.7. ORC格式	40
3.8. 管理OSS投递作业(新版)	43
3.9. 为OSS投递作业(新版)设置告警	45
3.10. 时区列表	51
4.投递日志到MaxCompute(旧版)	54
5.投递日志到MaxCompute(新版)	62

	5.1. MaxCompute投递新旧版对比	62
	5.2. 稳定性说明与使用限制	62
	5.3. 准备权限	65
	5.3.1. 授权概述	65
	5.3.2. 授予RAM用户操作MaxCompute投递的权限	66
	5.3.3. 通过默认角色读取Logstore数据	67
	5.3.4. 通过自定义角色读取Logstore数据	68
	5.3.5. 通过默认角色写数据到MaxCompute	70
	5.3.6. 通过自定义角色写数据到MaxCompute	73
	5.3.7. 通过RAM用户写数据到MaxCompute	77
	5.4. 创建MaxCompute投递作业(新版)	78
	5.5. 为MaxCompute投递作业(新版)设置告警	83
	5.6. 管理MaxCompute投递作业(新版)	89
6.	将日志服务数据投递到AnalyticDB MySQL	91
7.	将日志服务数据投递到TSDB	93
8.	管理日志投递任务	97
9.	数据投递仪表盘	99
10	).配置监控与告警	101
11	l.最佳实践	108
	11.1. 投递-对接数据仓库	108
12	2.常见问题	110
	12.1. 日志投递MaxCompute后,如何检查数据完整性	110
	12.2. 投递日志到OSS失败	111
	12.3. MaxCompute投递作业(新版)常见报错与问题	112

# 1.数据投递概述

日志服务提供投递功能,支持通过控制台将数据实时投递至OSS、MaxCompute等阿里云产品。本文介绍数据投递功能的概念、功能优势、应用场景、计费规则、投递目标等信息。

## 数据投递

数据投递是指将日志服务采集的数据通过控制台投递至其他阿里云产品,便于您存储数据或联合其它系统消费数据。启用 数据投递后,日志服务将定时将采集到的数据投递至对应的云产品。



# 应用场景

数据投递适用于数据存储、离线分析等场景。数据投递的实时性较弱,通常为5分钟~30分钟。数据存储时间依赖于存储系 统。

#### 功能优势

数据投递具有以下优势:

● 操作简易

您只需要在控制台上完成简单的配置,即可将日志服务的数据投递到OSS等云产品。

● 数据集中

日志服务已完成不同机器上的数据集中化,您只需将采集到的数据投递到OSS等云产品。

• 分类管理

您可以利用日志服务的数据分类管理功能,将不同项目、不同类型的数据投递到不同的OSS Bucket、MaxCompute表等 云产品目标中。

#### 计费规则

数据投递涉及多个计费项,包括读写流量、数据投递费用、请求费用等。更多信息,请参见计费项。

#### 投递目标

日志服务支持的数据投递目标如下表所示。

目标	说明
	日志服务支持将数据投递至对象存储OSS。 更多信息,请参见投 递日志到OSS(旧版)、投递日志到OSS(新版)。
对象存储OSS	<ul> <li>⑦ 说明</li> <li>推荐您使用E-MapReduce转换OSS数据格式。更多信息,请参见什么是E-MapReduce。</li> <li>投递完成后,支持通过云原生数据湖分析DLA分析数据。更多信息,请参见什么是云原生数据湖分析。</li> </ul>

目标	说明	
MaxCompute	日志服务支持通过以下方式将数据投递至MaxCompute。 • 日志服务 通过日志服务投递至MaxCompute。具体操作,请参见投递日 志到MaxCompute(旧版)、投递日志到MaxCompute(新 版)。 • DataWorks 通过DataWorks数据集成投递至MaxCompute。具体操作, 请参见通过数据集成投递日志服务数据。	
云原生数据仓库AnalyticDB MySQL版	日志服务支持将数据投递至云原生数据仓库AnalyticDB MySQL 版。具体操作,请参见 <del>将日志服务数据投递到AnalyticDB</del> MySQL。	
时序数据库TSDB	日志服务支持将数据投递至时序数据库TSDB。具体操作,请参 见 <mark>将日志服务数据投递到TSDB</mark> 。	
表格存储TableStore	日志服务支持通过表格存储的传送服务将数据拉取至表格存储 TableStore。具体操作,请参见 <mark>通过传送服务投递日志服务数</mark> 据。	

# 2.投递日志到OSS(旧版) 2.1. 将日志服务数据投递到OSS

日志服务采集到日志后,支持将日志投递至OSS进行存储与分析。本文介绍将日志投递至OSS的操作步骤。

#### 前提条件

- 已创建Project和Logstore。更多信息,请参见创建Project和Logstore。
- 已采集到日志。更多信息,请参见数据采集。
- 已在日志服务Project所在的地域创建OSS Bucket。更多信息,请参见创建存储空间。
- 已完成云资源访问授权。

如果您要跨阿里云账号或使用RAM用户配置投递规则,请参见RAM授权。

×
<b>~</b>

#### 背景信息

日志服务支持将Logstore中的数据自动归档到OSS,以发挥更多的日志价值。

- OSS支持自由设置生命周期,可以长期存储日志。
- 您可以通过数据处理平台(例如E-MapReduce和DLA)或自建程序消费OSS数据。

### 支持的地域说明

日志服务投递数据到OSS为同地域投递,即日志服务Project所在的地域和OSS Bucket所在地域相同。

↓ 注意 如果Project属于杭州金融云,则只支持投递到杭州金融云公网地域(oss-cn-hzfinance),不支持华东1
 金融云地域(oss-cn-hzjbp)。

# 投递数据

⑦ 说明 目前华东1(杭州)、华北2(北京)、西南1(成都)、华南1(深圳)地域支持OSS投递(新版)。更多 信息,请参见创建OSS投递作业(新版)。此处以华东2(上海)地域的界面为例。

- 1. 登录日志服务控制台。
- 2. 在Project列表区域,单击目标Project。
- 3. 在日志存储 > 日志库页签中, 单击目标Logstore左侧的>, 选择数据处理 > 导出 > OSS(对象存储)。
- 4. 单击开启投递。
- 5. 在投递提示对话框中, 单击直接投递。
- 在OSS投递功能面板,配置如下参数,然后单击确定。 重要参数配置说明如下所示。

### ↓ 注意

- 开启日志投递后,日志服务并发执行投递任务,每个Shard都会根据投递大小、投递时间决定任务生成的频率,当任一条件满足时,即会创建投递任务。
- 创建投递任务后,您可以通过投递任务的状态和投递到OSS的数据确认投递规则是否符合预期结果。

参数	说明
OSS投递名称	投递规则的名称。
OSS Bucket	OSS Bucket名称。 〇) 注意 必须是已存在的OSS Bucket名称,且该OSS Bucket与日志服务Project位 于相同地域。
文件投递目录	OSS Bucket中的目录。目录名不能以正斜线(/)或者反斜线(\)开头。 创建OSS投递任务后,Logstore中的数据将投递到目标OSS Bucket的此目录中。
分区格式	按照投递任务的创建时间动态生成OSS Bucket目录,不能以正斜线(/)开头,默认值 为%Y/%m/%d/%H/%M,相关示例请参见 <mark>分区格式</mark> ,参数详情请参见 <mark>strptime API</mark> 。
写OSS RAM角色	授予OSS投递任务将数据写入到OSS Bucket的权限。 • <b>默认角色</b> : 授权OSS投递任务使用阿里云系统角色AliyunLogDefaultRole将数据写入到 OSS Bucket中。即输入AliyunLogDefaultRole的ARN。 • 自定义角色: 授权OSS投递任务使用自定义角色将数据写入到OSS Bucket中。您需先授 予自定义角色将数据写入到OSS Bucket的权限,然后在写OSS RAM角色中输入您自定义 角色的ARN。 • 如果Logstore和OSS Bucket属于同一阿里云账号,请参见步骤二:授予RAM角色写 OSS Bucket的权限。 • 如果Logstore和OSS Bucket属于不同的阿里云账号,请参见步骤二:授予账号B下的 RAM角色与GOSS Bucket的权限。
投递大小	每个Shard的投递大小。通过该值控制OSS Object大小(以未压缩计算),取值范围为 5~256,单位为MB。 当每个Shard投递数据大小达到此处设置的大小时,会自动创建一个新的投递任务。
存储格式	日志数据投递到OSS后,支持存储为不同的文件格式,包括JSON格式、CSV格式和Parquet 格式。更多信息,请参见 <mark>CSV格式</mark> 。
是否压缩	OSS数据存储的压缩方式。 <ul> <li>不压缩:表示原始数据不压缩。</li> <li>压缩(snappy):表示使用snappy算法压缩数据,可减少OSS Bucket存储空间。</li> </ul>
投递时间	每个Shard的投递周期。取值范围为300~900,默认值为300,单位为秒。 当每个Shard投递周期达到此处设置的大小时,会自动创建一个新的投递任务。

# 查看OSS数据

将日志投递到OSS成功后,您可以通过OSS控制台、API、SDK或其它方式访问OSS数据。更多信息,请参见文件管理。 OSS Object地址格式如下所示: oss://OSS-BUCKET/OSS-PREFIX/PARTITION-FORMAT\_RANDOM-ID

OSS-BUCKET为OSS Bucket名称,OSS-PREFIX为目录前缀,PARTITION-FORMAT为分区格式(由投递任务的创建时间通 过strptime API计算得到),RANDOM-ID是投递任务的唯一标识。

⑦ 说明 OSS Bucket 目录是按照投递任务创建时间设置的。例如: 2016-06-23 00:00:00创建投递任务,投递的是
 2016-06-22 23:55后写入日志服务的数据,假设5分钟投递一次数据到OSS。如果您要分析2016-06-22全天日志,除
 了查看2016/06/22目录下的全部object以外,还需要检查2016/06/23/00/目录下前十分钟的Object是否包含2016-06-22的日志。

## 分区格式

一个投递任务对应一个OSS Bucket 目录,目录格式为*oss://OSS-BUCKET/OSS-PREFIX/PARTITION-FORMAT\_RANDOM-ID*。PARTITION-FORMAT是根据投递任务的创建时间格式化而得到的,以创建时间为2017/01/20 19:50:43的投递任务为例,介绍分区格式,如下表所示。

OSS Bucket	OSS Prefix	分区格式	OSS文件路径
test-bucket	test-table	%Y/%m/%d/%H/%M	oss://test-bucket/test- table/2017/01/20/19/50_14849 13043351525351_2850008
test-bucket	log_ship_oss_example	year=%Y/mon=%m/day=%d/lo g_%H%M%s	<i>oss://test- bucket/log_ship_oss_example/y ear=2017/mon=01/day=20/log_ 195043_1484913043351525351_ 2850008.parquet</i>
test-bucket	log_ship_oss_example	ds=%Y%m%d/%H	oss://test- bucket/log_ship_oss_example/ ds=20170120/19_148491304335 1525351_2850008.snappy
			oss://test- bucket/log_ship_oss_example/ 20170120/_14849130433515253 51_2850008
test-bucket	log_ship_oss_example	%Y%m%d/	⑦ 说明 此格式会导致 Hive等平台无法解析对应的 OSS内容,建议您不要使用该 格式。
			occ://toct
test-bucket	log_ship_oss_example	%Y%m%d%H	<i>uss://test-</i> <i>bucket/log_ship_oss_example/</i> <i>2017012019_1484913043351525</i> <i>351_2850008</i>

使用Hive、MaxCompute等大数据平台或阿里云DLA产品分析OSS数据时,如果您希望使用Partition信息,可将 PARTITION-FORMAT设置为key=value格式。例如:*oss://test-*

*bucket/log\_ship\_oss\_example/year=2017/mon=01/day=20/log\_195043\_1484913043351525351\_2850008.parquet*, 设置为三层分区列,分别为: year、mon、day。

### 相关操作

创建投递任务后,您可以在**OSS投递管理**页面,执行修改投递规则、关闭投递、查看投递任务状态及错误信息、重试投递 任务等操作。 ● 修改投递规则

单击投递配置,修改投递规则,参数详情请参见本文中的投递数据。

- 关闭投递
   单击关闭投递,即可关闭投递。
- 查看投递任务状态及错误信息

日志服务支持查看过去两天内的所有日志投递任务及其投递状态。

○ 任务状态

状态	说明
成功	投递任务正常运行。
进行中	投递任务进行中,请稍后查看是否投递成功。
失败	因外部原因而无法重试的错误导致投递任务失败,请根据错误信息进行排查并重试。

#### ○ 错误信息

如果投递任务出现错误,控制台上会显示相应的错误信息。

错误信息	错误原因	处理方法
UnAuthorized	没有权限。	请确认以下配置: OSS Bucket拥有者是否已创建 AliyunLogDefaultRole角色。 角色描述中配置的阿里云账号ID是否正确。 AliyunLogDefaultRole角色是否被授予OSS Bucket写权限。 RAM角色标识是否配置正确。
ConfigNotExist	配置不存在。	一般是由于关闭投递导致的。请在重新开启投递 并配置投递规则后,通过重试解决。
InvalidOssBucket	OSS Bucket不存在。	请确认以下配置: ■ OSS Bucket所在地域与日志服务Project所在地 域是否相同。 ■ Bucket名称是否配置正确。
InternalServerError	日志服务内部错误。	通过重试解决。

#### o 重试任务

日志服务会按照策略默认为您重试,您也可以手动重试。日志服务默认重试最近两天之内所有失败的任务,重试等待的最小间隔是15分钟。当任务执行失败时,第一次失败需要等待15分钟再进行重试,第二次失败需要等待30分钟再进行重试,以此类推。

如果您需要立即重试失败任务,请单击**重试全部失败任务**、目标任务右侧的**重试**或通过API、SDK指定任务进行重 试。

## 常见问题

如何获取RAM角色标识?

- 1. 登录RAM 控制台。
- 2. 在左侧导航栏中,单击RAM 角色管理。
- 3. 在RAM角色列表中,单击AliyunLogDefault Role。

#### 4. 在基本信息区域,获取ARN。

← AliyunLogDefaultRole			
基本信息			
RAM 角色名称	AliyunLogDefaultRole	创建时间	2020年6月1日 10:39:49
备注	日志服务默认使用此角色来访问您在其他云产品中的资源	ARN	acs:ram r 🗋 复制
最大会话时间	3600秒 編輯		

# 2.2. JSON格式

日志服务将日志投递到OSS后,支持存储为不同的文件格式,本文介绍JSON格式。

在配置投递规则时,如果选择存储格式为json,对应生成的OSS文件地址如下所示。

压缩类型	文件后缀	OSS文件地址示例	说明
不压缩	无	oss://oss-shipper- shenzhen/ecs_test/2016/01/26/20/54 _1453812893059571256_937	<pre>未压缩的JSON文件可下载到本地,以文本 方式打开查看。文件样例如下所示。 {"time":1453809242,"top ic":"","_source":10.170 .***.***","ip":"10.200.**.*** ","time":"26/Jan/2016:19:54:0 2 +0800","url":"POST</pre>
snappy	.snappy	oss://oss-shipper- shenzhen/ecs_test/2016/01/26/20/54 _1453812893059571256_937.snappy	如果使用了snappy压缩,解压缩的详细 说明请参见解 <mark>压Snappy压缩文件</mark> 。

# 2.3. CSV格式

日志服务将日志投递到OSS后,支持存储为不同的文件格式。本文介绍CSV格式。

## 参数配置

在配置投递规则时,如果选择存储格式为csv,对应的参数配置如下所示。

* CSV字段:	Key名称	操作
	_source_	删除
	time	删除
	+ 如何使用oss	shipper生成csv文件?
* 分隔符:	逗号	$\checkmark$
* 转义符:	双引号	$\checkmark$
无效字段内容:	指定字段名称不存在时投递内容,默认为空	
* 投递字段名称:	表示是否将字段名称写入CSV文件,默认为不写入	

#### 参数说明如下所示。更多信息,请参见CSV标准和postgresql CSV说明。

参数	说明
	您可以在日志服务的 <b>原始日志</b> 页签中查看日志字段的Key,将您需要投递到OSS的字段名 (Key)有序填入。 除了日志内容的Key和Value外,日志服务还提供保留字 段time、topic、source,保留字段更多信息,请参见 <mark>保留字段</mark> 。
CSV字段	<ul> <li>⑦ 说明</li> <li>● 同一个Key在CSV字段中只能配置一次,不支持多次使用。</li> <li>● 如果您的日志中存在同名字段(例如都为request_time),则日志服务会将其中一个字段名显示为request_time_0,底层存储的字段名仍为request_time。因此您在投递时,只能使用原始字段名request_time。</li> <li>存在同名字段时,系统只随机投递其中一个字段的值。请尽量避免日志中使用同名字段。</li> </ul>
分隔符	支持半角逗号(,)、竖线( )、空格和制表符,用于分割不同字段。
转义符	当字段内出现分隔符时,需要使用转义符包裹该字段,避免读数据时造成字段错误分割。
无效字段内容	当CSV字段中指定的Key不存在时,投递此参数配置的内容。
投递字段名称	打开 <b>投递字段名称</b> 开关后,将字段名称写入到CSV文件。

# OSS文件地址

#### 投递到OSS后, OSS文件地址样例如下所示。

压缩类型	文件后缀	OSS文件地址举例	说明
无压缩	.CSV	oss://oss-shipper- shenzhen/ecs_test/2016/01/26/20 /54_1453812893059571256_937.csv	未压缩的CSV文件可下载到本地,以文 本方式打开查看。

压缩类型	文件后缀	OSS文件地址举例	说明
Snappy	.snappy.csv	oss://oss-shipper- shenzhen/ecs_test/2016/01/26/20 /54_1453812893059571256_937.sn appy.csv	Snappy压缩文件的打开方式,请参 见 <mark>打开Snappy压缩文件</mark> 。

# 2.4. Parquet格式

日志服务将日志投递到OSS后,支持存储为不同的文件格式,本文介绍Parquet格式。

# 参数配置

在配置投递规则时,如果选择存储格式为Parquet,对应的参数配置如下所示。



#### 相关参数说明如下表所示。

参数	说明
Key名称	<ul> <li>您可以在日志服务的<b>原始日志</b>页签中查看日志字段的Key,将您需要投递到OSS的字段 名(Key)有序填入,在投递时将按照此顺序组织Parquet数据,并使用该key作为 Parquet数据列名。除了日志内容的Key外,日志服务还提供保留字 段_time_、_topic_、_source_,保留字段详情请参见保留字段。如果遇到如下 两种情况时,Parquet数据列的值为null。</li> <li>此处配置的key在日志服务的日志中不存在。</li> <li>字段由string类型转换非string类型(如double、int64等)失败。</li> <li>⑦ 说明</li> <li>同一个Key在Parquet字段中只能配置一次,不支持多次使用。</li> <li>如果您的日志中存在同名字段(例如都为request_time),则日志服务会 将其中一个字段名显示为request_time_0,底层存储的字段名仍 为request_time。因此您在投递时,只能使用原始字段名request_time。 存在同名字段时,系统只随机投递其中一个字段的值。请尽量避免日志中 使用同名字段。</li> </ul>
类型	Parquet存储支持6种类型:string、boolean、int32、int64、float、double。 日志投递过程中,会将日志服务中的日志字段由string类型转换为Parquet目标类型。如 果转换到非string类型失败,则该列数据为null。

### OSS文件地址

投递到OSS后,OSS文件地址示例如下表所示。

压缩类型	文件后缀	OSS文件地址示例	说明
不压缩	.parquet	oss://oss-shipper- shenzhen/ecs_test/2016/01/26/20 /54_1453812893059571256_937.par quet	下载到本地,使用 <mark>parquet-tools</mark> 工具 打开。
压缩(snappy)	.snappy.parquet	oss://oss-shipper- shenzhen/ecs_test/2016/01/26/20 /54_1453812893059571256_937.sn appy.parquet	下载到本地,使用 <mark>parquet-tools</mark> 工具 打开。

## 数据消费

- 通过E-MapReduce、Spark、Hive消费数据,详情请参见社区文档。
- 通过单机校验工具消费数据

开源社区提供的parquet-tools可以用来文件级别验证Parquet格式、查看schema、读取数据内容。您可以自行编译该 工具或者单击parquet-tools-1.6.0rc3-SNAPSHOT,下载日志服务提供的版本。

。 查看Parquet文件schema

```
$ java -jar parquet-tools-1.6.0rc3-SNAPSHOT.jar schema -d 00_1490803532136470439_124353.snappy.parq
uet | head -n 30
message schema {
 optional int32 __time__;
 optional binary ip;
 optional binary __source_;
 optional binary method;
 optional binary __topic__;
 optional double seq;
 optional int64 status;
 optional binary time;
 optional binary url;
 optional boolean ua;
}
creator: parquet-cpp version 1.0.0
file schema: schema
ip: OPTIONAL BINARY R:0 D:1
```

• • • • • • • •

。 查看Parquet 文件全部内容

```
$ java -jar parquet-tools-1.6.0rc3-SNAPSHOT.jar head -n 2 00 1490803532136470439 124353.snappy.parq
uet
time = 1490803230
ip = 10.200.98.220
_____source___ = *.*.*.*
method = POST
topic =
seq = 1667821.0
status = 200
time = 30/Mar/2017:00:00:30 +0800
time = 1490803230
ip = 10.200.98.220
__source__ = *.*.*.*
method = POST
__topic__ =
seq = 1667822.0
status = 200
time = 30/Mar/2017:00:00:30 +0800
```

更多用法请执行java - jar parquet-tools-1.6.0rc3-SNAPSHOT.jar - h命令查看帮助。

# 2.5. 解压Snappy压缩文件

日志服务将日志投递到OSS时,支持通过Snappy压缩OSS文件。投递成功后,您可以通过C++ Lib、Java Lib、Python Lib、Linux工具等方式解压OSS文件。

#### 使用C++ Lib解压

- 1. 下载Snappy Lib。
- 2. 使用 Snappy.Uncompress 解压。

### 使用Java Lib解压

1. 选择以下任意一种方式下载Java Lib。

② 说明 1.1.2.1版本存在Bug,可能无法解压部分压缩文件。1.1.2.6及以上版本已修复该问题,建议您使用 1.1.2.6及以上版本的Java Lib。关于该Bug的更多信息,请参见bad handling of the MAGIC HEADER。

○ 手动方式

xerial snappy-java

○ Maven方式

```
<dependency>
<groupId>org.xerial.snappy</groupId>
<artifactId>snappy-java</artifactId>
<version>1.0.4.1</version>
<type>jar</type>
<scope>compile</scope>
</dependency>
```

#### 2. 使用以下任意一种方法解压。

- ? 说明 不支持使用SnappyFramedInputStream。
- Snappy.Uncompress

#### 示例代码如下:

```
String fileName = "C:\\我的下载\\36_1474212963188600684_4451886.snappy";
RandomAccessFile randomFile = new RandomAccessFile(fileName, "r");
int fileLength = (int) randomFile.length();
randomFile.seek(0);
byte[] bytes = new byte[fileLength];
int byteread = randomFile.read(bytes);
System.out.println(fileLength);
System.out.println(byteread);
byte[] uncompressed = Snappy.uncompress(bytes);
String result = new String(uncompressed, "UTF-8");
System.out.println(result);
```

• Snappy.SnappyInputStream

#### 示例代码如下:

```
String fileName = "C:\\我的下载\\36_1474212963188600684_4451886.snappy";
SnappyInputStream sis = new SnappyInputStream(new FileInputStream(fileName));
byte[] buffer = new byte[4096];
int len = 0;
while ((len = sis.read(buffer)) != -1) {
    System.out.println(new String(buffer, 0, len));
}
```

## 使用Python Lib解压

- 1. 下载Python Lib。
- 2. 使用 snappy.uncompress 解压。
  - Python 2示例代码如下:

```
import snappy
compressed = open('/tmp/temp.snappy').read()
snappy.uncompress(compressed)
```

Python 3示例代码如下:

```
import snappy
compressed = open('/tmp/temp.snappy','rb').read()
print(snappy.uncompress(compressed).decode(encoding='utf-8',errors="ignore"))
```

#### ? 说明

- 如果您在Windows系统下,使用snappy.uncompress解压出现报错,例如 UnicodeDecodeError: 'gbk'
   codec can't decode byte 0xff in position 0: illegal multibyte sequence ,可能是Windows文件
   系统以BOM开头导致的。建议您使用类Unix系统或者在open函数的encoding中设置正确的文件编码。
- 不支持通过以下命令行工具解压snappy压缩文件,命令行模式下仅支持hadoop模式 (hadoop\_stream\_decompress)与流模式(stream\_decompress)。

python -m snappy -d compressed\_file.snappy uncompressed\_file

#### 使用Linux工具解压

针对Linux环境,日志服务提供可以解压Snappy文件的工具。

#### 1. 下载snappy\_tool。

2. 执行以下命令解压。

请根据实际情况替换03\_1453457006548078722\_44148.snappy和03\_1453457006548078722\_44148。

./snappy\_tool 03\_1453457006548078722\_44148.snappy 03\_1453457006548078722\_44148

#### 返回结果示例:

compressed.size: 2217186
snappy::Uncompress return: 1
uncompressed.size: 25223660

### 使用开源PHP工具解压

您可以使用开源的PHP工具解压Snappy文件。

1. 下载php-ext-snappy源码。

您也可以执行以下命令行下载php-ext-snappy源码。

git clone --recursive --depth=1 https://github.com/kjdev/php-ext-snappy.git

#### 2. 编译源码。

% cd php-ext-snappy % phpize % ./configure % make

% make install

#### 3. 在php.inf文件中配置扩展格式。

extension=snappy.so

完成以上配置后,就可以参考以下代码压缩和解压snappy格式文件。

```
$file_path = "test.snappy";
if (file_exists($file_path)) {
    $str = file_get_contents($file_path); //将整个文件内容读入到一个字符串中。
    $uncompressed = snappy_uncompress($str);
    echo $uncompressed;
}
```

### 更多参考

您可以访问GitHub上Snappy的开源项目,获取更多支持。更多信息,请参见开源Snappy。

# 2.6. RAM授权

当您使用RAM用户或跨阿里云账号投递日志服务数据到OSS时,需先完成RAM授权。本文介绍不同场景下的RAM授权步骤。

#### 前提条件

OSS Bucket拥有者已完成授权。具体操作,请参见云资源访问授权。

授权完成后,日志服务有权限扮演角色(STS机制)对OSS Bucket进行写入操作。

#### 概述

不同场景下,配置投递任务所需完成的RAM授权。

• 如果您需要对OSS Bucket进行更细粒度的访问控制,请参见修改授权策略。

- 如果日志服务Project和OSS Bucket不在同一个阿里云账号,跨阿里云账号投递的具体操作,请参见跨阿里云账号投 递。
- 如果您要通过RAM用户将日志服务中的日志投递到OSS Bucket , 请参见通过RAM用户投递(同账号)。
- 如果您要通过RAM用户将日志服务中的日志投递到其他阿里云账号下的OSS Bucket,请参见通过RAM用户投递(跨账号)。

#### 修改授权策略

通过云资源访问授权后,AliyunLogDefaultRole角色被授予AliyunLogRolePolicy权限,日志服务具有所有OSS Bucket的写入权限。如果您需要更精细的访问控制粒度,请解除AliyunLogDefaultRole角色的AliyunLogRolePolicy权限,为 AliyunLogDefaultRole角色授予自定义权限策略。具体操作,请参见RAM Policy概述。

#### 跨阿里云账号投递

如果您的日志服务Project和OSS Bucket不在同一个阿里云账号时需完成RAM授权才能进行投递操作,例如将阿里云账号A 下的日志服务中的日志投递至阿里云账号B下的OSS Bucket中,您可以按照以下步骤进行授权。

⑦ 说明 创建OSS投递任务时,只能指定一个OSS Bucket。故投递任务支持将阿里云账号A的日志投递到阿里云账号B,不支持将阿里云账号A的日志同时投递到阿里云账号B和C。如果您需要同时投递到两个阿里云账号的OSS Bucket,可以创建两个投递任务。

- 1. 阿里云账号B通过云资源访问授权创建角色AliyunLogDefaultRole。
- 2. 使用阿里云账号B登录RAM 控制台。
- 3. 在左侧导航栏中,选择身份管理>角色。
- 4. 在RAM角色列表中,单击AliyunLogDefaultRole。
- 5. 在信任策略管理页签中, 单击修改信任策略。

在Service配置项中添加**账号A对应的阿里云账号ID@log.aliyuncs.com**,其中请根据实际情况替换账号A对应的阿 里云账号ID,您可以在<mark>账号中心</mark>中查看阿里云账号ID。该策略表示阿里云账号A有权限通过日志服务获取临时Token来 操作阿里云账号B的资源。

```
{
"Statement": [
{
   "Action": "sts:AssumeRole",
   "Effect": "Allow",
   "Principal": {
        "Service": [
            "账号A对应的阿里云账号ID@log.aliyuncs.com",
            "log.aliyuncs.com"
        ]
      }
   }
},
"Version": "1"
}
```

6. 获取RAM角色标识(ARN)。

在该角色的基本信息中查看,例如acs:ram::13234:role/logrole。

使用阿里云账号A配置投递任务时,在RAM角色中输入此步骤中获取的RAM角色标识。

#### 通过RAM用户投递(跨账号)

如果阿里云账号A下的RAM用户A1创建投递任务,要将阿里云账号A下的日志服务中的数据投递至阿里云账号B下的OSS Bucket中,需要由阿里云账号A为RAM用户A1授予PassRole权限。

- 1. 阿里云账号B已完成跨阿里云账号投递中的配置。
- 2. 使用阿里云账号A登录RAM控制台。

- 3. 创建RAM用户A1。具体操作,请参见步骤一:创建RAM用户。
- 4. 为RAM用户A1授予AliyunRAMFullAccess权限。
  - i. 在左侧导航栏中, 单击**身份管理 > 用户**。
  - ii. 找到RAM用户A1, 单击添加权限。
  - iii. 选中系统策略下的AliyunRAMFullAccess, 单击确定。

授权成功后, RAM用户A1具备访问控制 (RAM) 的所有权限。

如果您需要控制RAM用户A1的权限范围,只授予投递到OSS的必要权限,可以为RAM用户A1授予自定义权限策略,策略脚本如下所示,其中Resource中配置为账号B的AliyunLogDefault Role的角色标识。创建自定义权限策略的操作步骤请参见通过RAM用户投递(同账号)。

```
{
   "Statement": [
   {
    "Action": "ram:PassRole",
    "Effect": "Allow",
    "Resource": "acs:ram::llllll:role/aliyunlogdefaultrole"
   }
  ],
  "Version": "1"
}
```

5. 获取RAM角色标识(ARN)。

在该角色的基本信息中查看,例如acs:ram::13234:role/logrole。

使用RAM用户A1配置投递任务时,RAM角色中输入此步骤中获取的RAM角色标识。

## 通过RAM用户投递(同账号)

如果您要使用RAM用户创建投递任务,需由阿里云账号为RAM用户授予相关权限。

- 1. 使用阿里云账号登录RAM 控制台。
- 2. 创建权限策略。
  - i. 在左侧导航栏中, 单击权限管理 > 权限策略管理。
  - ii. 单击创建权限策略。

#### iii. 在新建自定义权限策略页面中,配置如下参数,并单击确定。

参数	说明
策略名称	配置策略名称。
配置模式	选择 <b>脚本配置</b> 。
	将配置框中的原有脚本替换为如下内容。
策略内容	⑦ 说明 权限策略中需要包含PassRole权限。 {     "Version": "1",     "Statement": [     {         "Effect": "Allow",         "Action": "log:*",         "Resource": "*"     },     {         "Effect": "Allow",         "Action": "ram:PassRole",         "Resource": "*"     }     ] }

3. 创建RAM用户。具体操作,请参见步骤一:创建RAM用户。

4. 为RAM用户授权。

i. 在左侧导航栏中, 单击**身份管理 > 用户**。

- ii. 找到目标RAM用户,单击**添加权限**。
- iii. 单击自定义策略,选中步骤中创建的策略,然后单击确定。

# 3.投递日志到OSS(新版) 3.1. OSS投递新旧版本对比

本文介绍OSS投递新旧版本的区别以及如何将旧版本升级到新版本。

### 新旧版对比

对比项	OSS投递(旧版)	OSS投递(新版)
投递数据时间范围	支持从创建OSS投递作业时开始投递数据。	<ul><li> 支持从创建OSS投递作业时开始投递数据。</li><li> 支持投递历史数据。</li></ul>
作业数	在一个Logstore中,只可创建一个OSS投递作 业。	在一个Logstore中,可创建多个OSS投递作 业。
日期目录支持时区	不支持	支持
压缩类型	snappy	snappy、gzip、zstd
文件格式	JSON、CSV、Parquet	JSON、CSV、Parquet、ORC
观测仪表盘	不支持	支持
监控、告警	不支持	支持
暂停、启动任务	不支持	支持

# 迁移到新版

OSS投递新旧版本为并行关系,操作入口相同。您可以参见如下步骤完成迁移。

- 1. 停止OSS投递作业(旧版)。
- 2. 创建OSS投递作业(新版)。

具体操作,请参见创建OSS投递作业(新版)。

○ 注意 您在设置新版作业中的开始时间时,建议该时间早于停止旧版作业时间5分钟~15分钟(具体由旧版 作业中的投递时间参数决定)。该设置主要为了避免数据丢失,但可能导致投递少量的重复数据。

例如旧版作业中的**投递时间**为300秒,且您于2021-12-20 12:00:00停止旧版作业,那么建议将**开始时间**设置为 2021-12-20 11:55:00。

# 3.2. 准备权限

# 3.2.1. 授权概述

您使用OSS投递(新版)功能时,涉及OSS投递操作权限和数据访问权限。

- 如果您使用的是阿里云账号, 只需完成OSS投递作业访问数据的授权。
- 如果您使用的是RAM用户,则需完成OSS投递操作授权和OSS投递作业访问数据的授权。

↓ 注意 为了保障您的云资源安全,建议使用RAM用户。

### OSS投递操作授权

操作权限包括创建、删除、修改、查看OSS投递作业等。

- 阿里云账号:具备日志服务管理权限(AliyunLogFullAccess)。您无需为阿里云账号授予OSS投递操作权限。
- (推荐) RAM用户: 使用RAM用户操作OSS投递时, 您需要先使用阿里云账号为其授予OSS投递操作权限。更多信息, 请参见授予RAM用户操作OSS投递的权限。

#### OSS投递作业访问数据授权

OSS投递作业在读取源Logstore数据和写数据到OSS Bucket 时涉及数据访问。您可以通过默认角色或自定义角色完成数据 访问授权,详细说明如下:

● 默认角色: 授予OSS投递作业使用阿里云系统角色AliyunLogDefaultRole来读取源Logstore中的数据以及将数据写入到 OSS Bucket中。

阿里云系统角色AliyunLogDefaultRole具备读取Logstore数据和写数据到OSS Bucket的权限。更多信息,请参见通过默认角色访问数据。

• 自定义角色: 授予OSS投递作业使用自定义角色来读取源Logstore中的数据以及将数据写入到OSS Bucket中。

您需使用阿里云账号授予自定义角色读取Logstore数据和写数据到OSS Bucket的权限。更多信息,请参见通过自定义角 色访问数据(同账号场景)、通过自定义角色访问数据(跨账号场景)。

# 3.2.2. 授予RAM用户操作OSS投递的权限

当您使用RAM用户操作OSS投递(新版)功能时,需参见本文中的步骤完成授权。

#### 前提条件

已创建RAM用户。具体操作,请参见步骤一:创建RAM用户。

#### 操作步骤

- 1. 使用阿里云账号登录RAM控制台。
- 2. 创建权限策略。
  - i. 在左侧导航栏中,选择权限管理 > 权限策略。
  - ii. 单击创建权限策略。
  - iii. 在创建权限策略页面的脚本编辑页签中,将配置框中的原有脚本替换为如下内容,然后单击下一步。

请根据实际情况替换脚本中的Project名称和Logstore名称。

↓ 注意 如果您要使用RAM用户配置OSS投递作业告警,还需授予RAM用户告警操作权限。更多信息,请参见授予RAM用户告警操作权限。

```
{
 "Version": "1",
 "Statement": [{
   "Effect": "Allow",
   "Action": [
     "log:GetLogStore",
     "log:GetIndex",
     "log:GetLogStoreHistogram",
     "log:GetLogStoreLogs"
   ],
   "Resource": [
     "acs:log:*:*:project/Project名称/logstore/Logstore名称",
     "acs:log:*:*:project/Project名称/logstore/internal-diagnostic_log"
   ]
 },
   {
     "Effect": "Allow",
     "Action": [
       "log:CreateJob",
       "log:UpdateJob",
       "log:DeleteJob",
       "log:ListJobs",
       "log:GetJob"
     ],
     "Resource": "acs:log:*:*:project/Project名称/job/*"
   },
   {
     "Effect": "Allow",
     "Action": [
       "log:ListLogStores",
       "log:ListDashboard",
       "log:ListSavedSearch"
     ],
     "Resource": "acs:log:*:*:project/Project名称/*"
   },
   {
     "Effect": "Allow",
     "Action": [
       "ram:PassRole",
       "ram:GetRole",
       "ram:ListRoles"
     ],
     "Resource": "*"
   },
   {
     "Action": "ram:CreateServiceLinkedRole",
     "Resource": "*",
     "Effect": "Allow",
     "Condition": {
       "StringEquals": {
         "ram:ServiceName": "audit.log.aliyuncs.com"
       }
     }
   }
 ]
}
```

ⅳ. 设置**名称**,然后单击**确定**。

```
3. 为RAM用户授权。
```

- i. 在左侧导航栏中,选择**身份管理 > 用户**。
- ii. 找到目标RAM用户,单击添加权限。
- iii. 单击自定义策略,选中步骤中创建的策略。
- iv. 单击**确定**。

# 3.2.3. 通过默认角色访问数据

阿里云系统角色AliyunLogDefaultRole具备读取Logstore数据及将数据投递到OSS Bucket中的权限。您可以授予OSS投递 作业(新版)使用AliyunLogDefaultRole角色来读取源Logstore中的数据并将数据投递到OSS Bucket中。

#### 同账号投递

如果您的日志服务Logstore和OSS Bucket属于同一个账号,则只需使用该账号完成云资源访问授权。授权链接为<mark>云资源访</mark> 问授权。

授权完成后,OSS投递作业(新版)将使用AliyunLogDef ault Role角色来读取源Logstore中的数据并将数据投递到OSS Bucket中。即您在创建OSS投递作业(新版)时,需在**写OSS RAM角色**和读Logstore RAM角色中输入 AliyunLogDef ault Role角色的ARN(例如acs:ram::10\*\*\*\*12:role/aliyunlogdef ault role)。如何获取,请参见获取 AliyunLogDef ault Role的ARN。

*写OSS RAM角色:	<ul> <li>默认角色</li> <li>自定义角色</li> </ul>
	acs:ram::10 )612:role/aliyunlogdefaultrole
* 读Logstore RAM角色:	
	acs:ram::10 0612:role/aliyunlogdefaultrole
	分别用于OSS写入和Logstore读取的访问权限控制。如 "acs:ram:13234:role/logrole",细节请参考 <b>帮助</b>

#### 跨账号投递

如果日志服务Logstore和OSS Bucket不属于同一个阿里云账号,例如Logstore属于阿里云账号A,OSS Bucket属于账号 B,则两个账号分别完成云资源访问授权(授权链接为云资源访问授权)后,还需进行如下配置。

- 1. 使用阿里云账号B登录RAM 控制台。
- 2. 在左侧导航栏中,选择身份管理>角色。
- 3. 修改AliyunLogDefaultRole角色的信任策略。
  - i. 在RAM角色列表中, 单击AliyunLogDefault Role。
  - ii. 在信任策略管理页签中,单击修改信任策略。

iii. 修改信任策略。

在Service配置项中添加**阿里云账号A的ID@log.aliyuncs.com**,并根据实际情况替换该值。您可以在<del>账号中</del> 心查看阿里云账号ID。

该策略表示账号A有权限通过日志服务获取临时Token来操作账号B中的资源。

```
{
  "Statement": [
  {
    "Action": "sts:AssumeRole",
    "Effect": "Allow",
    "Principal": {
        "Service": [
            "账号A对应的阿里云账号ID@log.aliyuncs.com",
            "log.aliyuncs.com"
        ]
      }
    }
    ,
    "Version": "1"
}
```

#### iv. 单击确定。

授权完成后,OSS投递作业(新版)将使用账号A下的AliyunLogDefaultRole角色来读取源Logstore中的数据,使用账号B下的AliyunLogDefaultRole角色将数据投递到OSSBucket中。即您在创建OSS投递作业(新版)时,需在**写OSSRAM角色**中输入账号B的AliyunLogDefaultRole角色的ARN(例如

acs:ram::11\*\*\*\*13:role/aliyunlogdefaultrole),在读Logstore RAM角色中输入账号A的AliyunLogDefaultRole角 色的ARN(例如acs:ram::10\*\*\*\*12:role/aliyunlogdefaultrole)。如何获取,请参见获取AliyunLogDefaultRole的 ARN。

*写OSS RAM角色:	○ 默认角色 ● 自定义角色
	acs:ram::114 13:role/aliyunlogdefaultrole
* 读Logstore RAM角色:	<ul> <li></li></ul>
	acs:ram::10 10612:role/aliyunlogdefaultrole

#### 获取AliyunLogDefaultRole的ARN

- 1. 登录RAM 控制台。
- 2. 在左侧导航栏中,选择**身份管理 > 角色**。
- 3. 在RAM角色列表中,单击AliyunLogDefaultRole。
- 4. 在**基本信息**区域,获取ARN。

请记录该信息,如果您在创建OSS投递作业时使用的是默认角色,则需要输入该信息。

# 3.2.4. 通过自定义角色访问数据(同账号场景)

OSS投递作业(新版)运行时,将拉取Logstore中的数据并投递到OSS Bucket中,您可以使用自定义角色来完成以上操作。本文介绍如何对自定义角色进行授权。

#### 前提条件

已创建RAM角色。具体操作,请参见步骤一:创建RAM角色。

#### 步骤一:授予RAM角色读取Logstore数据的权限

授予RAM角色读取Logstore数据的权限后, OSS投递作业可以使用该角色读取Logstore中的数据。

- 1. 使用阿里云账号登录RAM控制台。
- 2. 创建具备读取Logstore数据权限的策略。
  - i. 在左侧导航栏,选择权限管理 > 权限策略。
  - ii. 单击创建权限策略。
  - iii. 在创建权限策略页面的脚本编辑页签中,将配置框中的原有脚本替换为如下内容,然后单击下一步。
    - 精确授权。

在实际场景中,请根据实际情况替换Project名称和Logstore名称。

```
{
    "Version":"1",
    "Statement":[
        {
            "Action":[
               "log:GetCursorOrData",
               "log:ListShards"
        ],
            "Resource":[
               "acs:log:*:*:project/Project名称/logstore/Logstore名称/*"
        ],
        "Effect":"Allow"
        }
    ]
}
```

■ 模糊匹配授权。

例如Project名称为log-project-dev-a、log-project-dev-b、log-project-dev-c等, Logstore名称为 website\_a\_log、website\_b\_log、website\_c\_log等,则您可以使用模糊匹配授权。在实际场景中,请根据实际情况替换Project名称和Logstore名称。

iv. 设置名称,然后单击确定。

例如设置策略名称为log-oss-export-source-policy。

- 3. 为RAM角色授权。
  - i. 在左侧导航栏中,选择身份管理>角色。
  - ii. 单击目标RAM角色对应的添加权限。
  - iii. 选择自定义策略,并选中步骤中创建的权限策略(例如log-oss-export-source-policy),单击确定。
  - i∨. 确认授权结果,单击**完成**。
- 获取RAM角色标识(ARN)。
   您可以在该角色的基本信息中查看RAM角色标识,例如acs:ram::13\*\*\*\*44:role/logrole。请记录该信息,如果您在

创建OSS投递作业时使用的是自定义角色,则需要在读Logstore RAM角色中输入该信息。

* 读日志服务授权:	○ 默认角色 ● 自定义角色	
	acs:ram::10 12:role/logrole	

### 步骤二: 授予RAM角色写OSS Bucket的权限

授予RAM角色写OSS Bucket的权限后, OSS投递作业可以使用该角色将Logstore中的数据写入到目标OSS Bucket中。

- 1. 使用阿里云账号登录RAM控制台。
- 2. 创建具备写OSS Bucket权限的策略。
  - i. 在左侧导航栏,选择权限管理 > 权限策略。
  - ii. 单击创建权限策略。
  - iii. 在创建权限策略页面的脚本编辑页签中,将配置框中的原有脚本替换为如下内容,然后单击下一步。

如果您有更细粒度的OSS权限控制需求,请参见RAM Policy概述。

iv. 设置名称,然后单击确定。

例如设置策略名称为log-oss-export-sink-policy。

- 3. 为RAM角色授权。
  - i. 在左侧导航栏中,选择**身份管理 > 角色**。
  - ii. 单击目标RAM角色对应的添加权限。
  - iii. 选择自定义策略,并选中步骤中创建的权限策略(例如log-oss-export-sink-policy),单击确定。
  - ⅳ. 确认授权结果,单击**完成**。
- 4. 获取RAM角色标识(ARN)。

您可以在该角色的基本信息中查看RAM角色标识,例如acs:ram::13\*\*\*\*44:role/ossrole。请记录该信息,如果您在创建OSS投递作业时使用的是自定义角色,则需要在写OSS RAM角色中输入该信息。

*写OSS RAM角色:	○ 默认角色 (●) 自定义角色
	acs:ram::134 1:role/ossrole
	用于访问权限控制,OSS Bucket拥有者创建角色的标示,如
	"acs:ram::13234:role/logrole"。帮助

# 3.2.5. 通过自定义角色访问数据(跨账号场景)

OSS投递作业(新版)运行时,将读取Logstore中数据并投递到OSS Bucket中。您可以授予OSS投递作业使用自定义角色 来完成以上操作。如果Logstore和OSS Bucket不属于同一个阿里云账号,您可参考本文档完成授权操作。

前提条件

已在阿里云账号A下创建RAM角色a,在阿里云账号B下创建RAM角色b。具体操作,请参见步骤一:创建RAM角色。

#### 步骤一:授予账号A下的RAM角色a读取Logstore数据的权限

授予账号A下的RAM角色a读取Logstore数据的权限后,OSS投递作业可以使用该角色读取账号A下Logstore中的数据。

- 1. 使用账号A登录RAM控制台。
- 2. 创建具备读取Logstore数据权限的策略。
  - i. 在左侧导航栏,选择**权限管理 > 权限策略**。
  - ii. 单击创建权限策略。
  - iii. 在创建权限策略页面的脚本编辑页签中,将配置框中的原有脚本替换为如下内容,然后单击下一步。
    - 精确授权。

在实际场景中,请根据实际情况替换Project名称和Logstore名称。

```
{
    "Version":"1",
    "Statement":[
        {
            "Action":[
               "log:GetCursorOrData",
               "log:ListShards"
              ],
              "Resource":[
               "acs:log:*:*:project/Project名称/logstore/Logstore名称/*"
             ],
            "Effect":"Allow"
        }
    ]
}
```

■ 模糊匹配授权。

例如Project名称为log-project-dev-a、log-project-dev-b、log-project-dev-c等, Logstore名称为 website\_a\_log、website\_b\_log、website\_c\_log等,则您可以使用模糊匹配授权。在实际场景中,请根据实际情况替换Project名称和Logstore名称。

iv. 设置名称,然后单击确定。

例如设置策略名称为log-oss-export-source-policy。

- 3. 为RAM角色a授权。
  - i. 在左侧导航栏中,选择**身份管理 > 角色**。
  - ii. 单击RAM角色a对应的添加权限。

iii. 选择自定义策略,并选中步骤中创建的权限策略(例如log-oss-export-source-policy),单击确定。

ⅳ. 确认授权结果,单击**完成**。

4. 获取RAM角色标识(ARN)。

您在该角色的基本信息中查看RAM角色标识,例如acs:ram::13\*\*\*\*44:role/logrole。请记录该信息,如果您在创建 OSS投递作业时使用的是自定义角色,则需要在读Logst ore RAM角色中输入该信息。

* 读日志服务授权:	○ 默认角色 ● 自定义角色	
	acs:ram::10	

#### 步骤二: 授予账号B下的RAM角色b写OSS Bucket的权限

授予账号B下的RAM角色b写OSS Bucket的权限后,OSS投递作业可以使用该角色将账号A下的Logstore数据投递到账号B下的OSS Bucket中。

- 1. 使用账号B登录RAM控制台。
- 2. 创建具备写OSS Bucket权限的策略。
  - i. 在左侧导航栏, 选择权限管理 > 权限策略。
  - ii. 单击创建权限策略。
  - iii. 在创建权限策略页面的脚本编辑页签中,将配置框中的原有脚本替换为如下内容,然后单击下一步。

如果您有更细粒度的OSS权限控制需求,请参见RAM Policy概述。

iv. 设置名称,然后单击确定。

例如设置策略名称为log-oss-export-sink-policy。

- 3. 为RAM角色b授权。
  - i. 在左侧导航栏中,选择**身份管理 > 角色**。
  - ii. 单击目标RAM角色b对应的添加权限。
  - iii. 选择自定义策略,并选中步骤中创建的权限策略(例如log-oss-export-sink-policy),单击确定。
  - iv. 确认授权结果, 单击**完成**。
- 4. 修改RAM角色b的信任策略。
  - i. 在RAM角色列表中, 单击RAM角色b。
  - ii. 在信任策略管理页签中,单击修改信任策略。

iii. 修改信任策略。

在Service配置项中添加**阿里云账号A的ID@log.aliyuncs.com**,并根据实际情况替换该值。您可以在<del>账号中</del> 心查看阿里云账号ID。

该策略表示账号A有权限通过日志服务获取临时Token来操作账号B中的资源。

iv. 单击确定。

5. 获取RAM角色标识(ARN)。

您可以在该角色的基本信息中查看RAM角色标识,例如acs:ram::11\*\*\*\*55:role/ossrole。请记录该信息,如果您在 创建OSS投递作业时使用的是自定义角色,则需要在写OSS RAM角色中输入该信息。

*写OSS RAM角色:	○ 默认角色 (●) 自定义角色
	acs:ram::114 55:role/ossrole
	用于访问权限控制,OSS Bucket拥有者创建角色的标示,如 "acs:ram::13234:role/logrole"。 <b>帮助</b>

# 3.3. 创建OSS投递作业(新版)

日志服务采集到日志后,支持将日志投递至OSS Bucket中进行存储与分析。本文介绍如何创建OSS投递作业(新版)。

#### 前提条件

- 已创建Project和Logstore。具体操作,请参见创建Project和Logstore。
- 已采集到日志。具体操作,请参见数据采集。
- 已在日志服务Project所在的地域创建Bucket。具体操作,请参见创建存储空间。

#### 支持的地域说明

日志服务投递数据到OSS为同地域投递,即日志服务Project所在的地域和OSS Bucket所在地域相同。

```
    ↓ 注意 目前只支持华东1(杭州)、华东2(上海)、华北2(北京)、华北3(张家口)、西南1(成都)、华南
    1(深圳)、中国(香港)。
```

### 创建投递作业

- 1. 登录日志服务控制台。
- 2. 在Project列表区域,单击目标Project。
- 3. 在日志存储 > 日志库页签中, 单击目标Logstore左侧的>, 选择数据处理 > 导出 > OSS(对象存储)。
- 4. 将鼠标悬浮OSS(对象存储)上,单击+。

# 5. 在OSS投递功能面板,配置如下参数,然后单击确定。

选择投递版本为新版,重要参数配置说明如下所示。

#### ◯ 注意

- 创建OSS投递作业后,每个Shard都会根据投递大小、投递时间决定数据投递的频率。当任一条件满足时,即会执行一次投递。
- 创建OSS投递作业后,您可以通过投递作业的状态和投递到OSS的数据确认投递规则是否符合预期结果。

参数	说明
OSS投递名称	投递作业的名称。
	OSS Bucket名称。
OSS Bucket	↓ 注意 必须是已存在的OSS Bucket名称,且该OSS Bucket与日志服务Project位 于相同地域。
ᅕᄮᄿᄬᄆᆿ	OSS Bucket中的目录。目录名不能以正斜线(/)或者反斜线(\)开头。
又忤投递目录	创建OSS投递作业后,Logstore中的数据将投递到目标OSS Bucket的此目录中。
文件后缀	如果您未设置文件后缀,则日志服务会根据存储格式和压缩类型自动生成后缀。例如 suffix。
分区格式	按照投递作业的创建时间动态生成OSS Bucket的目录,不能以正斜线(/)开头,默认值 为%Y/%m/%d/%H/%M,相关示例请参见 <mark>分区格式</mark> ,参数详情请参见 <mark>strptime API</mark> 。
	授予OSS投递作业将数据写入到OSS Bucket的权限。
	<ul> <li>默认角色:授权OSS投递作业使用阿里云系统角色AliyunLogDefaultRole将数据写入到 OSS Bucket中。即输入AliyunLogDefaultRole的ARN。如何获取ARN,请参见通过默认 角色访问数据。</li> </ul>
	• 自定义角色:授权OSS投递作业使用自定义角色将数据写入到OSS Bucket中。
写OSS RAM角色	您需先授予自定义角色将数据写入到OSS Bucket的权限,然后在 <b>写OSS RAM角色</b> 中输 入您自定义角色的ARN。如何获取ARN,请参见如下说明:
	<ul> <li>如果Logstore和OSS Bucket属于同一阿里云账号,请参见步骤二:授予RAM角色写 OSS Bucket的权限。</li> </ul>
	■ 如果Logstore和OSS Bucket属于不同的阿里云账号,请参见步骤二:授予账号B下的 RAM角色b写OSS Bucket的权限。
	授予OSS投递作业读取Logstore数据的权限。
	<ul> <li>默认角色:授权OSS投递作业使用阿里云系统角色AliyunLogDefaultRole来读取</li> <li>Logstore中的数据。即输入AliyunLogDefaultRole的ARN。如何获取ARN,请参见通过</li> <li>默认角色访问数据。</li> </ul>
	• 自定义角色:授权OSS投递作业使用自定义角色来读取Logstore中的数据。
读Logstore RAM角色	您需先授予自定义角色读取Logstore数据的权限,然后在 <b>读Logstore RAM角色</b> 中输 入您自定义角色的ARN。如何获取ARN,请参见如下说明:
	■ 如果Logstore和OSS Bucket属于同一阿里云账号,请参见步骤一:授予RAM角色读 取Logstore数据的权限。
	■ 如果Logstore和OSS Bucket属于不同的阿里云账号,请参见步骤一:授予账号A下的 RAM角色a读取Logstore数据的权限。

参数	说明
投递大小	每个Shard的投递大小。通过该值控制OSS Object大小(以未压缩计算),取值范围为 5~256,单位为MB。
存储格式	日志数据投递到OSS后,支持存储为不同的文件格式。更多信息,请参见CSV格式、JSON 格式、Parquet格式和ORC格式。
是否压缩	<ul> <li>OSS数据存储的压缩方式。</li> <li>不压缩(none):原始数据不压缩。</li> <li>压缩(snappy):使用snappy算法压缩数据,减少OSS Bucket的空间。更多信息,请参见snappy。</li> <li>压缩(zstd):使用zstd算法压缩数据,减少OSS Bucket的空间。</li> <li>压缩(gzip):使用gzip算法压缩数据,减少OSS Bucket的空间。</li> </ul>
投递时间	每个Shard的投递周期。取值范围为300~900,默认值为300,单位为秒。
开始时间	投递作业从该时间开始拉取Logstore中的数据。
时区选择	该时区用于格式化时间。 如果您设置了 <b>时区选择</b> 和 <b>分区格式</b> ,系统将根据您的设置生成OSS Bucket的目录。

## 查看OSS数据

将日志投递到OSS成功后,您可以通过OSS控制台、API、SDK或其它方式访问OSS数据。更多信息,请参见文件管理。 OSS Object地址格式如下所示:

oss://OSS-BUCKET/OSS-PREFIX/PARTITION-FORMAT\_RANDOM-ID

其中, OSS-BUCKET 为OSS Bucket名称, OSS-PREFIX 为目录前缀, PARTITION-FORMAT 为分区格式(由投递作业的 创建时间通过strptime API计算得到), RANDOM-ID 是一次投递行为的唯一标识。

⑦ 说明 OSS Bucket 目录是按照投递作业创建时间设置的。例如: 2022-01-23 00:00:00创建投递作业, 投递的是 2022-01-22 23:55后写入日志服务的数据, 假设5分钟投递一次数据到OSS Bucket。如果您要分析2022-01-22全天日志,除了查看 2022/01/22目录下的全部Object以外, 还需要检查 2022/01/23/00/目录下前十分钟的Object是否包含 2022-01-22的日志。

## 分区格式

一个投递作业对应一个OSS Bucket 目录,目录格式为*oss://OSS-BUCKET/OSS-PREFIX/PARTITION-FORMAT\_RANDOM-ID*。以创建时间为2022/01/20 19:50:43的投递作业为例,介绍分区格式,如下表所示。

OSS Bucket	OSS Prefix	分区格式	文件后缀	OSS文件路径
test-bucket	test-table	%Y/%m/%d/%H/%M	suffix	oss://test-bucket/test- table/2022/01/20/19/50_1 484913043351525351_2850 008.suffix
test-bucket	log_ship_oss_example	year=%Y/mon=%m/day=% d/log_%H%M%s	suffix	<i>oss://test- bucket/log_ship_oss_exam ple/year=2022/mon=01/da y=20/log_195043_1484913 043351525351_2850008.suf fix</i>

OSS Bucket	OSS Prefix	分区格式	文件后缀	OSS文件路径
test-bucket	log_ship_oss_example	ds=%Y%m%d/%H	suffix	oss://test- bucket/log_ship_oss_exam ple/ds=20220120/19_1484 913043351525351_2850008 .suffix
test-bucket	log_ship_oss_example	%Y%m%d/	suffix	oss://test- bucket/log_ship_oss_exam ple/20220120/_148491304 3351525351_2850008.suffix
				⑦ 说明 此格式会导 致Hive等平台无法解析对 应的OSS内容,建议您不 要使用该格式。
test-bucket	log_ship_oss_example	%Y%m%d%H	suffix	oss://test- bucket/log_ship_oss_exam ple/2022012019_14849130 43351525351_2850008.suffi x

使用Hive、MaxCompute等大数据平台或阿里云DLA产品分析OSS数据时,如果您希望使用Partition信息,可将文件名中的PARTITION-FORMAT设置为key=value格式。例如: *oss://test-*

*bucket/log\_ship\_oss\_example/year=2022/mon=01/day=20/log\_195043\_1484913043351525351\_2850008.parquet*, 设置为三层分区列,分别为: year、mon、day。

# 3.4. JSON格式

日志服务将数据投递到OSS后,支持存储为不同的文件格式。本文介绍JSON格式。

# 参数配置

在创建OSS投递作业(新版)时,如果选择存储格式为json,对应的参数配置如下所示。

* 是否投递tag: 🦳 是 💿 否	
参数	说明
是否投递tag	是否投递日志tag字段,例如tag:abc。

# OSS文件地址

投递到OSS后, OSS文件地址样例如下所示。

 ⑦ 说明

 ● 在创建OSS投递作业时,如果您自定义了文件后缀,则文件后缀由您自定义的后缀决定。

 ● 在创建OSS投递作业时,如果您未自定义文件后缀,则文件后缀由压缩类型决定。

 压缩类型
 文件后缀
 OSS文件地址示例
 说明

# 数据投递·投递日志到OSS(新版)

压缩类型	文件后缀	OSS文件地址示例	说明	
	如果您自定义了文件后缀,则以您 的设置为准,例如.json。	oss://oss-shipper- chengdu/ecs_test/2022/01/26/ 20/54_1453812893059571256_9 37.json	未压缩的JSON文件可下载到本地, 以文本方式打开查看。文件样例如 下所示。 {"time":1453809242,"	
不压缩	如果未自定义文件后缀,则无文件 后缀。	oss://oss-shipper- chengdu/ecs_test/2022/01/26/ 20/54_1453812893059571256_9 37	<pre>topic":"", "source_ _":"10.170.***.***", "ip" :"10.200.**.***", "time": "26/Jan/2016:19:54:02 +0800", "url":"POST</pre>	
压缩 (snappy)	如果您自定义了文件后缀,则以您 的设置为准,例如.json。	oss://oss-shipper- chengdu/ecs_test/2022/01/26/ 20/54_1453812893059571256_9 37.json	如果使用了snappy压缩,解压缩的 详细说明请参见 <mark>解压Snappy压缩</mark> 文件。	
	如果未自定义文件后缀,则文件后 缀为.snappy。	oss://oss-shipper- chengdu/ecs_test/2022/01/26/ 20/54_1453812893059571256_9 37.snappy		
压缩(gzip)	如果您自定义了文件后缀,则以您 的设置为准,例如.json。	oss://oss-shipper- chengdu/ecs_test/2022/01/26/ 20/54_1453812893059571256_9 37.json	如果使用了gzip压缩,则您可以安	
	如果未自定义文件后缀,则文件后 缀为.gz 。	oss://oss-shipper- chengdu/ecs_test/2022/01/26/ 20/54_1453812893059571256_9 37.gz	装并使用gzip命令进行解压。	
压缩(zstd)	如果您自定义了文件后缀,则以您 的设置为准,例如.json。	oss://oss-shipper- chengdu/ecs_test/2022/01/26/ 20/54_1453812893059571256_9 37.json	如果使用了zstd压缩,则您可以安	
	如果未自定义文件后缀,则文件后 缀为.zst。	oss://oss-shipper- chengdu/ecs_test/2022/01/26/ 20/54_1453812893059571256_9 37.zst	装并使用zstd命令进行解压。	

# 3.5. CSV格式

日志服务将日志投递到OSS后,支持存储为不同的文件格式。本文介绍CSV格式。

# 参数配置

* CSV字段:	Key名称	操作
	source	删除
	time	删除
	+ 如何使用oss	s shipper生成csv文件?
* 分隔符:	逗号	$\checkmark$
* 转义符:	双引号	$\sim$
无效字段内容:	指定字段名称不存在时投递内容,默认为空	
* 投递字段名称:		
	表示是否将字段名称写入CSV文件, 默认为不写入	

在创建OSS投递作业(新版)时,选择存储格式为csv,对应的参数配置如下所示。

参数说明如下所示。更多信息,请参见CSV标准和postgresql CSV说明。

参数	说明		
	在 <b>原始日志</b> 页签中查看日志字段,将需要投递到OSS的字段名有序填入。 日志服务还提供保留字段time、topic、source等。保留字段的更多信息,请参 见 <mark>保留字段</mark> 。		
CSV字段	<ul> <li>⑦ 说明</li> <li>同一个日志字段在CSV字段中只能配置一次,不支持多次使用。</li> <li>如果您的日志中存在同名字段(例如都为request_time),则日志服务会将其中 一个字段名显示为request_time_0,底层存储的字段名仍为request_time。因 此您在投递时,只能使用原始字段名request_time。</li> <li>存在同名字段时,系统只随机投递其中一个字段的值。请尽量避免日志中使用同 名字段。</li> </ul>		
分隔符	用于分割不同的日志字段。支持半角逗号(,)、竖线( )、空格和制表符。		
转义符	当字段内出现分隔符时,需要使用转义符包裹该字段,避免读数据时造成字段错误分割。		
无效字段内容	当CSV字段中指定的日志字段不存在时,投递此参数中配置的内容。		
投递字段名称	打开 <b>投递字段名称</b> 开关后,将字段名称写入到CSV文件中。		

# OSS文件地址

投递到OSS后, OSS文件地址样例如下所示。
#### ? 说明

- 在创建OSS投递作业时,如果您自定义了文件后缀,则文件后缀由您自定义的后缀决定。
- 在创建OSS投递作业时,如果您未自定义文件后缀,则文件后缀由压缩类型决定。

压缩类型	文件后缀	OSS文件地址举例	说明	
无压缩	如果您自定义了文件 后缀,则以您的设置 为准,例如.suffix。	oss://oss-shipper- chengdu/ecs_test/2022/01/26/20/ 54_1453812893059571256_937.suff ix	未压缩的CSV文件可下载到本地,以文	
	如果未自定义文件后 缀,则文件后缀 为.csv。	oss://oss-shipper- chengdu/ecs_test/2022/01/26/20/ 54_1453812893059571256_937.csv	<b>华万式打开宣</b> 有。	
压缩 (snappy)	如果您自定义了文件 后缀,则以您的设置 为准,例如.suffix。	oss://oss-shipper- chengdu/ecs_test/2022/01/26/20/ 54_1453812893059571256_937.suff ix	Snappy压缩文件的打开方式,请参 见 <mark>打开Snappy压缩文件</mark> 。	
	如果未自定义文件后 缀,则文件后缀 为.snappy.csv。	oss://oss-shipper- chengdu/ecs_test/2022/01/26/20/ 54_1453812893059571256_937.sna ppy.csv		
压结(asin)	如果您自定义了文件 后缀,则以您的设置 为准,例如.suffix。	oss://oss-shipper- chengdu/ecs_test/2022/01/26/20/ 54_1453812893059571256_937.suff ix	如果使用了gzip压缩,则您可以安装 并使用gzip命令进行解压。	
压缩(gzip)	如果未自定义文件后 缀,则文件后缀 为.gz.csv。	oss://oss-shipper- chengdu/ecs_test/2022/01/26/20/ 54_1453812893059571256_937.gz.c sv		
压缩(zstd)	如果您自定义了文件 后缀,则以您的设置 为准,例如.suffix。	oss://oss-shipper- chengdu/ecs_test/2022/01/26/20/ 54_1453812893059571256_937.suff ix	如果使用了zstd压缩, 则您可以安装 并使用zstd命令进行解压。	
	如果未自定义文件后 缀,则文件后缀 为.zst.csv。	oss://oss-shipper- chengdu/ecs_test/2022/01/26/20/ 54_1453812893059571256_937.zst. csv		

# 3.6. Parquet格式

日志服务将日志投递到OSS后,支持存储为不同的文件格式。本文介绍Parquet格式。

参数配置

\* Parquet字段: Key名称 类型 操作 \_\_topic\_\_ string V 删除 concent string V 删除 + 如何使用oss shipper生成parquet文件?

#### 在创建OSS投递作业(新版)时,选择存储格式为parquet,对应的参数配置如下所示。



参数	说明
Key名称	<ul> <li>您可以在原始日志页签中查看日志字段,将需要投递到OSS的字段名有序填入。OSS投送作业将按照此顺序组织Parquet数据,并使用该字段名作为Parquet数据的列名。</li> <li>日志服务还提供保留字段_time_、_topic_、_source_。保留字段的更多信息,请参见保留字段。</li> <li>如果遇到如下两种情况时,Parquet数据的列值为null。</li> <li>此处配置的字段名在Logstore中不存在。</li> <li>字段的数据类型由string类型转换非string类型(如double、int64等)失败。</li> <li>⑦ 说明</li> <li>同一个字段名在Parquet字段中只能配置一次,不支持多次使用。</li> <li>如果您的日志中存在同名字段(例如都为request_time),则日志服务会将其中一个字段名显示为request_time_0,底层存储的字段名仍为request_time。因此您在投递时,只能使用原始字段名request_time。存在同名字段时,系统只随机投递其中一个字段的值。请尽量避免日志中使用同名字段。</li> </ul>
	Parquet存储支持6种类型: string、boolean、int32、int64、float、double。
类型	日志投递过程中,会将日志服务中的日志字段由string类型转换为Parquet目标类型。如 果转换到非string类型失败,则该列数据为null。

#### OSS文件地址

投递到OSS后, OSS文件地址示例如下表所示。

? 说明

- 在创建OSS投递作业时,如果您自定义了文件后缀,则文件后缀由您自定义的后缀决定。
- 在创建OSS投递作业时,如果您未自定义文件后缀,则文件后缀由压缩类型决定。

压缩类型	文件后缀	OSS文件地址示例	说明
	如果您自定义了文件 后缀,则以您的设置 为准,例如.suffix。	oss://oss-shipper- chengdu/ecs_test/2022/01/26/20/ 54_1453812893059571256_937.suff ix	
不压缩			

#### 数据投递·投递日志到OSS (新版)

压缩类型	文件后缀	OSS文件地址示例	说明	
	如果未自定义文件后 缀,则文件后缀 为.parquet。	oss://oss-shipper- chengdu/ecs_test/2022/01/26/20/ 54_1453812893059571256_937.par quet		
	如果您自定义了文件 后缀,则以您的设置 为准,例如.suffix。	oss://oss-shipper- chengdu/ecs_test/2022/01/26/20/ 54_1453812893059571256_937.suff ix		
тан (зпарру)	如果未自定义文件后 缀,则文件后缀 为.snappy.parquet	oss://oss-shipper- chengdu/ecs_test/2022/01/26/20/ 54_1453812893059571256_937.sna ppy.parquet	下载到本地,使用 <mark>parquet-tools</mark> 工具 打开。	
	如果您自定义了文件 后缀,则以您的设置 为准,例如.suffix。	oss://oss-shipper- chengdu/ecs_test/2022/01/26/20/ 54_1453812893059571256_937.suff ix		
TEUR (GTID)	如果未自定义文件后 缀,则文件后缀 为.gz.parquet	oss://oss-shipper- chengdu/ecs_test/2022/01/26/20/ 54_1453812893059571256_937.gz. parquet		
压烷 (actd)	如果您自定义了文件 后缀,则以您的设置 为准,例如.suffix。	oss://oss-shipper- chengdu/ecs_test/2022/01/26/20/ 54_1453812893059571256_937.suff ix		
	如果未自定义文件后 缀,则文件后缀 为.zst.parquet	oss://oss-shipper- chengdu/ecs_test/2022/01/26/20/ 54_1453812893059571256_937.zst. parquet		

### 数据消费

- 通过E-MapReduce、Spark、Hive消费数据。更多信息,请参见社区文档。
- 通过单机校验工具消费数据。

开源社区提供的Parquet-MR可以用来文件级别验证Parquet格式、查看schema、读取数据内容。您可以自行编译该工具或者单击parquet-tools-1.6.0rc3-SNAPSHOT,下载日志服务提供的版本。

#### 。 查看Parquet文件schema

#### ■ 命令

\$ java -jar parquet-tools-1.6.0rc3-SNAPSHOT.jar schema -d 00\_1490803532136470439\_124353.snappy.pa
rquet | head -n 30

#### ■ 返回结果

```
message schema {
 optional int32 time ;
 optional binary ip;
 optional binary source ;
 optional binary method;
 optional binary topic ;
 optional double seq;
 optional int64 status;
 optional binary time;
 optional binary url;
 optional boolean ua;
}
creator: parquet-cpp version 1.0.0
file schema: schema
 time : OPTIONAL INT32 R:0 D:1
ip: OPTIONAL BINARY R:0 D:1
. . . . . . .
```

#### ○ 查看Parquet文件全部内容

#### ■ 命令

\$ java -jar parquet-tools-1.6.0rc3-SNAPSHOT.jar head -n 2 00\_1490803532136470439\_124353.snappy.pa
rquet

#### ■ 返回结果

```
__time__ = 1490803230
ip = 192.0.2.1
__source__ = *.*.*
method = POST
__topic__ =
seq = 1667821.0
status = 200
time = 30/Mar/2017:00:00:30 +0800
__time__ = 1490803230
ip = 192.0.2.1
__source__ = *.*.*
method = POST
__topic__ =
seq = 1667822.0
status = 200
time = 30/Mar/2017:00:00:30 +0800
url = /PutData?Category=YunOsAccountOpLog&AccessKeyId=********&Date=Fri%2C%2028%20Jun%202013
```

更多用法请执行java - jar parquet - tools - 1.6.0rc3 - SNAPSHOT. jar - h命令查看帮助。

## 3.7. ORC格式

日志服务将日志投递到OSS后,支持存储为不同的文件格式。本文介绍ORC格式。

#### 参数配置

在创建OSS投递作业(新版)时,选择存储格式为orc,对应的参数配置如下所示。

压缩(snappy)			$\sim$
OSS数据存储的压缩方式,支持: none、snappy、gzip、zstd等。其中, none 表示原始数据不压缩, snappy/gzip/zstd表示使用对应算法对数据做压缩,可 以减少OSS Bucket存储空间使用量(不同存储格式所支持的压缩类型略有不 同)			
Key名称	类型		操作
topic	string	$\sim$	删除
concent	string	$\mathbf{\nabla}$	血味
	压缩(snappy) OSS数据存储的压缩方式,支持 表示原始数据不压缩,snappy// 以减少OSS Bucket存储空间使用 同) Key名称	圧缩(snappy)         OSS数据存储的压缩方式,支持: none、snappy,gzip, 表示原始数据不压缩, snappy/gzip/zstd表示使用对应算 以减少OSS Bucket存储空间使用量(不同存储格式所支 同)         Key名称       类型        topic       string	压缩(snappy) OSS数据存储的压缩方式,支持: none、snappy,gzip、zstd等。 表示原始数据不压缩,snappy/gzip/zstd表示使用对应算法对数据 以减少OSS Bucket存储空间使用量(不同存储格式所支持的压缩数 同) Key名称 类型

#### 相关参数说明如下表所示。

参数	说明
Key名称	<ul> <li>您可以在原始日志页签中查看日志字段,将需要投递到OSS的字段名有序填入。OSS投送作业将按照此顺序组织ORC数据,并使用该字段名作为ORC数据的列名。</li> <li>日志服务还提供保留字段_time_、_topic_、_source_。保留字段的更多信息,请参见保留字段。</li> <li>如果遇到如下两种情况时,ORC数据的列值为null。</li> <li>此处配置的字段名在Logstore中不存在。</li> <li>字段的数据类型由string类型转换非string类型(如double、int64等)失败。</li> <li>⑦ 说明 <ul> <li>同一个字段名在Orc字段中只能配置一次,不支持多次使用。</li> <li>如果您的日志中存在同名字段(例如都为request_time),则日志服务会将其中一个字段名显示为request_time),则日志服务会将其中一个字段名显示为request_time),则日志服务会有有关。</li> <li>亦在同名字段时,系统只随机投递其中一个字段的值。请尽量避免日志中使用同名字段。</li> </ul> </li> </ul>
类型	ORC存储支持6种类型:string、boolean、int32、int64、float、double。 日志投递过程中,会将日志服务中的日志字段由string类型转换为ORC目标类型。如果转 换到非string类型失败,则该列数据为null。

#### OSS文件地址

投递到OSS后, OSS文件地址示例如下表所示。

#### ? 说明

- 在创建OSS投递作业时,如果您自定义了文件后缀,则文件后缀由您自定义的后缀决定。
- 在创建OSS投递作业时,如果您未自定义文件后缀,则文件后缀由压缩类型决定。

压缩类型	文件后缀	OSS文件地址示例	说明
不压缩	如果您自定义了文件 后缀,则以您的设置 为准,例如.suffix。	oss://oss-shipper- chengdu/ecs_test/2022/01/26/20/ 54_1453812893059571256_937.suff ix	
	如果未自定义文件后 缀,则文件后缀 为.orc。	oss://oss-shipper- chengdu/ecs_test/2022/01/26/20/ 54_1453812893059571256_937.orc	
压缩(snappy)	如果您自定义了文件 后缀,则以您的设置 为准,例如.suffix。	oss://oss-shipper- chengdu/ecs_test/2022/01/26/20/ 54_1453812893059571256_937.suff ix	下载到本地,使用 <mark>orc-tools</mark> 工具打 开。
	如果未自定义文件后 缀,则文件后缀 为.snappy.orc	oss://oss-shipper- chengdu/ecs_test/2022/01/26/20/ 54_1453812893059571256_937.sna ppy.orc	
	如果您自定义了文件 后缀,则以您的设置 为准,例如.suffix。	oss://oss-shipper- chengdu/ecs_test/2022/01/26/20/ 54_1453812893059571256_937.suff ix	
压细 (ZSUU)	如果未自定义文件后 缀,则文件后缀 为.zst.orc	oss://oss-shipper- chengdu/ecs_test/2022/01/26/20/ 54_1453812893059571256_937.zst. orc	

### 数据消费

- 通过E-MapReduce、Spark、Hive消费数据。更多信息,请参见社区文档。
- 通过单机校验工具消费数据。

orc tool可用于查看ORC文件的Meta信息、读取数据等。你也可以单击orc-tools-1.7.2-uber.jar,下载日志服务提供的版本进行验证。

#### ○ 查看Meta信息

#### ■ 命令

java -jar ~/Downloads/orc-tools-1.7.2-uber.jar meta -p file.orc

#### ■ 返回结果

```
Processing data file /Users/xx/file.orc [length: 200779]
Structure for /Users/xx/file.orc
File Version: 0.12 with ORC CPP ORIGINAL by ORC C++ 1.7.2
Rows: 124022
Compression: ZSTD
Compression size: 65536
Calendar: Julian/Gregorian
Type: struct<bucket:string,bucket_region:string>
Stripe Statistics:
 Stripe 1:
   Column 0: count: 124022 hasNull: false
   Column 1: count: 124022 hasNull: false min: bucket0 max: sls-training-data sum: 1468133
   Column 2: count: 0 hasNull: true
File Statistics:
  Column 0: count: 124022 hasNull: false
 Column 1: count: 124022 hasNull: false min: bucket0 max: sls-training-data sum: 1468133
 Column 2: count: 0 hasNull: true
Stripes:
  Stripe: offset: 3 data: 199856 rows: 124022 tail: 97 index: 578
   Stream: column 0 section ROW INDEX start: 3 length 102
   Stream: column 1 section ROW INDEX start: 105 length 367
   Stream: column 2 section ROW INDEX start: 472 length 109
   Stream: column 0 section PRESENT start: 581 length 25
   Stream: column 1 section PRESENT start: 606 length 25
   Stream: column 1 section LENGTH start: 631 length 38989
   Stream: column 1 section DATA start: 39620 length 160794
   Stream: column 2 section PRESENT start: 200414 length 23
   Stream: column 2 section LENGTH start: 200437 length 0
   Stream: column 2 section DATA start: 200437 length 0
   Encoding column 0: DIRECT
   Encoding column 1: DIRECT V2
   Encoding column 2: DIRECT V2
File length: 200779 bytes
Padding length: 0 bytes
Padding ratio: 0%
```

#### ○ 读取数据

```
■ 命令
```

java -jar ~/Downloads/orc-tools-1.7.2-uber.jar data -n 5 file.orc

#### ■ 返回结果

```
Processing data file /Users/xx/file.orc [length: 200779]
{"bucket":"bucket3","bucket_region":"cn-hangzhou"}
{"bucket":"bucket4","bucket_region":"cn-hangzhou"}
{"bucket":"dashboard-bucket","bucket_region":"cn-hangzhou"}
{"bucket":"bucket2","bucket_region":null}
```

更多用法请执行java - jar orc-tools-1.7.2-uber.jar命令或者参见orc tool查看帮助。

## 3.8. 管理OSS投递作业(新版)

您可以在数据投递概览页面管理OSS投递作业,包括查看OSS投递作业的基础信息、统计报表,修改配置并重启作业、删除OSS投递作业等操作。

#### 前提条件

- 已创建OSS投递作业。具体操作,请参见创建OSS投递作业(新版)。
- 已开启作业运行日志。具体操作,请参见开启作业运行日志。

#### 操作入口

- 1. 登录日志服务控制台。
- 2. 在Project列表区域,单击目标Project。
- 3. 在左侧导航栏,选择作业>数据投递。
- 4. 单击目标OSS投递作业。

#### 查看OSS投递作业基础信息

您可以在基础信息区域查看OSS投递作业的创建时间、修改时间、源日志库、OSS Bucket等信息。

基础信息			
源日志库:	test-oss	状态:	<ul> <li>已終止</li> </ul>
任务名称:	job-163 -449072	调度ID:	e42ff2afi 74976edd494
规则修改时间:	2021-12-13 17:34:21	规则创建时间:	2021-12-13 17:04:56
OSS Bucket:	sis	OSS Prefix:	sls
攒批设置:	大小: 256MB 或时间: 300秒	存储格式:	json

## 查看统计图表

创建OSS投递作业后,日志服务默认创建一个仪表盘,您可以在仪表盘中查看OSS投递作业的运行指标。



图表名称	说明
读失败条数	从源Logstore中读取失败的日志条数。
读公网流量 (MB)	目前不支持公网,因此数据为0。
写成功条数	投递到目标OSS Bucket成功的日志条数。
写失败条数	投递到目标OSS Bucket失败的日志条数。
写公网流量(MB)	目前不支持公网,因此数据为0。
处理速率 (events/s)	每秒处理的日志条数。
进度落后	展示该Logstore下每个Shard的延时信息。
运行异常	展示OSS投递作业运行过程中的异常信息。 您可以在目标Project的internal-diagnostic_log Logstore中,通过关键词ERROR或者WARN 查看完整的错误日志。该目标Project为您在开启作业运行日志时所指定的Project。更多信 息,请参见 <mark>开通服务日志</mark> 。
运行状态	展示OSS投递作业的运行状态。包括投递时间、成功条数,失败条数,平均响应时间等信息。

#### 修改配置并重启作业

您可以单击修改投递配置,修改OSS投递作业的相关配置。

保存修改后,日志服务会自动重启作业。重启过程大概需要几十秒钟,您可以单击**刷新**,确认重启是否完成。重启后,请 务必确认作业已正常启动,并检查投递到OSS Bucket的数据符合预期。

⑦ 说明 无法修改OSS投递名称和开始时间。

### 删除OSS投递作业

如果您不再需要运行该OSS投递作业,您可以在数据投递概览页面的右上角选择更多 > 删除。

□ **警告** 删除OSS投递作业后,不可恢复,请谨慎操作。

## 3.9. 为OSS投递作业(新版)设置告警

日志服务已内置监控规则模板,您只需添加对应的告警实例即可实时监控OSS投递作业,并可通过钉钉等渠道接收到告警 通知。本文介绍设置告警的相关操作。

#### 前提条件

已创建OSS投递作业。具体操作,请参见创建OSS投递作业(新版)。

#### 步骤一:开启作业运行日志

- 1. 登录日志服务控制台。
- 2. 在Project列表区域,单击目标Project。 该Project为OSS投递作业所在的Project。
- 3. 在页面左上方,单击 🞧 图标。
- 4. 进入开通作业运行日志页面。
  - 如果您开未通过该Project的详细日志,则在服务日志页签中,单击开通服务日志。

- 如果您已开通过该Project的详细日志,则在服务日志页签中,单击 
   ◎图标。
- 5. 设置如下参数,然后单击确定。

参数	说明
作业运行日志	打开 <b>作业运行日志</b> 开关后,系统将在您指定的Project中自动创建一个名为internal- diagnostic_log的Logstore,用于存储Scheduled SQL、MaxCompute投递、OSS投递、 数据导入等作业的运行日志与错误日志。日志字段说明,请参见 <mark>数据导入、投递等作业的</mark> 运行日志。
日志存储位置	开通作业运行日志功能后,需要选择日志的存储位置,即需要指定Project。可以设置为: <ul> <li>自动创建(推荐)。</li> <li>当前Project。</li> <li>同一地域下的其他Project。</li> </ul>

开通作业运行日志后,您可以在指定Project下的internal-diagnostic\_log Logstore中查看OSS投递作业的运行日志与 错误日志,其日志主题(\_\_topic\_\_)为etl\_metrics。您也可以通过投递作业名称查询目标投递作业的运行日志与错 误日志,对应的查询语句为 job\_name:*作业名称*,例如 job\_name:job-1646\*\*\*\*946 。

#### 步骤二:配置行动策略

- 1. 登录日志服务控制台。
- 2. 进入行动策略管理页面。
  - i. 在Project列表区域,单击任意的Project。
  - ii. 在左侧导航栏中, 单击**告**警。
  - iii. 选择告警管理 > 行动策略。
- 3. 找到目标行动策略(sls.app.export.builtin),单击修改。

您也可以创建新的行动策略用于告警通知。具体操作,请参见创建行动策略。

 在编辑行动策略页面的第一行动列表页签中,将请求地址修改为钉钉群机器人的Webhook地址。其他选项,保持 默认配置。

如何获取钉钉群机器人的Webhook地址,请参见<mark>钉钉-自定义。</mark>您也可以根据业务需求,使用其他告警渠道。具体操 作,请参见<mark>通知渠道说明</mark>。

5. 单击确认。

#### 步骤三:添加告警实例

日志服务已内置5种监控规则模板。更多信息,请参见<mark>告警投递监控规则</mark>。您只需根据业务需求,添加对应的告警实例即 可。此处以添加数据投递延迟监控规则对应的告警实例为例。

- 1. 登录日志服务控制台。
- 2. 在Project列表区域,单击目标Project。

该Project为您internal-diagnostic\_log Logstore所在的Project。

- 3. 在左侧导航栏中, 单击告警。
- 4. 在规则/事务页签中,单击SLS数据投递。

规则	/事务	告警管理 ∨			Ø -
🄁 来源:		自走义告誓 (旧版) (0) 自走义告誓 (新版) (0) 🗌 系统内置告誓(19)			■ 规则视图 ▼
⑦ 状态:	=	规则状态 🗌 已开启 (0) 🗌 未开启 (19) 🗌 暫停中 (0) 🗌 运行中 (0)	□ 有新版本 (0) □ 有异常(0) □ 无异常(19)		
🔛 类别:	74	▶ 产品 🗌 💲 SLS数据加工(5) 🔽 💲 SLS数据投递(5) 🗌 💲 SLS Logtail(8)	□ SLS消费组(1) 全部类别 ~		
新建告诉	*	开启 关闭 临时关闭 恢复 升级 删除		C test-lv 🗸	请选择 ✓ 输入模版ID或模版描述筛选 C
		监控规则	类别	状态	操作
	+	数据投递延迟监控 ⑦	(内置告誓) SLS数据投递) 投递延迟监控 展开	● 未创建	添加 开启
	+	数据投递流量(绝对值)监控 ①	(内置告答) SLS数据投递 投递流量监控 展开	● 未创建	添加 开启
	+	数据投递异常报播监控 ①	(内置告容) SLS数据投递 投递报错监控 展开	● 未创建	汤动 开加
	+	数据投递失败条数监控 ⑦	内置告答)(SLS数据投递)(投递失败监控) 展开	● 未创建	<b>添加 开廊</b>
	+	数据投递流量(日同比) 监控 ⑦	内置告替 SLS数据投递 投递流量监控 展开	● 未创建	添加 开启

#### 5. 在规则列表中,单击数据投递延迟监控对应的添加。

#### 6. 在参数设置对话框中,配置监控规则,然后单击设置并开启。

参数	说明
告警名称	告警名称,支持自定义。
监控的投递作业名称	需要监控的OSS投递作业的名称。 <ul> <li>默认值为 .* ,表示监控当前阿里云账号下所有的OSS投递作业。</li> <li>多个投递作业之间可以使用竖线( )分隔。您还可以使用正则表达式 .* 进行配置,例如sche.*,表示监控以sche开头的投递作业。</li> </ul>
监控阈值	当OSS投递作业的延迟时间超过该值时,触发告警。默认值为300秒。
行动策略	当前告警所绑定的行动策略,日志服务将通过该行动策略给指定用户发送告警通知。 默认为 sls.app.export.builtin (SLS数据投递内置行动策略)。您也可以自定义 行动策略。具体操作,请参见 <mark>创建行动策略</mark> 。
严重度	定义告警消息的严重度。
静默期	告警静默期,即重复通知的间隔。重复的告警在静默期内不会被重复通知。例如1d(1 天)、2h(2小时)、3m(3分钟)。

#### 添加完成后,您可以在监控规则列表中,单击数据投递延迟执行规则对应的+,查看已开启的告警实例。

_	数据投递延迟监控 ③	(内置吉蓉) SLS数据投递) 投递延迟监控 展开	● 已开启(1)		屏蔽列表 ⑦	添加 关闭 临时关闭 删除 查看
	告醫名称		行动策略	状态	外部配置	操作
	数据投递延迟监控 ⑦		sls.app.export.bu	●已开启	屏蔽列表 ③	关闭 临时关闭 删除 设置 查看 关注

### 相关操作

为OSS投递作业设置告警后,您还可以进行如下操作。

操作	说明
屏蔽列表	屏蔽列表中的OSS投递作业不会触发告警。
关闭告警实例	关闭告警实例后,告警实例不会再触发告警, <b>状态</b> 变更为 <b>未开启</b> 。 该操作不会删除规则参数中已设置的信息。需要再次监控时,无需重新设置规则参数。

操作	说明
临时关闭告警实例	临时关闭告警实例后,在指定时间内不再触发告警。
恢复告警实例	处于临时关闭状态的监控实例,可随时恢复告警。
删除告警实例	删除告警实例, <b>状态</b> 变更为 <b>未创建</b> 。 该操作会删除规则参数中已设置的信息。需要再次监控时,需要重新设置规则参数。
设置告警实例	修改告警实例的配置参数。

## 告警投递监控规则

日志服务已内置如下5种监控规则模板,用于监控OSS投递作业。

● 数据投递延迟监控规则

项目	说明		
规则名称	数据投递延迟监控。		
作用	用于监控OSS投递作业中Shard消费延迟情况。当投递延迟时长大于监控阈值时,触发告警。		
参数配置	<ul> <li>监控的投递作业名称:需要监控的OSS投递作业的名称。</li> <li>默认值为 .* ,表示监控当前阿里云账号下所有的OSS投递作业。</li> <li>多个投递作业之间可以使用竖线 () 分隔。您还可以使用正则表达式 .* 进行配置,例如sche.*,表示监控以sche开头的投递作业。</li> <li>监控阈值:当OSS投递作业的延迟时间超过该值时,触发告警。默认值为300秒。</li> <li>行动策略:当前告警所绑定的行动策略,日志服务将通过该行动策略给指定用户发送告警通知。</li> <li>严重度:告警消息的严重度。</li> <li>静默期:告警静默期,即重复通知的间隔。重复的告警在静默期内不会被重复通知。例如1d(1天)、2h(2小时)、3m(3分钟)。</li> </ul>		
关联仪表盘	<b>数据流量观测</b> 仪表盘中的 <b>进度落后</b> 图表。		
消除方法	<ul> <li>请按照如下方法进行处理。</li> <li>. 确认源Logstore的数据量是否明显增长,处理速率是否同时提高。</li> <li>如果源Logstore的数据量增长的同时处理速率也提高,则说明源Logstore数据在增长,OSS投递作业在自动扩容资源。请先观察5分钟,如果延迟降时间到正常范围,则说明问题解除。否则请执行下一步。</li> <li>如果源Logstore的数据量增长的同时处理速率未提高或者Shard消费延迟还处于上涨趋势,则说明可能是因为源Logstore的Shard数目不足,导致数据投递资源扩展受限,您需要手动分裂源Logstore的Shard数目不足,导致数据投递资源扩展受限,您需要手动分裂源Logstore的Shard。具体操作,请参见分裂Shard。完成分裂后,请先观察5分钟。如果延迟降时间到正常范围,则说明问题解除。否则请执行下一步。</li> <li>如果都不是,则执行下一步。</li> <li>如果存在,请按照数据投递异常报错监控的消除方法进行处理。处理完成后,观察5分钟。如果延迟降时间到正常范围,则说明问题解除。否则请执行下一步。</li> <li>如果不存在,则执行下一步。</li> <li>派备Project、Logstore和OSS投递作业名称等信息,提交工单联系日志服务技术支持人员。</li> </ul>		

#### • 数据投递异常报错监控规则

项目			
规则名称	数据投递异常报错监控。		
作用	用于监控OSS投递作业中的异常报错。当OSS投递作业出现异常报错后,触发告警。		
参数配置	<ul> <li>监控的投递作业名称:需要监控的OSS投递作业的名称。</li> <li>默认值为 * ,表示监控您当前阿里云账号下所有的OSS投递作业。</li> <li>多个投递作业之间可以使用竖线 () 分隔。您还可以使用正则表达式 * 进行配置,例如sche.*,表示监控以sche开头的投递作业。</li> <li>行动策略:当前告警所绑定的行动策略,日志服务将通过该行动策略给指定用户发送告警通知。</li> <li>严重度:告警消息的严重度。</li> <li>静默期:告警静默期,即重复通知的间隔。重复的告警在静默期内不会被重复通知。例如1d(1天)、2h(2小时)、3m(3分钟)。</li> </ul>		
关联仪表盘	<b>数据流量观测</b> 仪表盘中的运行异常图表。		
消除方法	<ul> <li>请按照如下方法进行处理。</li> <li>如果错误信息中包含Unauthorized、InvalidAccessKeyld或SignatureNotMatch,则说明OSS投递作业读取源Logstore数据或者写入数据到OSS Bucket的权限不足。请检查授权操作。更多信息,请参见授权概述。</li> <li>如果错误信息中包含ProjectNotExist或LogStoreNotExist,则说明OSS投递作业涉及的Project或Logstore不存在。请登录日志服务控制台确认。</li> <li>如果错误信息中包含TransformError,则说明源Logstore中的数据不满足当前的数据投递逻辑。请从错误信息中找到原始数据的位置,然后修改OSS投递作业的配置并重启作业。如何修改OSS投递作业,请参见修改配置并重启作业。</li> <li>如果问题仍未解决,请准备Project、Logstore和OSS投递作业名称等信息,提交工单联系日志服务技术支持人员。</li> </ul>		

#### • 数据投递流量(绝对值)监控规则

项目	说明		
规则名称	数据投递流量(绝对值)监控。		
作用	用于监控OSS投递作业在5分钟内的平均投递条数。当平均投递的日志条数低于监控阈值时触 发告警。		
参数配置	<ul> <li>监控的投递作业名称:需要监控的OSS投递作业的名称。</li> <li>默认值为 * ,表示监控您当前阿里云账号下所有的OSS投递作业。</li> <li>多个投递作业之间可以使用竖线 () 分隔。您还可以使用正则表达式 * 进行配置,例如sche.*,表示监控以sche开头的投递作业。</li> <li>监控阈值:当平均投递条数低于该值时,触发告警。默认值为40000行/秒。</li> <li>行动策略:当前告警所绑定的行动策略,日志服务将通过该行动策略给指定用户发送告警通知。</li> <li>严重度:告警消息的严重度。</li> <li>静默期:告警静默期,即重复通知的间隔。重复的告警在静默期内不会被重复通知。例如1d(1天)、2h(2小时)、3m(3分钟)。</li> </ul>		
关联仪表盘	数据流量观测仪表盘中的处理速率(events/s)图表。		

项目	说明
消除方法	<ul> <li>请按照如下方法进行处理。</li> <li>i.确认投递流量变化趋势与源Logstore的数据量增长或减少趋势是否一致。</li> <li>如果一致,则说明是源Logstore数据量增长或减少导致的。无需处理。</li> <li>如果不一致,则执行下一步。</li> <li>ii.确认是否存在数据投递延迟监控告警。</li> <li>如果存在,请处理延迟告警。处理完成后,先观察15分钟。</li> <li>如果数据投递延迟回到1分钟以内,且投递流量与源Logstore数据量增长或减少的 趋势一致,则说明问题解除。</li> <li>如果数据投递延迟回到1分钟以内,但是投递流量与源Logstore数据量增长或减少的 的趋势仍然不一致,则进行下一步。</li> <li>如果不存在,则执行下一步。</li> <li>iii.准备Project、Logstore和OSS投递作业名称等信息,提交工单联系日志服务技术支持 人员。</li> </ul>

#### • 数据投递流量(日同比)监控规则

项目	说明
规则名称	数据投递流量(日同比)监控。
作用	用于监控OSS投递作业在5分钟内相比昨日的增长阈值和下跌阈值情况。当增长大于日同比增 长阈值或下跌超过日同比下跌阈值时触发告警。
参数配置	<ul> <li>监控的投递作业名称:需要监控的OSS投递作业的名称。</li> <li>默认值为 * ,表示监控您当前阿里云账号下所有的OSS投递作业。</li> <li>多个投递作业之间可以使用竖线 (])分隔。您还可以使用正则表达式 * 进行配置,例如sche.*,表示监控以sche开头的投递作业。</li> <li>日同比增长阈值:当数据投递量日增长大于该值时,触发告警。默认值为40%。</li> <li>日同比下跌阈值:当数据投递量日下跌大于该值时,触发告警。默认值为20%。</li> <li>行动策略:当前告警所绑定的行动策略,日志服务将通过该行动策略给指定用户发送告警通知。</li> <li>严重度:告警消息的严重度。</li> <li>静默期:告警静默期,即重复通知的间隔。重复的告警在静默期内不会被重复通知。例如1d(1天)、2h(2小时)、3m(3分钟)。</li> </ul>
关联仪表盘	数据流量观测仪表盘中的处理速率(events/s)图表。

项目	说明
消除方法	<ul> <li>请按照如下方法进行处理。</li> <li>i. 确认投递流量变化趋势与源Logstore的数据量增长或减少趋势是否一致。</li> <li>如果一致,则说明是源Logstore数据量增长或减少导致的。无需处理。</li> <li>如果不一致,则执行下一步。</li> <li>ii. 确认是否存在数据投递延迟监控告警。</li> <li>如果存在,请处理延迟告警。处理完成后,先观察15分钟。</li> <li>如果数据投递延迟回到1分钟以内,且投递流量与源Logstore数据量增长或减少的 趋势一致,则说明问题解除。</li> <li>如果数据投递延迟回到1分钟以内,但是投递流量与源Logstore数据量增长或减少的 的趋势仍然不一致,则进行下一步。</li> <li>如果不存在,则执行下一步。</li> <li>iii. 准备Project、Logstore和OSS投递作业名称等信息,提交工单联系日志服务技术支持 人员。</li> </ul>

#### • 数据投递失败条数监控规则

项目	说明
规则名称	数据投递失败条数监控。
作用	用于监控OSS投递作业在15分钟内的投递失败情况。当投递失败的数据条数大于监控阈值时,触发告警。
参数配置	<ul> <li>监控的投递作业名称:需要监控的OSS投递作业的名称。</li> <li>默认值为 * ,表示监控您当前阿里云账号下所有的OSS投递作业。</li> <li>多个投递作业之间可以使用竖线 () 分隔。您还可以使用正则表达式 * 进行配置,例如sche.*,表示监控以sche开头的投递作业。</li> <li>监控阈值:当投递失败的数据条数超过该值时,触发告警。默认值为10条。</li> <li>行动策略:当前告警所绑定的行动策略,日志服务将通过该行动策略给指定用户发送告警通知。</li> <li>严重度:告警消息的严重度。</li> <li>静默期:告警静默期,即重复通知的间隔。重复的告警在静默期内不会被重复通知。例如1d(1天)、2h(2小时)、3m(3分钟)。</li> </ul>
关联仪表盘	数据流量观测仪表盘中的读(写)失败条数图表。
消除方法	<ul> <li>请按照如下方法进行处理。</li> <li>i.确认是否存在数据投递异常报错告警。</li> <li>如果存在,请按照数据投递异常报错监控的消除方法进行处理。处理完成后,观察5 分钟。如果投递失败的条数降到正常范围,则说明问题解除。否则请执行下一步。</li> <li>如果不存在,则执行下一步。</li> <li>ii. 准备Project、Logstore和OSS投递作业名称等信息,提交工单联系日志服务技术支持 人员。</li> </ul>

# 3.10. 时区列表

本文介绍OSS投递(新版)所涉及的时区格式以及常见的时区列表。该时区主要用于格式化时间。

### 时区格式设置

OSS投递(新版)中的时区格式为 {±偏移小时数}{分钟数} 。

- 偏移小时数的取值范围: [-12,+14]。
- 分钟数的取值范围: 00、30、45。
- 整体取值范围: [-1200,+1400]。

⑦ 说明 小时和分钟都必须为两位数字。

您可以在创建OSS投递作业(新版)时,设置时区选择。具体操作,请参见创建OSS投递作业(新版)。

x) <u>x7x</u> 3)+	时区选择:	+0800	$\sim$
-------------------	-------	-------	--------

### 常见时区列表

时区格式	全称	中文名称
	China Standard Time	中国标准时间
	Hong Kong Time	中国香港时间
	Australian Western Time	澳大利亚西部时间
	Korea Standard Time	韩国标准时间
. 0000	Malaysia Time	马来西亚时间
+0800	Philippile Time	菲律宾时间
	Singapore Time	新加坡时间
	Central Indonesian Time	印度尼西亚中部时间
	Ulaanbaatar Time	乌兰巴托时间
	Choibalsan Time	乔巴山时间
10000	Japan Standard Time	日本标准时间
+0900	Ulaanbaatar Summer Time	乌兰巴托夏令时
+0930	Australian Central Standard Time	澳大利亚中部标准时间
+1200	New Zealand Standard Time	新西兰标准时间
+0530	India Standard TIme	印度标准时间
0000	Greenwich Mean Time	格林威治标准时间
-0000	Western Europen Time	西欧时间
-0400	Atlantic Standard Time	大西洋标准时间(美国)
-0500	Eastern Standard Time	东部标准时间(美国)
-0600	Central Standard Time	中部标准时间 (美国)
-0700	Mountain Standard Time	山区标准时间(美国)

### 数据投递·投递日志到OSS(新版)

时区格式	全称	中文名称
-0800	Pacific Standard Time	太平洋标准时间 (美国)

# 4.投递日志到MaxCompute(旧版)

您可以在日志服务控制台实时查询所采集到的日志,并将日志投递到MaxCompute,进行进一步的BI分析及数据挖掘。本 文介绍通过日志服务投递日志到MaxCompute的操作步骤。

#### 前提条件

已开通MaxCompute,并创建表。更多信息,请参见创建表。

#### 使用限制

- 只有阿里云账号能够创建投递任务,不支持RAM账号操作。
- 不同Logstore中的数据请勿投递到同一张MaxCompute表中,可能造成多个投递任务的数据写入到一张MaxCompute表中。
- 投递不支持日志时间(对应保留字段\_\_time\_\_)距离当前时间14天以前的数据,在投递过程中自动丢弃超过14天的数据。
- 通过日志服务投递日志到MaxCompute,暂不支持DECIMAL、DATETIME、DATE、TIMESTAMP数据类型。更多信息,请参见2.0数据类型版本。
- 支持的地域如下所示,其他地域请使用DataWorks进行数据同步。更多信息,请参见日志服务通过数据集成投递数据。

日志服务Project所在地域	MaxCompute所在地域
华北1(青岛)	华东2(上海)
华北2(北京)	华北2(北京)、华东2(上海)
华北3(张家口)	华东2(上海)
华北5(呼和浩特)	华东2(上海)
华东1(杭州)	华东2(上海)
华东2(上海)	华东2(上海)
华南1(深圳)	华南1(深圳)、华东2(上海)
中国(香港)	华东2(上海)

#### 步骤一: 创建投递任务

1.

- 2. 在Project列表区域,单击目标Project。
- 3. 在日志存储 > 日志库页签中,单击目标Logstore左侧的>,选择数据处理 > 导出 > MaxCompute(原ODPS)。
- 4. 单击开启投递。
- 5. 在投递提示对话框中, 单击直接投递。
- 6. 初始化数加平台。

选中我已同意《数加平台服务协议》和我已同意《开放数据处理服务(ODPS)服务条款》,单击确定。

初始化开通需10~20秒左右,请耐心等待。如果已完成初始化,将跳过该步骤。

7. 在MaxCompute投递功能页面, 配置投递规则, 并单击确定。

MaxCompute投递功	力能
选择要投递的区域:	华东2(上海)
Project名称:	sisa
Logstore名称:	0:
投递属性 <b>(帮助)</b>	
*投递名称:	nginx-err-log
* 项目名:	TS_DL
	请输入要投递区域的项目名称
* MaxCompute表名:	test-odps
* MaxCompute普通列:	_time_ CD time string $\lor$
	•
* MaxCompute分区列:	_partition_time_ C> pt
	•
*时间分区格式:	yyyy_MM_dd_HH
•导入时间间隔:	1800s 🗸
投递注意事项:	投递注意事项 +

#### 重要参数配置如下所示。

参数	说明
选择要投递的区域	日志服务Project所在地域不同,所支持的MaxCompute地域不同。更多信息,请参见 <mark>使用限</mark> <mark>制</mark> 。
投递名称	投递任务的名称。命名规则如下: • 只能包含小写字母、数字、下划线(_)和短划线(-)。 • 必须以小写字母或者数字开头和结尾。 • 长度必须在3~63字符之间。
项目名	MaxCompute项目名称,支持新建或选择已创建的MaxCompute项目。
MaxCompute表名	MaxCompute表名称,支持新建或选择已创建的MaxCompute表。

参数	说明
MaxCompute普通列	<ul> <li>左边输入框中填写与MaxCompute表列相映射的日志字段名称,右边为MaxCompute表的列名称。更多信息,请参见数据模型映射。</li> <li> ↓ 注意 <ul> <li>日志服务投递日志到MaxCompute按照日志字段与MaxCompute表列的顺序进行映射,修改MaxCompute表列名不影响数据投递。如果更改MaxCompute表schema,请重新配置日志字段与MaxCompute表列映射关系。</li> <li>左边输入框的日志字段,不支持双引号("")、单引号(")和含有空格的字符串。</li> <li>如果您的日志中存在同名字段(例如都为request_time),则日志服务会将其中一个字段名显示为request_time_0,底层存储的字段名仍为request_time。因此您在投递时,只能使用原始字段名request_time。存在同名字段时,系统只随机投递其中一个字段的值。请尽量避免日志中使用同名字段。</li> </ul></li></ul>
MaxCompute分区列	左边输入框中填写与MaxCompute表分区列相映射的日志字段名称,右边为MaxCompute表 分区列名称。更多信息,请参见数据模型映射。 ⑦ 说明 最大配置3个分区列,请谨慎选择自定义字段作为分区列,保证一次投递任 务中生成的分区数目小于512个,否则会导致投递任务写数据到MaxCompute表失败, 整批数据无法投递。
时间分区格式	时间分区格式,配置示例和参数详细请参见参考示例和Java SimpleDateFormat。 ⑦ 说明 <ul> <li>仅当分区字段配置为partition_time时,时间分区格式才生效。</li> <li>请勿使用精确到秒的日期格式,易导致单表的分区数目超过限制(60000个)。</li> <li>单次投递任务的数据分区数目必须在512以内。</li> </ul>
导入时间间隔	投递任务的时长,默认值为1800,单位为秒。 当投递任务的时长达到此处设置的大小时,会自动创建一个新的投递任务。

开启投递后,一般情况下日志数据会在写入Logstore后的1个小时导入到MaxCompute,导入成功后即可在 MaxCompute内查看到相关日志数据。更多信息,请参见日志投递MaxCompute后,如何检查数据完整性。

### 步骤二: 查看MaxCompute数据

投递到MaxCompute成功后,您可以查看MaxCompute数据,数据样例如下所示。您可以使用已经与MaxCompute绑定的 大数据开发工具Data IDE来消费数据,进行可视化的BI分析及数据挖掘。

log_source   log_time   log_topic   time   ip   thread   log_extract_others   log_partition_time   st atus					
+++++++					
+					
10.10.*.*   1453899013     27/Jan/2016:20:50:13 +0800   10.10.*.*   414579208   {"url":"POST /PutData					
?Category=YunOsAccountOpLog&AccessKeyId=************************************					
OGMT&Topic=raw&Signature=************************************					
016_01_27_20_50   200					
+++++++					
+					

### 授予日志服务账号投递权限

在数加平台删除表后再重建,会导致默认授权失效,您需手动为日志服务投递数据操作重新授权。

- 1. 登录DataWorks控制台。
- 2. 在工作空间列表页面,单击目标工作空间对应的进入数据开发。
- 3. 新建业务流程。
  - i. 在数据开发页面,将鼠标悬停至+新建图标,单击业务流程。
  - ii. 在新建业务流程对话框中, 配置业务名称, 并单击新建。
- 4. 新建节点。
  - i. 在数据开发页面,将鼠标悬停至+ 新建图标,选择 Max Compute > ODPS SQL。
  - ii. 在新建节点对话框中,配置节点名称和目标文件夹,并单击提交。 其中,目标文件夹需配置为您在步骤中所创建的业务流程。
- 5. 在已创建的节点编辑框中,执行如下命令,完成授权。



命令	含义		
ADD USER aliyun\$shennong_open@aliyun.com;	在MaxCompute项目空间中添加用户。 其中shennong_open@aliyun.com为日志服务账号(固定 值,请勿修改)。		
<pre>GRANT Read, List ON PROJECT {ODPS_PROJECT_NAME } TO USER aliyun\$shennong_open@aliyun.com;</pre>	授予用户读和查询MaxCompute项目空间的权限。 其中 <i>{ODPS_PROJECT_NAME</i> 为MaxCompute项目空间名称, 请使用实际值替换。		
GRANT Describe, Alter, Update ON TABLE <b>{ODP5_T</b> ABLE_NAME} TO USER aliyun\$shennong_open@aliyun.c om;	授予用户Describe、Alter和Update权限。 其中 <i>{ODPS_TABLE_NAME]</i> 为MaxCompute表名称,请使用实 际值替换。		

命令	含义	
	确认授权是否成功。 如果返回如下信息表示授权成功。	
SHOW GRANTS FOR aliyun\$shennong_open@aliyun.co m;	<pre>A projects/{ODPS_PROJECT_NAME}: List   Read A projects/{{ODPS_PROJECT_NAME}/tables/{ODPS_T ABLE_NAME}: Describe   Alter   Update</pre>	

#### 相关操作

创建投递任务后,您可以在**MaxCompute(原ODPS)投递管理**页面,执行修改投递配置、关闭投递任务、查看投递任 务状态及错误信息、重试投递任务等操作。

● 修改投递配置

单击**投递配置**,修改投递配置,参数详情请参见本文中的<mark>步骤一:创建投递任务</mark>。其中如果想新增列,可以在大数据 计算服务MaxCompute修改投递的数据表列信息。

• 关闭投递任务

单击**关闭投递**,即可关闭投递任务。

• 查看投递任务状态及错误信息

日志服务支持查看过去两天内的所有日志投递任务及其投递状态。

○ 任务状态

状态	说明
成功	投递任务正常运行。
进行中	投递任务进行中,请稍后查看是否投递成功。
失败	因外部原因而无法重试的错误导致投递任务失败,请根据错误信息进行排查并重试。例如 MaxCompute表结构不符合日志服务规范、无权限等。 日志服务支持重试最近2天内所有失败的任务。

#### ○ 错误信息

如果投递任务出现错误,控制台上会显示相应的错误信息。

错误信息	建议方案
MaxCompute项目空间不存在	在MaxCompute控制台中确认您配置的MaxCompute项目是否存在,如果不存在则 需要创建。日志服务不会主动重试该错误,请在解决问题后手动重试。
MaxCompute表不存在	在MaxCompute控制台中确认您配置的MaxCompute表是否存在,如果不存在则需 要创建。日志服务不会主动重试该错误,请在解决问题后手动重试。
MaxCompute项目空间或表没有向日志 服务授权	在MaxCompute控制台中确认授权给日志服务账号的权限是否存在,如果不存在则 需要重新授权。更多信息,请参见 <mark>授予日志服务账号投递权限</mark> 。日志服务不会主动 重试该错误,请在解决问题后手动重试。
MaxCompute错误	显示投递任务收到MaxCompute错误,请参见MaxCompute相关文档或联系 MaxCompute团队解决。日志服务自动重试最近两天内的失败任务。
日志服务导入字段配置无法匹配 MaxCompute表的列	重新配置MaxCompute表的列与日志服务中日志字段的映射。日志服务不会主动重 试该错误,请在解决问题后手动重试。

o 重试任务

针对内部错误,日志服务支持按照策略自动重试。其他情况下,请您手动重试。自动重试的最小间隔是30分钟。当任 务失败后,等待30分钟再做重试。日志服务支持重试最近2天内所有失败的任务。

如果您需要立即重试失败任务,请单击重试全部失败任务或通过API、SDK指定任务进行重试。

### 参考信息

• \_\_\_partition\_time\_\_\_

将日志时间作为分区字段,通过时间筛选数据是MaxCompute常见的过滤数据的方法。

。 格式

\_\_partition\_time\_\_是根据日志服务中\_\_time\_\_字段的值计算得到的,结合分区时间格式,向下取整。为避免触发 MaxCompute单表分区数目的限制,日期分区列的值会根据投递间隔对齐。

例如:日志服务的日志时间为27/Jan/2016 20:50:13 +0800,日志服务据此计算出保留字段\_time\_为1453899013(Unix时间戳),不同配置下的时间分区列取值如下所示。

导入MaxCompute间隔	分区时间格式	partition_time	
1800	yyyy_MM_dd_HH_mm_00	2016_01_27_20_30_00	
1800	уууу-MM-dd HH:mm	2016-01-27 20:30	
1800	ууууMMdd	20160127	
3600	yyyyMMddHHmm	201601272000	
3600	yyyy_MM_dd_HH	2016_01_27_20	

。 使用方法

使用\_\_partition\_time\_\_ 筛选数据,可以避免全表扫描。例如查询2016年1月26日一天内日志数据,查询语句如下所示。

select \* from {ODPS\_TABLE\_NAME} where log\_partition\_time >= "2015\_01\_26" and log\_partition\_time < "
2016 01 27";</pre>

\_\_extract\_others\_\_

\_\_extract\_others\_\_是一个JSON字符串。例如要获取该字段的user-agent内容,查询语句如下所示。

select get\_json\_object(sls\_extract\_others, "\$.user-agent") from {ODPS\_TABLE\_NAME} limit 10;

? 说明

- get\_json\_object是MaxCompute提供的标准UDF,请联系MaxCompute团队开通使用该标准UDF的权限。更 多信息,请参见MaxCompute提供的标准UDF。
- 示例供参考,请以MaxCompute产品的建议为最终标准。

• 数据模型映射

将日志服务中的日志投递到MaxCompute时,涉及两个服务之间的数据模型映射问题,相关注意事项与示例如下所示。

- MaxCompute表至少包含一个数据列和一个分区列。
- 日志服务保留字段建议使用\_\_partition\_time\_\_、\_\_source\_\_、\_\_topic\_\_。
- 一个MaxCompute表的分区数最大值为60000个,当分区数超出最大值后无法再写入数据。
- 投递任务是批量执行的,请谨慎设置分区列及其类型,保证一个投递任务内处理的数据分区数小于512个,否则该批数据都无法写入MaxCompute。
- 系统保留字段\_\_extract\_others\_\_有曾用名\_extract\_others\_, 可兼容使用。
- MaxCompute分区列的值不支持配置为MaxCompute的保留字和关键字。更多信息,请参见保留字与关键字。
- MaxCompute分区列取值不支持配置为空,所以映射到分区列的字段必须为保留字段或日志字段,且可以通过cast运算符将string类型字段值转换为对应分区列类型,空分区列的日志会在投递中被丢弃。
- 日志服务中一个日志字段只能映射到一个MaxCompute表的列(数据列或分区列),不支持字段冗余,同一个字段名 第二次使用时其投递的值为null,如果null出现在分区列会导致数据无法被投递。

MaxCompute数据列、分区列与日志服务字段的映射关系示例如下所示,其中日志服务保留字段详情请参见保留字段。

MaxCompute 列类型	列名 (MaxCompute )	数据类型 (MaxCompute )	日志字段名称 (日志服务)	字段类型(日志 服务)	字段说明
	log_source	string	source	保留字段	日志来源。
	log_time	bigint	time	保留字段	日志的Unix时间戳(从 1970年1月1日开始所经过 的秒数),对应数据模型 中的Time域。
	log_topic	string	topic	保留字段	日志主题。
数据列	time	string	time	日志内容字段	解析自日志,对应数据模 型中的key-value。在很多 时候Logtail采集的数据的 time与time取值相 同。
	ip	string	ір	日志内容字段	解析自日志。
	thread	string	thread	日志内容字段	解析自日志。
	log_extract_ot hers	string	extract_othe rs	保留字段	未在配置中进行映射的其 他日志字段会通过key- value序列化到JSON中,该 JSON是一层结构,不支持 字段内部JSON嵌套。

MaxCompute 列类型	列名 (MaxCompute )	数据类型 (MaxCompute )	日志字段名称 (日志服务)	字段类型(日志 服务)	字段说明
	log_partition_ti me	string	partition_tim e	保留字段	由日志的time 字段对 齐计算而得,分区粒度可 配置。
נגסון	status	string	status	日志内容字段	解析自日志,该字段取值 支持枚举,保证分区数目 不超过上限。

# 5.投递日志到MaxCompute (新版) 5.1. MaxCompute投递新旧版对比

本文介绍MaxCompute投递新旧版本的区别以及如何将旧版本迁移到新版本。

#### 新旧版对比

对比项	MaxCompute投递(旧版)	MaxCompute投递(新版)
投递数据时间范围	支持从创建MaxCompute投递作业时开始 投递数据。	<ul> <li>支持从创建MaxCompute投递作业时 开始投递数据。</li> <li>支持投递历史数据。</li> </ul>
作业数	在一个Logstore中,只可创建一个 MaxCompute投递作业。	在一个Logstore中,可创建多个 MaxCompute投递作业。
2.0数据类型版本	不支持	支持
Tags	不支持	支持
DateTime映射支持时区	不支持	支持
观测仪表盘	不支持	支持
监控、告警	不支持	支持
暂停、启动任务	不支持	支持
时间分区格式	采用Class SimpleDateFormat格式	采用strftime格式
无效字段	显示为Null	如果不允许设置为Null,则丢弃整行。

### 迁移到新版

MaxCompute投递新旧版本为并行关系,操作入口相同。您可以参见如下步骤完成迁移。

- 1. 停止MaxCompute投递作业(旧版)。
- 2. 创建MaxCompute投递作业(新版)。

具体操作,请参见创建MaxCompute投递作业(新版)。

↓ 注意 您在设置新版作业中的起始时间时,建议该时间早于停止旧版作业时间5分钟~15分钟(具体由旧版 作业中的导入时间间隔参数决定)。该设置主要为了避免数据丢失,但可能导致投递少量的重复数据。

例如旧版作业中的**导入时间间隔**为1800秒,且您于2021-12-20 12:00:00停止旧版作业,那么建议将**起始时** 间设置为2021-12-20 11:55:00。

## 5.2. 稳定性说明与使用限制

本文介绍MaxCompute投递(新版)的稳定性与使用限制。

#### 稳定性说明

读日志服务

稳定项	说明
可用性	可用性较高。 如果日志服务出错,无法读取数据,MaxCompute投递作业会在内部至少重试10次。如 果仍然失败,作业执行会报错,然后作业重启。

## 写MaxCompute

稳定项	说明
并发度	按照日志服务Shard进行分区并创建投递实例,支持快速扩容。 如果日志服务源Logstore进行Shard分裂,可以在数秒以内完成投递实例的扩容,加快 数据导出速度。
数据不丢失	MaxCompute投递作业基于 <mark>消费组</mark> 进行扩展,提供一致性保证。投递完成后,才会提交 offset,因此可以保证数据写入MaxCompute之前,offset不被提交,即保证投递数据 不丢失。
Schema变更	如果投递过程中在MaxCompute表中添加了新列,则新列只会被写入新分区,不会被写 入旧分区以及当前分区。

## 处理脏数据

错误项	是否计入失败条数	说明
分区错误	是	常见场景为分区不合法或分区列不存在。该条数据不会写入 MaxCompute。
数据列不合法	否	常见场景类型为不匹配或者类型转换失败。该列数据不会写入 MaxCompute,其余列正常写入MaxCompute。
数据列过长	否	常见场景为数据超出string类型或者varchar长度限制。该列数据 经过截断后写入MaxCompute,其余列正常写入MaxCompute。

## 监控告警

稳定项	说明
监控告警	数据投递有完善的监控,可实时追踪投递作业的延迟、流量等指标。您可以根据业务需 求,配置自定义告警,及时发现投递问题(例如导出实例不足、网络Quota限制等)。 具体操作,请参见为MaxCompute <mark>投递作业(新版)设置告警</mark> 。

## 重启作业

稳定项	说明
分区数过多	作业重启时,因为分区数过多(5分钟仍未完成写入),可能导致数据重复。
数据写入失败	作业重启且数据写入MaxCompute失败(授权错误、网络错误)时,可能导致数据部分 重复。

## 使用限制

网络

限制项	说明
同地域投递的网络	同地域投递时,数据通过阿里云内网传输,因此网络稳定性和速度更有保障。
跨地域投递的网络	跨地域投递时,数据通过公网传输,因此网络稳定性和速度没有保障。 网络出现问题时,内部会重试至少10次,如果仍然失败,作业执行会报错,导致作业重 启。

## 读流量

限制项	说明
读流量	单个Project以及单个Shard存在最高流量限制。更多信息,请参见数据读写。 如果超过最高流量限制,请分裂Shard或者申请扩容Project读流量限制。超过限制,会 导致MaxCompute投递作业读取数据失败,并在内部至少重试10次,如果仍然失败,作 业执行会报错,然后作业重启。

## 写入性能

限制项	说明
并发实例	支持最大的导出并发实例为64。 如果日志服务Shard数量超过64,则会合并多个Shard到一个实例中进行导出处理,并且 尽量保证每个实例中的Shard数相同。
写入阈值	MaxCompute单分区写入上限为10 MB/s。 超过限制时,写数据到MaxCompute会不稳定。

## 权限管理

限制项	说明
写授权	MaxCompute写授权支持RAM用户以及RAM角色两种方式,并且需要在MaxCompute侧 进行单独操作。

## 数据类型

### ● 普通列

类型	示例	说明
string	"hello"	最大长度为8 MB。
datetime	"2021-12-22 05:00:00.123456"	日志服务中的数据需满足MaxCompute 的数据格式要求。
date	"2021-12-22"	日志服务中的数据需满足MaxCompute 的数据格式要求。
timestamp	1648544867	毫秒级以及秒级精度。
decimal	1.2	日志服务中的数据需满足MaxCompute 的数据格式要求。
char	"hello"	最大长度为255字节。

类型	示例	说明
varchar	"hello"	最大长度为65535字节。
binary	"hello"	最大长度为8 MB。
bigint	123	最大支持int64。
boolean	1	<ul> <li>1、t、T、true、TRUE、True解析为 True。</li> <li>0、f、F、false、FALSE、False解析 为False。</li> </ul>
double	1.2	最大支持64位浮点数。
float	1.2	最大支持32位浮点数。
integer	123	最大支持int32。
smallint	12	最大支持int16。
tinyint	12	最大支持int8。

#### ● 分区列

限制项	说明
分区列	按照字符串处理,需要满足MaxCompute分区列的格式要求。

## 5.3. 准备权限

## 5.3.1. 授权概述

您使用MaxCompute投递(新版)功能时,涉及MaxCompute投递操作权限和数据访问权限。

您可以使用阿里云账号或RAM用户操作MaxCompute投递作业。两者的授权说明如下:

↓ 注意 为了保障您的云资源安全,建议使用RAM用户。

- 阿里云账号:需完成MaxCompute投递作业访问数据的授权。
- RAM用户:需完成MaxCompute投递操作授权和MaxCompute投递作业访问数据的授权。

#### MaxCompute投递操作授权

操作权限包括创建、删除、修改、查看MaxCompute投递作业等。

- 阿里云账号:具备日志服务管理权限(AliyunLogFullAccess)。您无需为阿里云账号授予MaxCompute投递操作权限。
- RAM用户:使用RAM用户操作MaxCompute投递时,您需要先使用阿里云账号为其授予MaxCompute投递操作权限。更多信息,请参见授予RAM用户操作MaxCompute投递的权限。

#### MaxCompute投递作业数据访问授权

MaxCompute投递作业在读取源Logstore数据和写数据到MaxCompute表时涉及数据访问。您可以通过默认角色、自定义 角色或RAM用户完成数据访问授权,详细说明如下表所示。

操作	授权方式	说明
法Logstoro的权明可要	通过默认角色读取Logstore数 据	授权MaxCompute投递作业使用阿里云系统角色 AliyunLogDefaultRole读取源Logstore中的数据。
展LOGSTOTEBY 次限配量	通过自定义角色读取Logstore 数据	授权MaxCompute投递作业使用自定义角色读取源Logstore中 的数据。
	通过默认角色写数据到 MaxCompute	授权MaxCompute投递作业使用阿里云系统角色 AliyunLogDefaultRole将读取到的Logstore数据写入到 MaxCompute表中。
写MaxCompute的权限配置	通过自定义角色写数据到 MaxCompute	授权MaxCompute投递作业使用自定义角色将读取到的 Logstore数据写入到MaxCompute表中。
	通过RAM用户写数据到 MaxCompute	授权MaxCompute投递作业使用RAM用户(AccessKey方式)将 读取到的Logstore数据写入到MaxCompute表中。

## 5.3.2. 授予RAM用户操作MaxCompute投递的权限

当您使用RAM用户操作MaxCompute投递(新版)功能时,需参见本文中的步骤完成授权。

#### 前提条件

已创建RAM用户。具体操作,请参见步骤一:创建RAM用户。

#### 操作步骤

- 1. 使用阿里云账号登录RAM控制台。
- 2. 创建权限策略。
  - i. 在左侧导航栏中,选择**权限管理 > 权限策略**。
  - ii. 单击创建权限策略。
  - iii. 在**创建权限策略**页面的**脚本编辑**页签中,将配置框中的原有脚本替换为如下内容,然后单击下一步。 请根据实际情况替换脚本中的*Project名称*和*Logstore名称*。

↓ 注意 如果您要使用RAM用户配置MaxCompute投递作业告警,还需授予RAM用户告警操作权限。更多 信息,请参见授予RAM用户告警操作权限。

```
"Version": "1",
   "Statement": [{
           "Effect": "Allow",
           "Action": [
               "log:GetLogStore",
               "log:GetIndex",
               "log:GetLogStoreHistogram",
               "log:GetLogStoreLogs"
           ],
           "Resource": [
               "acs:log:*:*:project/Project名称/logstore/Logstore名称",
               "acs:log:*:*:project/Project名称/logstore/internal-diagnostic_log"
           ]
       },
       {
           "Effect": "Allow",
           "Action": [
               "log:CreateJob",
               "log:UpdateJob",
               "log:DeleteJob",
               "log:ListJobs",
               "log:GetJob"
           ],
           "Resource": "acs:log:*:*:project/Project名称/job/*"
       },
        {
           "Effect": "Allow",
            "Action": [
              "log:ListLogStores",
               "log:ListDashboard",
               "log:ListSavedSearch"
           ],
           "Resource": "acs:log:*:*:project/Project名称/*"
       },
        {
           "Effect": "Allow",
            "Action": [
               "ram:PassRole",
               "ram:GetRole",
               "ram:ListRoles"
           ],
           "Resource": "*"
       }
   ]
}
```

iv. 设置**名称**,然后单击确定。

- 3. 为RAM用户授权。
  - i. 在左侧导航栏中,选择**身份管理 > 用户**。
  - ii. 找到目标RAM用户,单击添加权限。
  - iii. 单击自定义策略,选中步骤中创建的策略。
  - ⅳ. 单击确定。

## 5.3.3. 通过默认角色读取Logstore数据

MaxCompute投递作业(新版)运行时,需要读取Logstore中的数据,您可以使用默认角色(AliyunLogDefaultRole角 色)完成该操作。本文介绍如何生成默认角色。

### 操作步骤

您只需使用当前账号完成云资源访问授权,生成AliyunLogDefaultRole角色。AliyunLogDefaultRole角色默认具备读取Logstore数据的权限。

- 1. 在MaxCompute投递功能面板中,选择读日志服务授权为默认角色。
- 2. 单击您需要授权系统角色。

MaxCompute投递功能	ŧ	$\times$
* 项目名:	454045666	
	请输入要投递区域的项目各称	
* MaxCompute表名:	220404	
*读日志服务授权:	<ul> <li>         ・         ・         ・</li></ul>	

3. 在云资源访问授权页面,单击同意授权。

完成上述操作后, MaxCompute投递作业(新版)将使用AliyunLogDefaultRole角色读取源Logstore中的数据。

## 5.3.4. 通过自定义角色读取Logstore数据

MaxCompute投递作业(新版)运行时,需要读取Logstore中的数据,您可以使用自定义角色完成该操作。本文介绍如何 对自定义角色进行授权。

#### 前提条件

已创建RAM角色(例如MaxcomputeShipRole)。具体操作,请参见创建RAM角色。

#### 操作步骤

授予RAM角色读取Logstore数据的权限后,MaxCompute投递作业可以使用该角色读取Logstore中的数据。

- 1. 登录RAM控制台。
- 2. 创建具备读取Logstore数据权限的策略。
  - i. 在左侧导航栏,选择权限管理 > 权限策略。
  - ii. 单击创建权限策略。

- iii. 在创建权限策略页面的脚本编辑页签中,将配置框中的原有脚本替换为如下内容,然后单击下一步。
  - 精确授权。

在实际场景中,请根据实际情况替换Project名称和Logstore名称。

```
{
    "Version":"1",
    "Statement":[
        {
            "Action":[
               "log:GetCursorOrData",
               "log:ListShards"
            ],
            "Resource":[
               "acs:log:*:*:project/Project名称/logstore/Logstore名称/*"
            ],
            "Effect":"Allow"
        }
    ]
}
```

■ 模糊匹配授权。

例如Project名称为log-project-dev-a、log-project-dev-b、log-project-dev-c等, Logstore名称为 website\_a\_log、website\_b\_log、website\_c\_log等,则您可以使用模糊匹配授权。在实际场景中,请根据实际情况替换Project名称和Logstore名称。

```
{
    "Version":"1",
    "Statement":[
        {
            "Action":[
                "log:GetCursorOrData",
                xuxu:ListShards"
               ],
               "Resource":[
                "acs:log:*:*:project/log-project-dev-*/logstore/website_*_log*"
               ],
               "Effect":"Allow"
               }
        ]
}
```

iv. 设置名称, 然后单击确定。

例如设置策略名称为log-maxcompute-export-source-policy。

- 3. 为RAM角色授权。
  - i. 在左侧导航栏中,选择**身份管理 > 角色**。
  - ii. 单击目标RAM角色对应的添加权限。
  - iii. 选择自定义策略,并选中步骤中创建的权限策略(例如log-maxcompute-export-source-policy),单击确定。
  - iv. 确认授权结果, 单击**完成**。

完成上述操作后,您可以授权MaxCompute投递作业(新版)使用自定义RAM角色读取源Logstore中的数据,即您在创建 MaxCompute投递作业(新版)时,选择**读日志服务授权**为**自定义角色**,然后输入自定义RAM角色的ARN(例如 acs:ram::10\*\*\*\*12:role/maxcomputeshiprole)。如何获取ARN,请参见查看RAM角色基本信息。

* 读日志服务授权:		
	acs:ram::10 12:role/maxcomputeshiprole	

## 5.3.5. 通过默认角色写数据到MaxCompute

MaxCompute投递作业(新版)运行时,需要将读取到的Logstore数据投递到MaxCompute表中,您可以使用默认角色完成该操作。本文介绍如何对默认角色进行授权。

#### 前提条件

如果您使用的是RAM用户,请确保该用户具备操作RAM角色授权的权限。

#### 同账号投递

如果您的日志服务和MaxCompute属于同一个阿里云账号,则使用该账号完成<mark>云资源访问授权</mark>后,还需添加 AliyunLogDefaultRole角色为MaxCompute工作空间成员,具体操作如下:

您可以通过可视化界面或命令行进行RAM角色授权。

#### 通过可视化界面完成授权

- 1. 修改RAM角色信任策略。
  - i. 登录RAM控制台。
  - ii. 在左侧导航栏中,选择**身份管理 > 角色**。
  - iii. 在RAM角色列表中,单击AliyunLogDefault Role角色。
  - iv. 在信任策略管理页签下,单击修改信任策略。
  - v. 在修改信任策略面板中, 将原有的信任策略替换为如下内容, 然后单击确定。

```
{
   "Statement": [
   {
      "Action": "sts:AssumeRole",
      "Effect": "Allow",
      "Principal": {
           "Service": [
            "log.aliyuncs.com",
            "dataworks.aliyuncs.com"
        ]
     }
   }
   }
   /
   //
   //
   //
   //
   //
   //
   //
}
```

- 2. 添加RAM角色为工作空间成员。
  - i. 登录MaxCompute控制台。
  - ii. 在页面左上角,选择地域。
  - iii. 在项目管理页签中,单击目标项目对应的成员管理。

iv. 在成员管理页面,单击添加成员。

ø	≡ 工作空间配置	成员管理 角色管	里						添加成员 查看	涌色权限
巫	成员管理	请输入成员名或云账号	Q	自定义角色: 请证	选择	~				
0	权限列表	全部数据分析师	0 项目所有者 1	空间管理员 1 开	FF发 0 运维 0	部署 0 访客	) 安全管理员 0	模型设计师 0		
~	MaxCompute高级配置	成员		云账号		角色			加入时间	操作
۲	数据源管理	C	vunid.com	CH. (1995)(1995)(1995)	nid.com	空间管理员 ×		~	2020-09-22 10:33:19	所有著

v. 在**添加成员**对话框中,选中当前登录账号和AliyunLogDefaultRole角色,根据页面提示完成添加。 此处需在**批量设置角色**中选中**开发**。更多信息,请参见<mark>添加工作空间成员并设置角色</mark>。

- 3. 授予AliyunLogDefaultRole角色操作MaxCompute表的权限。
  - i. 在左侧导航栏中, 单击MaxCompute高级配置。
  - ii. 设置MaxCompute项目选择为生产环境下的目标项目,然后单击自定义用户角色。
  - iii. 在角色列表中,单击role\_project\_admin角色对应的成员管理。

0	= 工作空间配置	MaxCompute 项目选择:	生产环境definelline		
-	成员管理	基本设置			新増角色
0	权限列表	自定义用户角色			
~	MaxCompute高级配置		输入角色名称进行搜索		
۲	数据源管理		角色名称	角色类型	操作
۲	扩展程序配置		admin	project	查看详情 成员管理
	开源集群管理		role_project_admin	project	查看详情 成员管理 权限管理

- iv. 在成员管理对话框中,选中当前登录账号和AliyunLogDefaultRole角色账号,根据页面提示完成添加。
- v. 单击role\_project\_admin角色对应的权限管理。
- vi. 在角色授权对话框的表页签下,单击添加表,然后选择目标MaxCompute表,选中Describe、Alter和Update。

○ 注意 上述授权只针对指定的MaxCompute表。如果您希望目标角色可以操作当前MaxCompute空间项目下所有的表,则您可以为当前登录账号和AliyunLogDefaultRole角色账号添加admin角色权限。即在角色列表中, 单击admin角色对应的成员管理,然后在成员管理对话框中,选中当前登录账号和AliyunLogDefaultRole角色 账号,根据页面提示完成添加。

#### 通过命令行完成授权

- 1. 登录MaxCompute控制台。
- 2. 在页面左上角,选择地域。
- 3. 在项目管理页签中,单击目标项目对应的数据开发。
- 4. 新建业务流程。
  - i. 在数据开发页面,选择新建 > 新建业务流程。
  - ii. 在新建业务流程对话框中,设置业务名称,然后单击新建。
- 5. 新建节点。
  - i. 在数据开发页面,选择新建 > 新建节点 > MaxCompute > ODPS SQL。
  - ii. 在新建节点对话框中,设置节点名称和路径,然后单击提交。
    - 其中, 路径需设置为您在步骤中所创建的业务流程。
- 6. 在已创建的节点编辑框中,执行如下命令,完成授权。

odps sql author:calive:calive	*********** ltrole`; RAM\$cwaliyunid.com:`role/aliyunlogdefaultrole`; ogdefaultrole`;
命令	说明
USE <i>project-name</i> ;	指定MaxCompute项目,需与您在创建MaxCompute投递作 业(新版)时设置的MaxCompute项目保持一致。更多信息, 请参见 <mark>创建MaxCompute投递作业(新版)</mark> 。
ADD USER RAM\$ ****.aliyunid.com:`role/aliyunlogdef aultrole`;	在MaxCompute项目中添加用户。 • ****.aliyunid.com为MaxCompute项目所属的阿里云账 号。您可以通过 list users; 命令查看对应的阿里云 账号。 • aliyunlogdefaultrole为AliyunLogDefaultRole角色的名 称,必须为小写格式。
GRANT DESCRIBE, ALTER, UPDATE ON TABLE <b>table-na</b> <i>me</i> TO user RAM\$****. <i>aliyunid.com</i> :`role/aliyunlogde faultrole`;	授予用户查看、修改、更新目标MaxCompute表的权限。 ⑦ 说明 该授权只针对指定的MaxCompute表,如果 您希望目标用户可以操作当前MaxCompute项目下的所 有表,请执行 GRANT admin to user RAM\$****.aliyu nid.com:`role/aliyunlogdefaultrole`; 命令完 成授权。
SHOW GRANTS FOR `RAM\$ <i>****.aliyunid.com</i> :role/aliyunlogdefaultrole`;	确认授权是否成功。

#### 如果返回如下类似信息表示授权成功。

Authorization Type: ACL [user/RAM\$\*\*\*\*.aliyunid.com:role/aliyunlogdefaultrole] A projects/default\_project\_\*\*\*\*/tables/\*\*\*\*: Describe | Alter | Update

#### 在授权过程中,可能发生如下报错。

- • 在操作过程中出现 FAILED: ODPS-0130013:Authorization exception Authorization Failed [4003], You ha
   ve NO privilege to do the PROJECT SECURITY OPERATION for {acs:odps:\*:projects/xxxxx/authorization/u
   sers}. Context ID:1111-11111-11111. 错误,表示当前的操作用户无MaxCompute账户操作或者授
   权权限。更多信息,请参见出现操作用户无权限错误时,如何解决。

完成上述操作后,您可以授权MaxCompute投递作业(新版)使用AliyunLogDefaultRole角色将数据投递到MaxCompute 表中。即您在创建MaxCompute投递作业(新版)时,选择**写MaxCompute授权**为**默认角色**。

写入授权方式:	角色密钥		
* 写MaxCompute授权:			
	acs:ram::13 44:role/aliyunlogdefaultrole		

#### 跨账号投递
如果您的日志服务和MaxCompute不属于同一个阿里云账号,例如日志服务属于账号A,MaxCompute属于账号B,则两个 账号分别完成<mark>云资源访问授权</mark>后,还需进行如下操作。

- 1. 修改账号B下的AliyunLogDefaultRole角色的信任策略。
  - i. 使用账号B登录RAM 控制台。
  - ii. 在左侧导航栏中,选择**身份管理 > 角色**。
  - iii. 在RAM角色列表中, 单击AliyunLogDefault Role。
  - iv. 在信任策略管理页签中, 单击修改信任策略。
  - v. 在修改信任策略面板中,修改信任策略,然后单击确定。

在Service配置项中添加**阿里云账号A的ID@log.aliyuncs.com**和**dataworks.aliyuncs.com**。其中**阿里云账号** A的ID,需根据实际情况替换,您可以在账号中心查看阿里云账号ID。

该策略表示账号A有权限通过日志服务获取临时Token来操作账号B中的资源。

{
"Statement": [
{
"Action": "sts:AssumeRole",
"Effect": "Allow",
"Principal": {
"Service": [
"账号A对应的阿里云账号ID@log.aliyuncs.com",
"dataworks.aliyuncs.com",
"log.aliyuncs.com"
]
}
}
],
"Version": "1"
}

2. 添加账号B下的AliyunLogDefaultRole角色为MaxCompute工作空间成员。

您可以通过可视化界面或命令行进行RAM角色授权,即使用账号B登录MaxCompute控制台,完成授权操作。具体操作,请参见通过可视化界面完成授权或通过命令行完成授权。

完成上述操作后,您可以授权MaxCompute投递作业(新版)使用账号B下的AliyunLogDefaultRole角色将数据投递到 MaxCompute表中。即您在创建MaxCompute投递作业(新版)时,选择写MaxCompute授权为自定义角色,然后输入 账号B下的AliyunLogDefaultRole角色的ARN(例如acs:ram::11\*\*13:role/aliyunlogdefaultrole)。如何获取ARN,请参 见查看RAM角色基本信息。

写入授权方式:	角色	密钥	
* 写MaxCompute授权:	() 默认	角色	● 自定义角色
	acs:rai	m::114	13:role/aliyunlogdefaultrole

## 5.3.6. 通过自定义角色写数据到MaxCompute

MaxCompute投递作业(新版)运行时,需要将读取到的Logstore中数据投递到MaxCompute的表中,您可以使用自定义角色完成该操作。本文介绍如何对自定义角色进行授权。

### 前提条件

- 已创建RAM角色(例如MaxComputeShipRole)。具体操作,请参见创建RAM角色。
- 如果您使用的是RAM用户,请确保该用户具备操作RAM角色授权的权限。

### 同账号投递

授予RAM角色写MaxCompute的权限后,MaxCompute投递作业可以使用该角色将Logstore中的数据写入到目标 MaxCompute的表中。此处通过添加RAM角色为MaxCompute工作空间成员的方式完成授权。

您可以通过可视化界面或命令行进行RAM角色授权。

### 通过可视化界面完成授权

- 1. 修改RAM角色信任策略。
  - i. 登录RAM控制台。
  - ii. 在左侧导航栏中,选择**身份管理>角色**。
  - iii. 在RAM角色列表中, 单击目标RAM角色。
  - iv. 在信任策略管理页签下,单击修改信任策略。
  - v. 在修改信任策略面板中,将原有的信任策略替换为如下内容,然后单击确定。

```
{
  "Statement": [
   {
      "Action": "sts:AssumeRole",
      "Effect": "Allow",
      "Principal": {
           "Service": [
            "log.aliyuncs.com",
            "dataworks.aliyuncs.com"
           ]
        }
    }
    ,
    "Version": "1"
}
```

- 2. 添加RAM角色为工作空间成员。
  - i. 登录MaxCompute控制台。
  - ii. 在页面左上角,选择地域。
  - iii. 在**项目管理**页签中,单击目标项目对应的成员管理。
  - iv. 在成员管理页面, 单击添加成员。

ø	≡ 工作空间配置	成员管	理 角色管理									添加成员 查	角色权限
巫	成员管理	请输入质	成员名或云账号	Q	自定义角色:	请选择		~					
9	权限列表	全部	数据分析师 0	项目所有者 1	空间管理员 1	开发 0 〕	运维 0	部署 0	访客 0	安全管理员 0	模型设计师 0		
~	MaxCompute高级配置		成员		云账号			角色				加入时间	操作
٢	数据源管理		C	vunid.com	COLUMN DESIGN	nid.com		空间管理	员×		~	2020-09-22 10:33:19	所有著

v. 在**添加成员**对话框中,选中当前登录账号和目标RAM角色,根据页面提示完成添加。

此处需在批量设置角色中选中开发。更多信息,请参见添加工作空间成员并设置角色。

- 3. 授予RAM角色操作MaxCompute表的权限。
  - i. 在左侧导航栏中, 单击MaxCompute高级配置。
  - ii. 设置MaxCompute项目选择为生产环境下的目标项目,然后单击自定义用户角色。

iii. 在角色列表中,单击role\_project\_admin角色对应的成员管理。

ø	三	MaxCompute 项目选择:	生产环境def		
-	成员管理	基本设置	白定义田户备色		新增角色
9	权限列表	自定义用户角色			
N.	MaxCompute高级配置		输入角色名称进行搜索		
٢	数据源管理		角色名称	角色类型	操作
۲	扩展程序配置		admin	project	查看详情 成员管理
	开源集群管理		role_project_admin	project	查看详情 成员管理 权限管理

- iv. 在成员管理对话框中,选中当前登录账号和目标RAM角色账号,根据页面提示完成添加。
- v. 在角色授权对话框的表页签下,单击添加表,然后选择目标MaxCompute表,选中Describe、Alter和Update。

○ 注意 上述授权只针对指定的MaxCompute表。如果您希望目标角色可以操作当前MaxCompute空间项目下所有的表,则您可以为当前登录账号和目标RAM角色账号添加admin角色权限。即在角色列表中,单击admin角色对应的成员管理,然后在成员管理对话框中,选中当前登录账号和目标RAM角色账号,根据页面提示完成添加。

### 通过命令行完成授权

- 1. 登录MaxCompute控制台。
- 2. 在页面左上角,选择地域。
- 3. 在项目管理页签中,单击目标项目对应的数据开发。
- 4. 新建业务流程。
  - i. 在数据开发页面,选择新建 > 新建业务流程。
  - ii. 在新建业务流程对话框中,设置业务名称,然后单击新建。
- 5. 新建节点。
  - i. 在数据开发页面,选择新建 > 新建节点 > MaxCompute > ODPS SQL。
  - ii. 在新建节点对话框中,设置节点名称和路径,然后单击提交。
    - 其中,路径需配置为您在步骤中所创建的业务流程。
- 6. 在已创建的节点编辑框中,执行如下命令,完成授权。

odps sql author:c .aliyunid.com create time:2022-05-23 15:55:50 USE ADD USER RAM\$cwaliyunid.com;`role/maxcomputesh GRANT DESCRIBE, ALTER, UPDATE ON TABLE e to user SHOW GRANTS FOR `RAM\$cwaliyunid.com:role/maxco	hiprole`; AMM\$cwaliyunid.com:`role/maxcomputeshiprole`; pmputeshiprole`;
命令	说明
USE <i>project-name;</i>	指定MaxCompute项目,需与您在创建MaxCompute投递作 业(新版)时设置的MaxCompute项目保持一致。更多信息, 请参见 <mark>创建MaxCompute投递作业(新版)</mark> 。
ADD USER RAM\$****.aliyunid.com:`role/maxcomputes	在MaxCompute项目空间中添加用户。 o ****.aliyunid.com为MaxCompute项目所属的阿里云账 号。您可以通过 list users; 命令查看对应的阿里云 账号。 o maxcomputeshiprole为自定义RAM角色的名称,必须为小 写格式。

命令	说明		
	授予用户查看、修改、更新目标MaxCompute表的权限。		
GRANT DESCRIBE, ALTER, UPDATE ON TABLE <i>table-na</i> <i>me</i> to user RAM\$ ****. <i>aliyunid.com</i> :`role/ <i>maxcompute</i> <i>shiprole</i> `;	⑦ 说明 该授权只针对指定的MaxCompute表,如果 您希望目标用户可以操作当前MaxCompute项目下的所 有表,请执行 GRANT admin to user RAM\$****.aliyu nid.com:`role/maxcomputeshiprole`; 命令完成授 权。		
SHOW GRANTS FOR `RAM\$****.aliyunid.com:role/maxc omputeshiprole`;	确认授权是否成功。		

如果返回如下类似信息表示授权成功。

```
Authorization Type: ACL
[user/RAM$****.aliyunid.com:role/maxcomputeshiprole]
A projects/default_project_****/tables/****: Describe | Alter | Update
```

在授权过程中,可能发生如下报错。

- 在操作过程中出现 FAILED: ODPS-0130013:Authorization exception Authorization Failed [4003], You ha
   ve NO privilege to do the PROJECT SECURITY OPERATION for {acs:odps:\*:projects/xxxxx/authorization/u
   sers}. Context ID:1111-1111-1111-1111. 错误,表示当前的操作用户无MaxCompute账户操作或者授
   权权限。更多信息,请参见出现操作用户无权限错误时,如何解决。

完成上述操作后,您可以授权MaxCompute投递作业(新版)使用自定义RAM角色将数据投递到MaxCompute表中。即您 在创建MaxCompute投递作业(新版)时,选择**写MaxCompute授权**为**自定义角色**,然后输入自定义RAM角色的 ARN(例如acs:ram::10\*\*12:role/maxcomputeshiprole)。如何获取ARN,请参见查看RAM角色基本信息。

写入授权方式:	角色 密钥
* 写MaxCompute授权:	
	acs:ram::10-0012:role/maxcomputeshiprole

### 跨账号投递

如果您的日志服务和MaxCompute不属于同一个阿里云账号,例如日志服务属于账号A,MaxCompute属于账号B,则需授 予账号B下的RAM角色b写MaxCompute的权限。授权完成后,MaxCompute投递作业可以使用该RAM角色将Logstore中的 数据写入到目标MaxCompute表中。

1. 修改账号B下RAM角色b的信任策略。

- i. 使用账号B登录RAM控制台。
- ii. 在左侧导航栏中,选择**身份管理 > 角色**。
- iii. 在RAM角色列表中,单击RAM角色b。
- iv. 在信任策略管理页签中, 单击修改信任策略。

v. 修改信任策略, 然后单击确定。

在Service配置项中添加**阿里云账号A的ID@log.aliyuncs.com**和**dataworks.aliyuncs.com**。其中**阿里云账号** A**的ID**,需根据实际情况替换,您可以在<u>账号中心</u>查看阿里云账号ID。

该策略表示账号A有权限通过日志服务获取临时Token来操作账号B中的资源。

```
ł
   "Statement": [
       {
           "Action": "sts:AssumeRole",
           "Effect": "Allow",
            "Principal": {
                "Service": [
                    "log.aliyuncs.com",
                    "dataworks.aliyuncs.com",
                    "阿里云账号A的ID@log.aliyuncs.com"
               1
            }
        }
   1,
   "Version": "1"
}
```

2. 添加账号B下的RAM角色b为MaxCompute工作空间成员。

您可以通过可视化界面或命令行进行RAM角色授权,即使用账号B登录MaxCompute控制台,完成授权操作。具体操作,请参见通过可视化界面完成授权或通过命令行完成授权。

完成上述操作后,您可以授权MaxCompute投递作业(新版)使用账号B下的自定义RAM角色将数据投递到MaxCompute 表中。即您在创建MaxCompute投递作业(新版)时,选择**写MaxCompute授权**为**自定义角色**,然后输入账号B下的自 定义RAM角色的ARN(例如acs:ram::11\*\*13:role/maxcomputeshiprole)。如何获取ARN,请参见查看RAM角色基本信息。

写入授权方式:	角色密钥
* 写MaxCompute授权:	○ 默认角色 (●) 自定义角色
	acs:ram::11 13:role/maxcomputeshiprole

## 5.3.7. 通过RAM用户写数据到MaxCompute

MaxCompute投递作业(新版)运行时,需要将读取到的Logstore数据投递到MaxCompute表中,您可以使用RAM用户完成该操作。本文介绍如何对RAM用户进行授权。

### 前提条件

已创建RAM用户。具体操作,请参见创建RAM用户。

### 操作步骤

- 1. 添加RAM用户为MaxCompute工作空间成员。
  - i. 登录MaxCompute控制台。
  - ii. 在页面左上角,选择地域。
  - iii. 在项目管理页签中,单击目标项目对应的成员管理。
  - iv. 在成员管理页面, 单击添加成员。
  - v. 在添加成员对话框中,选中目标RAM用户,根据页面提示完成添加。

此处需在**批量设置角色**中选中开发。更多信息,请参见添加工作空间成员并设置角色。

2. 创建RAM用户的AccessKey。

#### 具体操作,请参见为RAM用户创建访问密钥。

□ 注意

- RAM用户的AccessKey Secret只在创建时显示,不支持查询,请妥善保管。
- 若AccessKey泄露或丢失,则需要创建新的AccessKey,最多可以创建2个AccessKey。

完成上述操作后,您可以授权MaxCompute投递作业(新版)使用RAM用户将数据投递到MaxCompute表中。即您在创建 MaxCompute投递作业(新版)时,选择**写入授权方式**为密钥,然后输入RAM用户的AccessKey。

写入授权方式:	角色密钥
AccessKey ID:	LTA
AccessKey Secret:	Y4

## 5.4. 创建MaxCompute投递作业(新版)

日志服务采集到日志后,支持将日志投递至MaxCompute的表中进行存储与分析。本文介绍如何创建MaxCompute投递作业(新版)。

### 前提条件

- 已创建Project和Logstore。具体操作,请参见创建Project和Logstore。
- 已采集日志。具体操作,请参见数据采集。
- 已在MaxCompute中创建表。更多信息,请参见创建表。

### 注意事项

↓ 注意 目前仅限华北2(北京)地域的白名单用户使用MaxCompute投递(新版)。如果您是华北2(北京)地域的非白名单用户,请提交工单申请。

- 对于char类型或varchar类型的字段,如果其值长度超过指定长度,那么投递到MaxCompute后,超过部分会被截断。
   例如长度限制为3,字段值为012345,则投递到MaxCompute后,值为012。
- 对于string类型、char类型或varchar类型的字段,如果其值为空字符串,那么投递到MaxCompute后,值为Null。
- 对于datetime类型的字段,其值格式必须为YYYY-MM-DD HH:mm:ss(DD和HH之间可以存在多个空格)。字段值格式 错误时,不会导致投递错误,但投递到MaxCompute后,值为Null。
- 对于date类型的字段,其值格式错误时,不会导致投递错误,但投递到MaxCompute后,值为Null。
- 对于decimal类型的字段,如果其值中的小数位长度超过指定长度,会被四舍五入截断;如果整数位超过指定长度,系统会将整条日志作为脏数据丢弃,并增加错误计数。
- 投递过程中,默认丢弃脏数据。
- 对于日志中不存在的值,投递到MaxCompute后,可能为默认值或Null。
  - 如果创建MaxCompute表时指定了默认值,那么投递到MaxCompute后,值为默认值。
  - 如果创建MaxCompute表时未指定默认值,但是允许值为Null,那么投递到MaxCompute后,值为Null。
- 由于MaxCompute的限制,最大投递并发数为64,即最多支持64个并发同时写MaxCompute。MaxCompute单分区的 最大流量为10 MB/s。

#### 操作步骤

1.

2. 在Project列表区域,单击目标Project。

- 3. 在日志存储 > 日志库页签中,单击目标Logstore左侧的>,选择数据处理 > 导出 > MaxCompute(原ODPS)。
- 4. 将鼠标悬浮在MaxCompute (原ODPS)上, 单击+。
- 5. 在MaxCompute投递功能面板中,配置如下参数,然后单击确定。

选择投递版本为新版(推荐),重要参数配置说明如下所示。

参数	说明
投递名称	投递任务的名称。命名规则如下: • 只能包含小写字母、数字、下划线(_)和短划线(-)。 • 必须以小写字母或者数字开头和结尾。 • 长度必须在3~63字符之间。
投递区域	目标MaxCompute表所在地域。
	送需要使用公网,可能存在网络不稳定的风险,请谨慎使用。
MaxCompute Endpoint	MaxCompute地域对应的Endpoint。更多信息,请参见 <mark>Endpoint</mark>
Tunnel Endpoint	MaxCompute地域对应的Tunnel Endpoint。更多信息,请参见Endpoint。
项目名	目标MaxCompute表所在的MaxCompute项目。
MaxCompute表名	MaxCompute表名称。
读日志服务授权	<ul> <li>授予MaxCompute投递作业读取Logstore数据的权限。</li> <li>默认角色:授权MaxCompute投递作业使用阿里云系统角色AliyunLogDefaultRole来读取Logstore中的数据。更多信息,请参见通过默认角色读取Logstore数据。</li> <li>自定义角色:授权MaxCompute投递作业使用自定义角色来读取Logstore中的数据。</li> <li>您需先授予自定义角色读取Logstore数据的权限,然后在读日志服务授权中输入您自定义角色的ARN。更多信息,请参见通过自定义角色读取Logstore数据。</li> </ul>
写入授权方式	您可以通过RAM用户的AccessKey或RAM角色,完成写MaxCompute的授权。
写MaxCompute授权	授予MaxCompute投递作业将数据写入到MaxCompute表中的权限。 • 默认角色: 授权MaxCompute投递作业使用阿里云系统角色AliyunLogDefaultRole将数 据写入到MaxCompute表中。更多信息,请参见通过默认角色写数据到MaxCompute。 • 自定义角色: 授权MaxCompute投递作业使用自定义角色将数据写入到MaxCompute表 中。 您需先授予自定义角色将数据写入到MaxCompute表中的权限,然后在写MaxCompute 授权中输入您自定义角色的ARN。 • 如果日志服务和MaxCompute属于同一阿里云账号,请参见同账号投递。 • 如果日志服务和MaxCompute属于不同的阿里云账号,请参见跨账号投递。 • RAM用户的AccessKey: 授权MaxCompute投递作业使用RAM用户将数据写入到 MaxCompute表中。更多信息,请参见通过RAM用户写数据到MaxCompute。

参数	说明
	左边输入框中填写与MaxCompute表列相映射的日志字段名称,右边为MaxCompute表的列 名称。更多信息,请参见 <mark>数据模型映射</mark> 。
MaxCompute普通列	<ul> <li>✔ 注意</li> <li>● 日志服务投递日志到MaxCompute按照日志字段与MaxCompute表列的顺序进行映射,修改MaxCompute表列名不影响数据投递。如果更改MaxCompute表的Schema,请重新配置日志字段与MaxCompute表列映射关系。</li> <li>● 左边输入框的日志字段,不支持双引号("")、单引号(")和含有空格的字符串。</li> <li>● 如果您的日志中存在同名字段(例如都为request_time),则日志服务会将其中一个字段名显示为request_time_0,底层存储的字段名仍为request_time。因此您在投递时,只能使用原始字段名request_time。</li> <li>存在同名字段时,系统只随机投递其中一个字段的值。请尽量避免日志中使用同名字段。</li> </ul>
MaxCompute分区列	左边输入框中填写与MaxCompute表分区列相映射的日志字段名称,右边为MaxCompute表 分区列名称。更多信息,请参见数据模型映射。 ⑦ 说明 • 最大配置3个分区列,请谨慎选择自定义字段作为分区列,保证一次投递任务中 生成的分区数目小于512个,否则会导致投递任务写数据到MaxCompute表失 败,整批数据无法投递。 • 分区字段不支持extract_others、extract_others_all。
时间分区格式	时间分区格式,配置示例和参数详细请参见参考信息和Java SimpleDateFormat。 ⑦ 说明 <ul> <li>仅当MaxCompute分区列中的字段配置为partition_time时,时间分区格式才生效。</li> <li>请勿使用精确到秒的日期格式,易导致单表的分区数目超过限制(60000个)。</li> </ul>
时区选择	该时区用于格式化时间以及时间分区。更多信息,请参见 <mark>时区列表</mark> 。
投递模式	支持实时投递和批投递。 • <b>实时投递</b> :即时读取Logstore中的数据,并投递到MaxCompute。 • <b>批投递</b> :读取Logstore中早于当前时间5分钟~10分钟之间的数据,并投递到 MaxCompute中。 更多信息,请参见投递模式说明。
开始时间范围	投递作业从该时间开始拉取Logstore中的数据。

创建投递作业后,一般情况下日志数据会在写入Logstore后的1个小时导入到MaxCompute,导入成功后即可在 MaxCompute内查看到相关数据。更多信息,请参见日志投递MaxCompute后,如何检查数据完整性。

## 数据模型映射

将日志服务中的日志投递到MaxCompute表中时,涉及两个服务之间的数据模型映射问题,相关注意事项与示例如下所示。

- MaxCompute表至少包含一个数据列和一个分区列。
- 日志服务保留字段建议使用\_\_partition\_time\_\_、\_\_source\_\_、\_\_topic\_\_。
- 一个MaxCompute表的分区数最大值为60000个,当分区数超出最大值后无法再写入数据。
- 系统保留字段\_\_extract\_others\_\_有曾用名\_extract\_others\_, 可兼容使用。
- MaxCompute分区列的值不支持配置为MaxCompute的保留字和关键字。更多信息,请参见保留字与关键字。
- MaxCompute分区列取值不支持配置为空,所以映射到分区列的字段必须为保留字段或日志字段,且可以通过cast运算 符将string类型字段值转换为对应分区列类型,空分区列的日志会在投递中被丢弃。
- 日志服务中一个日志字段只能映射到一个MaxCompute表的列(数据列或分区列),不支持字段冗余。

MaxCompute数据列、分区列与日志服务字段的映射关系示例如下所示,其中日志服务保留字段详情请参见保留字段。

MaxCompute列 类型	列名 (MaxCompute )	数据类型 (MaxCompute )	日志字段名称 (日志服务)	字段类型(日志 服务)	字段说明
	log_source	string	source	保留字段	日志来源。
	log_time	bigint	time	保留字段	日志时间,Unix时间戳格 式,对应数据模型中的 Time域。
	log_topic	string	topic	保留字段	日志主题。
数据列	time	string	time	日志内容字段	解析自日志,对应数据模型 中的key-value。在很多时 候Logtail采集的数据的 time与time取值相 同。
	ip	string	ip	日志内容字段	解析自日志。
	thread	string	thread	日志内容字段	解析自日志。
	log_extract_ot hers	string	extract_other s	保留字段	未在配置中进行映射的其他 日志字段会通过key-value 序列化到JSON中,该JSON 是一层结构,不支持字段内 部JSON嵌套。
	log_partition_ti me	string	partition_tim e	保留字段	由日志中的time字段 对齐计算而得,分区粒度可 配置。

分区列 MaxCompute列 类型	列名 (MaxCompute )	数据类型 (MaxCompute )	日志字段名称 (日志服务)	字段类型(日志 服务)	字段说明
	status	string	status	日志内容字段	解析自日志,该字段取值支 持枚举,保证分区数目不超 过上限。

### 投递模式说明

目前, MaxCompute投递(新版)支持实时投递和批投递两种模式。

- 实时投递:即时读取Logstore中的数据,投递到MaxCompute。
- 批投递:读取Logstore中早于当前时间5分钟~10分钟之间的数据,并投递到MaxCompute中。

设置**投递模式为批投递**后,如果您要设置**开始时间范围中的起始时间或结束时间**,则必须按照5分钟对齐。例如 202 2-05-24 16:35:00 是正确设置, 2022-05-24 16:36:00 为非法设置。

* 投递模式	○ 实时 ● 批投递	
	如果是批役递只能选取对齐5分钟的时间。比如12:05是合法的,而12:06是非法的 (开始时间和结束时间均是如此)。模式的详细信息,请参考帮助文档。	
* 开始时间范围:	○ 所有 ○ 某时间开始 ● 特定时间范围	_
* 起始时间:	2022-05-24 16:35:00	
* 结束时间:	2022-05-24 16:45:00 節	

另外,批投递支持投递\_\_unique\_id\_\_字段和\_\_receive\_time\_\_字段。

○ \_\_unique\_id\_\_字段值为一个64位的字符串,用于唯一标识一条日志。

如果您要投递该字段,只能在MaxCompute普通列中添加该字段。

○ \_\_receive\_time\_\_字段表示日志被日志服务接收的时间,您可通过时间分区格式设置其格式,最大精确到半小时。关于时间分区格式的说明,请参见参考信息。

如果您要投递该字段,只能在MaxCompute分区列中添加该字段。

### 参考信息

• \_\_partition\_time\_\_字段

将日志时间作为分区字段,通过时间筛选数据是MaxCompute常见的过滤数据的方法。

。 格式

\_\_partition\_time\_\_是根据日志服务中\_\_time\_\_字段的值计算得到的,结合时区配置以及分区时间格式,生成时间字符串。为避免触发MaxCompute单表分区数目的限制,日期分区列的值按照1800秒(半小时)对齐。

例如:日志服务的日志时间为27/Jan/2022 20:50:13 +0800,日志服务据此计算出保留字段\_time\_为1643287813(Unix时间戳),不同配置下的时间分区列取值如下所示。

分区时间格式	partition_time
%Y_%m_%d_%H_%M_00	2022_01_27_20_30_00
%Y_%m_%d_%H_%M	2022_01_27_20_30
%Y%m%d	20220127

### ○ 使用方法

使用\_\_partition\_time\_\_ 筛选数据,可以避免全表扫描。例如查询2022年1月26日一天内日志数据,查询语句如下所示。

select \* from {ODPS\_TABLE\_NAME} where log\_partition\_time >= "2022\_01\_26" and log\_partition\_time < "
2022\_01\_27";</pre>

- \_\_extract\_others\_\_字段和\_\_extract\_others\_all\_\_字段
  - \_\_extract\_others\_\_字段中包含日志字段中未映射的所有字段(不包括\_\_topic\_\_、\_\_tag\_\_:\*和\_\_source\_\_)。
  - \_\_extract\_others\_all\_\_字段中包含日志字段中未映射的所有字段(包括\_\_topic\_\_、\_\_tag\_\_:\*和\_\_source\_\_)。

## 5.5. 为MaxCompute投递作业(新版)设置告警

日志服务已内置数据投递监控规则模板,您只需添加对应的告警实例即可实时监控MaxCompute投递作业,并可通过钉钉 等渠道接收到告警通知。本文介绍设置告警的相关操作。

#### 前提条件

已创建MaxCompute投递作业。具体操作,请参见创建MaxCompute投递作业(新版)。

## 步骤一:开启作业运行日志

- 1. 登录日志服务控制台。
- 在Project列表区域,单击目标Project。
   该Project为MaxCompute投递作业所在的Project。
- 3. 在页面左上方, 单击 🞧 图标。
- 4. 进入开通作业运行日志页面。
  - 如果您开未通过该Project的详细日志,则在服务日志页签中,单击开通服务日志。
  - 如果您已开通过该Project的详细日志,则在服务日志页签中,单击 
     資图标。
- 5. 设置如下参数,然后单击确定。

参数	说明
作业运行日志	打开 <b>作业运行日志</b> 开关后,系统将在您指定的Project中自动创建一个名为internal- diagnostic_log的Logstore,用于存储Scheduled SQL、MaxCompute投递、OSS投递、 数据导入等作业的运行日志与错误日志。日志字段说明,请参见数据导入、投递等作业的 运行日志。
日志存储位置	开通作业运行日志功能后,需要选择日志的存储位置,即需要指定Project。可以设置为: <ul> <li>自动创建(推荐)。</li> <li>当前Project。</li> <li>同一地域下的其他Project。</li> </ul>

开通作业运行日志后,您可以在指定Project下的internal-diagnostic\_log Logstore中查看MaxCompute投递作业的运行日志与错误日志,其日志主题(\_\_topic\_\_)为etl\_metrics。您也可以通过投递作业名称查询目标投递作业的运行日志与错误日志,对应的查询语句为 job\_name:*f业名称*,例如 job\_name:job-1646\*\*\*946 。

### 步骤二:配置行动策略

1. 登录日志服务控制台。

- 2. 进入行动策略管理页面。
  - i. 在Project列表区域,单击任意的Project。

- ii. 在左侧导航栏中, 单击**告**警。
- iii. 选择告警管理 > 行动策略。
- 3. 找到目标行动策略(sls.app.export.builtin),单击修改。

您也可以创建新的行动策略用于告警通知。具体操作,请参见创建行动策略。

4. 在**编辑行动策略**面板的**第一行动列表**页签中,将**请求地址**修改为钉钉群机器人的Webhook地址。其他选项,保持 默认配置。

如何获取钉钉群机器人的WebHook地址,请参见<mark>钉钉-自定义</mark>。您也可以根据业务需求,使用其他告警渠道。具体操 作,请参见<mark>通知渠道说明</mark>。

5. 单击确认。

### 步骤三:添加告警实例

日志服务已内置5种监控规则模板。更多信息,请参见<mark>数据投递监控规则</mark>。您只需根据业务需求,添加对应的告警实例即 可。此处以添加数据投递延迟监控规则对应的告警实例为例。

- 1. 登录日志服务控制台。
- 2. 在Project列表区域,单击目标Project。

该Project为您internal-diagnostic\_log Logstore所在的Project。

- 3. 在左侧导航栏中,单击告警。
- 4. 在规则/事务页签中,单击SLS数据投递。

规则/事	規則/事务 告警管理 ∨ ⑧・					
P 来源: ⑦ 状态:	▶ 注意: □ 自立文告号 (田板) (0) □ 自立文告号 (新版) (0) □ 系统内置告号(19) □ (本本本本本本本本本本本本本本本本本本本本本本本本本本本本本本本本本本本					
<b>器</b> 类别:	■ 产品 S SLS数据加工(5) S SLS数据投递(5) S SLS Logtail(8) [	SLS消費担(1) 全部送別 ∨				
新建告警	开启 关闭 临时关闭 恢复 开级 腳除		C test-lv V	请选择 > 输入模版ID或模版描述筛选 Q		
	监控规则	类别	状态	操作		
- +	- 数据投递延迟监控 ⑦	内置告誓 SLS数据投递 投递延迟监控 展开	● 未创建	添加 开启		
- +	- 数要投递流量(绝对值)监控 ①	(内置背容) SLS数据投递) 投递流量监控 展开	● 未创建	添加 开启		
- +	- 数据投递异常反情监控 ⑦	(内置告督) (SLS数据投递) 投递报错监控) 展开	● 未创建	漆加 开启		
- +	- 数据投递失败条数监控 ⑦	(内置告誓) SLS数据投递) 投递失败监控 展开	● 未创建	添加 开启		
- +	- 数据投递流量(日同比)监控 ③	(内置告答) (SLS数据投递) (投递流量监控) 展开	● 未创建	添加 开启		

- 5. 在规则列表中, 单击数据投递延迟监控对应的添加。
- 6. 在参数设置对话框中, 配置监控规则, 然后单击设置并开启。

参数	说明
告警名称	告警名称,支持自定义。
监控的投递作业名称	需要监控的MaxCompute投递作业的名称。 • 默认值为 .* ,表示监控当前阿里云账号下所有的MaxCompute投递作业。 • 多个投递作业之间可以使用竖线( )分隔。您还可以使用正则表达式 .* 进行配置, 例如sche.*,表示监控以sche开头的投递作业。
监控阈值	当MaxCompute投递作业的延迟时间超过该值时,触发告警。默认值为300秒。
行动策略	当前告警所绑定的行动策略,日志服务将通过该行动策略给指定用户发送告警通知。 默认为 sls.app.export.builtin (SLS数据投递内置行动策略)。您也可以自定义 行动策略。具体操作,请参见 <mark>创建行动策略</mark> 。

参数	说明
严重度	定义告警消息的严重度。
静默期	设置告警静默期,即设置重复通知的间隔。重复的告警在静默期内不会被重复通知。例如 1d(1天)、2h(2小时)、3m(3分钟)。

添加完成后,您可以在监控规则列表中,单击数据投递延迟执行规则对应的+,查看已开启的告警实例。

_	数据投递延迟监控 ⑦	内置告警 SLS数据投递 投递延迟监控 展开	● 已开启(1)	;	屏蔽列表 ⑦	添加 关闭 临时关闭 删除 查看
	告警名称		行动策略	状态	外部配置	操作
	数据投递延迟监控 ⑦		sls.app.export.bu	● 已开启	屏蔽列表 ②	关闭 临时关闭 删除 设置 查看 关注

## 相关操作

为MaxCompute投递作业设置告警后,您还可以进行如下操作。

操作	说明
屏蔽列表	屏蔽列表中的MaxCompute投递作业不会触发告警。
关闭告警实例	关闭告警实例后,告警实例不会再触发告警, <b>状态</b> 变更为 <b>未开启</b> 。 该操作不会删除规则参数中已设置的信息。需要再次监控时,无需重新设置规则参数。
临时关闭告警实例	临时关闭告警实例后,在指定时间内不再触发告警。
恢复告警实例	处于临时关闭状态的监控实例,可随时恢复告警。
删除告警实例	删除告警实例, <b>状态</b> 变更为 <b>未创建</b> 。 该操作会删除规则参数中已设置的信息。需要再次监控时,需要重新设置规则参数。
设置告警实例	修改告警实例的配置参数。

## 数据投递监控规则

日志服务内置的数据投递监控规则如下所示。

- 数据投递延迟监控规则
- 数据投递异常报错监控规则
- 数据投递流量(绝对值)监控规则
- 数据投递流量(日同比)监控规则
- 数据投递失败条数监控规则
- 数据投递延迟监控规则

项目	说明
规则名称	数据投递延迟监控。
作用	用于监控MaxCompute投递作业中Shard消费延迟情况。当投递延迟时长大于监控阈值时, 触发告警。

项目	说明
	<ul> <li>监控的投递作业名称:需要监控的MaxCompute投递作业的名称。</li> <li>默认值为 .* ,表示监控当前阿里云账号下所有的MaxCompute投递作业。</li> <li>多个投递作业之间可以使用竖线( )分隔。您还可以使用正则表达式 .* 进行配置,例如sche.*,表示监控以sche开头的投递作业。</li> </ul>
参数配置	<ul> <li><b>监控阈值</b>:当MaxCompute投递作业的延迟时间超过该值时,触发告警。默认值为300 秒。</li> <li>行动策略:当前告警所绑定的行动策略,日志服务将通过该行动策略给指定用户发送告警</li> </ul>
	通知。 <ul> <li><b>严重度</b>:告警消息的严重度。</li> </ul>
	<ul> <li>● 静默期:告警静默期,即重复通知的间隔。重复的告警在静默期内不会被重复通知。例如 1d(1天)、2h(2小时)、3m(3分钟)。</li> </ul>
关联仪表盘	数据流量观测仪表盘中的进度落后图表。
消除方法	<ul> <li>请按照如下方法进行处理。</li> <li>i.确认源Logstore的数据量是否明显增长,处理速率是否同时提高。</li> <li>如果源Logstore的数据量增长的同时处理速率也提高,则说明源Logstore数据在增长,MaxCompute投递作业在自动扩容资源。请先观察5分钟,如果延迟降时间到正常范围,则说明问题解除。否则请执行下一步。</li> <li>如果源Logstore的数据量增长的同时处理速率未提高或者Shard消费延迟还处于上涨趋势,则说明可能是因为源Logstore的Shard数目不足,导致数据投递资源扩展受限,您需要手动分裂源Logstore的Shard。具体操作,请参见分裂Shard。完成分裂后,请先观察5分钟。如果延迟降时间到正常范围,则说明问题解除。否则请执行下一步。</li> <li>如果都不是,则执行下一步。</li> <li>i. 确认是否存在数据投递异常报错告警。</li> <li>如果存在,请按照数据投递异常报错监控的消除方法进行处理。处理完成后,观察5分钟。如果延迟降时间到正常范围,则说明问题解除。否则请执行下一步。</li> <li>如果不存在,则执行下一步。</li> <li>ii. 准备Project、Logstore和MaxCompute投递作业名称等信息,提交工单联系日志服务技术支持人员。</li> </ul>

### • 数据投递异常报错监控规则

项目	说明
规则名称	数据投递异常报错监控。
作用	用于监控MaxCompute投递作业中的异常报错。当MaxCompute投递作业出现异常报错后, 触发告警。
参数配置	<ul> <li>监控的投递作业名称:需要监控的MaxCompute投递作业的名称。</li> <li>默认值为 .* ,表示监控您当前阿里云账号下所有的MaxCompute投递作业。</li> <li>多个投递作业之间可以使用竖线 () 分隔。您还可以使用正则表达式 .* 进行配置,例如sche.*,表示监控以sche开头的投递作业。</li> <li>行动策略:当前告警所绑定的行动策略,日志服务将通过该行动策略给指定用户发送告警通知。</li> <li>严重度:告警消息的严重度。</li> <li>静默期:告警静默期,即重复通知的间隔。重复的告警在静默期内不会被重复通知。例如1d(1天)、2h(2小时)、3m(3分钟)。</li> </ul>

项目	说明	
关联仪表盘	<b>数据流量观测</b> 仪表盘中的运行异常图表。	
消除方法	<ul> <li>请按照如下方法进行处理。</li> <li>如果错误信息中包含Unauthorized、InvalidAccessKeyld或SignatureNotMatch,则说明MaxCompute投递作业读取源Logstore数据或者写入数据到MaxCompute的权限不足。请检查授权操作。更多信息,请参见授权概述。</li> <li>如果错误信息中包含ProjectNotExist或LogStoreNotExist,则说明MaxCompute投递作业涉及的Project或Logstore不存在。请登录日志服务控制台确认。</li> <li>如果错误信息中包含TransformError,则说明源Logstore中的数据不满足当前的数据投递逻辑。请从错误信息中找到原始数据的位置,然后修改MaxCompute投递作业的配置并重启作业。如何修改MaxCompute投递作业,请参见修改配置并重启作业。</li> <li>如果问题仍未解决,请准备Project、Logstore和MaxCompute投递作业名称等信息,提交工单联系日志服务技术支持人员。</li> </ul>	

#### • 数据投递流量(绝对值)监控规则

项目	说明	
规则名称	数据投递流量(绝对值)监控。	
作用	用于监控MaxCompute投递作业在5分钟内的平均投递条数。当平均投递的日志条数低于监 控阈值时触发告警。	
参数配置	<ul> <li>监控的投递作业名称:需要监控的MaxCompute投递作业的名称。</li> <li>默认值为 .* ,表示监控您当前阿里云账号下所有的MaxCompute投递作业。</li> <li>多个投递作业之间可以使用竖线 ( ) 分隔。您还可以使用正则表达式 .* 进行配置,例如sche.*,表示监控以sche开头的投递作业。</li> <li>监控阈值:当平均投递条数低于该值时,触发告警。默认值为40000行/秒。</li> <li>行动策略:当前告警所绑定的行动策略,日志服务将通过该行动策略给指定用户发送告警通知。</li> <li>严重度:告警消息的严重度。</li> <li>静默期:告警静默期,即重复通知的间隔。重复的告警在静默期内不会被重复通知。例如1d(1天)、2h(2小时)、3m(3分钟)。</li> </ul>	
关联仪表盘	数据流量观测仪表盘中的处理速率(events/s)图表。	
消除方法	<ul> <li>请按照如下方法进行处理。</li> <li>i.确认投递流量变化趋势与源Logstore的数据量增长或减少趋势是否一致。</li> <li>如果一致,则说明是源Logstore数据量增长或减少导致的。无需处理。</li> <li>如果不一致,则执行下一步。</li> <li>ii.确认是否存在数据投递延迟监控告警。</li> <li>如果存在,请处理延迟告警。处理完成后,先观察15分钟。</li> <li>如果数据投递延迟回到1分钟以内,且投递流量与源Logstore数据量增长或减少的 趋势一致,则说明问题解除。</li> <li>如果数据投递延迟回到1分钟以内,但是投递流量与源Logstore数据量增长或减少 的趋势仍然不一致,则进行下一步。</li> <li>如果不存在,则执行下一步。</li> <li>iii.准备Project、Logstore和MaxCompute投递作业名称等信息,提交工单联系日志服务 技术支持人员。</li> </ul>	

● 数据投递流量(日同比)监控规则

项目	说明	
规则名称	数据投递流量(日同比)监控。	
作用	用于监控MaxCompute投递作业在5分钟内相比昨日的增长阈值和下跌阈值情况。当增长大 于日同比增长阈值或下跌超过日同比下跌阈值时触发告警。	
参数配置	<ul> <li>监控的投递作业名称:需要监控的MaxCompute投递作业的名称。</li> <li>默认值为 .* ,表示监控您当前阿里云账号下所有的MaxCompute投递作业。</li> <li>多个投递作业之间可以使用竖线 () 分隔。您还可以使用正则表达式 .* 进行配置,例如sche.*,表示监控以sche开头的投递作业。</li> <li>日同比增长阈值:当数据投递量日增长大于该值时,触发告警。默认值为40%。</li> <li>日同比下跌阈值:当数据投递量日下跌大于该值时,触发告警。默认值为20%。</li> <li>行动策略:当前告警所绑定的行动策略,日志服务将通过该行动策略给指定用户发送告警通知。</li> <li>严重度:告警消息的严重度。</li> <li>静默期:告警静默期,即重复通知的间隔。重复的告警在静默期内不会被重复通知。例如1d(1天)、2h(2小时)、3m(3分钟)。</li> </ul>	
关联仪表盘	数据流量观测仪表盘中的处理速率(events/s)图表。	
消除方法	<ul> <li>请按照如下方法进行处理。</li> <li>i.确认投递流量变化趋势与源Logstore的数据量增长或减少趋势是否一致。</li> <li>如果一致,则说明是源Logstore数据量增长或减少导致的。无需处理。</li> <li>如果不一致,则执行下一步。</li> <li>ii.确认是否存在数据投递延迟监控告警。</li> <li>如果存在,请处理延迟告警。处理完成后,先观察15分钟。</li> <li>如果数据投递延迟回到1分钟以内,且投递流量与源Logstore数据量增长或减少的 趋势一致,则说明问题解除。</li> <li>如果数据投递延迟回到1分钟以内,但是投递流量与源Logstore数据量增长或减少的 的趋势仍然不一致,则进行下一步。</li> <li>如果不存在,则执行下一步。</li> <li>iii.准备Project、Logstore和MaxCompute投递作业名称等信息,提交工单联系日志服务 技术支持人员。</li> </ul>	

### • 数据投递失败条数监控规则

项目	说明
规则名称	数据投递失败条数监控。
作用	用于监控MaxCompute投递作业在15分钟内的投递失败情况。当投递失败的数据条数大于监 控阈值时,触发告警。

项目	说明
参数配置	<ul> <li>监控的投递作业名称:需要监控的MaxCompute投递作业的名称。</li> <li>默认值为 .* ,表示监控您当前阿里云账号下所有的MaxCompute投递作业。</li> <li>多个投递作业之间可以使用竖线 (])分隔。您还可以使用正则表达式 .* 进行配置,例如sche.*,表示监控以sche开头的投递作业。</li> <li>监控阈值:当投递失败的数据条数超过该值时,触发告警。默认值为10条。</li> <li>行动策略:当前告警所绑定的行动策略,日志服务将通过该行动策略给指定用户发送告警通知。</li> <li>严重度:告警消息的严重度。</li> <li>静默期:告警静默期,即重复通知的间隔。重复的告警在静默期内不会被重复通知。例如1d(1天)、2h(2小时)、3m(3分钟)。</li> </ul>
关联仪表盘	数据流量观测仪表盘中的读(写)失败条数图表。
消除方法	请按照如下方法进行处理。 i. 确认是否存在数据投递异常报错告警。 • 如果存在,请按照数据投递异常报错监控的消除方法进行处理。处理完成后,观察5 分钟。如果投递失败的条数降到正常范围,则说明问题解除。否则请执行下一步。 • 如果不存在,则执行下一步。 ii. 准备Project、Logstore和MaxCompute投递作业名称等信息,提交工单联系日志服务 技术支持人员。

## 5.6. 管理MaxCompute投递作业(新版)

您可以在数据投递概览页面管理MaxCompute投递作业,包括查看MaxCompute投递作业的基础信息、统计报表,修改配置并重启作业、删除MaxCompute投递作业等操作。

## 前提条件

- 已创建MaxCompute投递作业。具体操作,请参见创建MaxCompute投递作业(新版)。
- 已开启作业运行日志。具体操作,请参见开启作业运行日志。

## 操作入口

- 1. 登录日志服务控制台。
- 2. 在Project列表区域,单击目标Project。
- 3. 在左侧导航栏,选择作业>数据投递。
- 4. 单击目标MaxCompute投递作业。

## 查看MaxCompute投递作业基础信息

您可以在基础信息区域查看MaxCompute投递作业的创建时间、修改时间、源日志库等信息。

基础信息			
源日志库:	source-log	状态:	⑦ 运行中
任务名称:	export-c	调度ID:	c8b5fa
投递名称:	teal-th	时区:	+0800
规则修改时间:	2022-01-21 11:10:22	规则创建时间:	2022-01-21 11:10:22
MaxCompute Project:	shi hijun	MaxCompute Table:	Shi
MaxCompute Endpoint:	http://service.cn-hangzhou.maxcompute.aliyun.com/api	Tunnel Endpoint:	http://dt.cn-hangzhou.maxcompute.aliyun.com

## 查看统计图表

创建MaxCompute投递作业后,日志服务默认创建一个仪表盘,您可以在仪表盘中查看MaxCompute投递作业的运行指标。

图表名称	说明
读成功条数	从源Logstore中读取成功的日志条数。
读失败条数	从源Logstore中读取失败的日志条数。
读公网流量(MB)	通过公网读取数据的流量。
写成功条数	投递到MaxCompute成功的日志条数。
写失败条数	投递到MaxCompute失败的日志条数。
写公网流量(MB)	通过公网写数据的流量。
处理速率(events/s)	每秒处理的日志条数。
进度落后	展示该Logstore下每个Shard的延时信息。
运行异常	展示MaxCompute投递作业运行过程中的异常信息。 您可以在目标Project的internal-diagnostic_log Logstore中,通过关键词ERROR或者WARN 查看完整的错误日志。该目标Project为您在开启作业运行日志时所指定的Project。更多信 息,请参见 <mark>开通服务日志</mark> 。
运行状态	展示MaxCompute投递作业的运行状态。包括投递时间、成功条数、失败条数、平均响应时间 等信息。

## 修改配置并重启作业

您可以单击修改投递配置,修改MaxCompute投递作业的相关配置。

保存修改后,日志服务会自动重启作业。重启过程大概需要几十秒钟,您可以单击**刷新**,确认重启是否完成。重启后,请 务必确认作业已正常启动,并检查投递到MaxCompute的数据符合预期。

⑦ 说明 无法修改MaxCompute投递名称和开始时间。

## 删除MaxCompute投递作业

如果您不再需要运行该MaxCompute投递作业,您可以在数据投递概览页面的右上角选择更多 > 删除。

☐ 警告 删除MaxCompute投递作业后,不可恢复,请谨慎操作。

## 6.将日志服务数据投递到AnalyticDB MySQL

日志服务采集到日志后,支持将日志投递至云原生数据仓库AnalyticDB MySQL进行存储与分析。本文档介绍将日志投递至 AnalyticDB MySQL的操作步骤。

### 前提条件

- 已采集日志到目标Logstore。更多信息,请参见数据采集。
- 在AnalyticDB MySQL中已完成以下准备工作:
  - i. 在日志服务Project所在地域,创建AnalyticDB MySQL集群。更多信息,请参见创建集群。
    - ⑦ 说明 目前日志服务仅支持同地域投递。
  - ii. 创建数据库账号。更多信息,请参见创建数据库账号。
  - iii. 创建数据库。更多信息,请参见创建数据库。
  - iv. 如果您需要通过外网地址连接AnalyticDB MySQL集群,还需要申请外网地址。更多信息,请参见申请和释放公网地 址。
  - v. 创建数据库表。更多信息,请参见CREATE TABLE。

#### 创建投递任务

- 1.
- 2. 在Project列表区域,单击目标Project。
- 3. 在日志存储 > 日志库页签中,依次展开目标Logstore下的数据处理 > 导出,单击AnalyticDB右侧的+。
- 4. 在投递提示对话框中, 单击直接投递。
- 5. 首次创建投递任务到AnalyticDB MySQL时,需创建服务关联角色AliyunServiceRoleForAnalyticDBForMySQL。
  - i. 在创建服务关联角色对话框中,单击AliyunServiceRoleForAnalyticDBForMySQL。

#### \rm 创建服务关联角色

执行此操作时,将会自动创建一个服务关联角色,以完成相应功能。 角色名称:AliyunServiceRoleForAnalyticDBForMySQL 角色权限策略:AliyunServiceRolePolicyForAnalyticDBForMySQL 权限说明:允许AnalyticDB for MySQL服务通过该角色操作日志,同步数据链路。 文档链接:参见

- ii. 在创建服务关联角色对话框中, 单击确定。
- 6. 在LogHub —— 数据投递页面,配置相关参数,然后单击确定。

LogHub —— 数据扮	通
选择区域 华东1(杭	
Project名	称 tu tu
Logstore名	你 。
*投递名	称 sls_adb
<b>*</b> 投递描	述 投递到ADB
◆集群版	本
*集群名	称 am- ixOue9 ~
*数据库名	称 adb_demo ~
*表	S test 🗸
<b>*</b> 账号名	称 Numerica Antonio Anto
*账号密	······

参数	说明	
投递名称	配置投递任务的名称,便于后续管理。	
投递描述	配置投递任务的描述,便于后续管理。	
集群版本	选择AnalyticDB MySQL集群版本,此处以选择3.0为例。	
集群名称	选择您已创建的AnalyticDB MySQL集群。	
数据库名称	选择您已创建的AnalyticDB MySQL集群中的数据库。	
表名	选择您已创建的AnalyticDB MySQL集群中的数据库表。	
账号名称	配置为您在AnalyticDB MySQL集群中创建的数据库账号名称。	
账号密码	配置为您在AnalyticDB MySQL集群中创建的数据库账号密码。	
字段映射	系统自动提取日志服务中最近10分钟的日志字段,同时自动映射对应的AnalyticDB MySQL数据 表中的字段。左边文本框为日志字段名称,右边为AnalyticDB MySQL数据库表中的字段。 ⑦ 说明 Timestamp类型的字段映射精确到秒级别。例如日志时间为2022-01-05 10:12:13.145,投递到AnalyticDB MySQL后,投递结果为2022-01-05 10:12:13。	
投递开始时间	配置投递开始时间。 当日志服务采集到日志后,日志服务可以实时将数据投递到AnalyticDB MySQL。	
是否过滤脏数据	脏数据包括数据类型转化失败和非空字段为空等数据。 <ul> <li>如果打开<b>是否过滤脏数据</b>开关,则遇到脏数据时,过滤掉脏数据。</li> <li>如果关闭<b>是否过滤脏数据</b>开关,则遇到脏数据时,投递任务中断。请谨慎选择。</li> </ul>	

创建投递任务后,您可以在日志服务控制台上管理数据投递任务,包括查看任务详情、修改投递规则以及启动、停止、删 除任务等操作。更多信息,请参见<mark>管理日志投递任务</mark>。

## 查看日志数据

将日志投递到AnalyticDB MySQL后,您就可以在数据库中通过SELECT语句查询日志数据。

# 7.将日志服务数据投递到TSDB

本文介绍如何通过日志服务将日志数据投递到TSDB,并映射为时间点进行时序存储,以满足其他应用场景需求。

## 背景信息

○ 注意 尊敬的用户您好,阿里云日志服务SLS投递ApsaraDB for Lindorm功能目前已暂停新增投递配置,如您有 投递ApsaraDB for Lindorm的需求,请提交工单与日志服务团队联系,我们将基于您的业务需求为您提供相应的解决 方案。感谢您的支持与理解。

日志作为一种特殊的数据,对处理历史数据、诊断问题以及了解系统活动等有着非常重要的作用,对数据分析人员、开发 人员以及运维人员而言,日志也是其工作过程中必不可缺的数据来源。

阿里云日志服务考虑到不同的应用场景,提供了从各种云产品进行日志采集并深度处理的完善服务,TSDB支持从日志服 务导入日志记录,并映射为时间点进行时序存储。

### 前提条件

- 已开通日志服务和TSDB。
- 日志服务和TSDB必须在同一个区域(region),目前暂不支持跨区域数据交互。

### 创建投递任务

按照如下步骤演示进行任务创建过程。

- 1. 登录日志服务控制台,单击选择需要导出的Project和LogStore。
- 2. 在日志库列表中,单击目标Logstore名称前的>,然后依次单击数据处理 > 导出 > TSDB。



3. 单击TSDB后的+,将会跳转到单独的导出任务配置页,请您在配置页继续完成导出任务的创建。

⑦ 说明 如果日志库尚未做任何数据加工, 日志服务会提示是否先完成日志的数据加工再进行投递, 如果日志 记录很规范, 也可以选择不经加工直接投递到TSDB。

4. 根据参数信息填写任务配置项。

LogHubProject名称:	
LogHubLogstore名称:	100.000.00
* 投递名称:	
* 投递描述:	
* TSDB 实例:	Υ
* 度量名称:	
* AccessKeyID:	
* AccessKeySecret:	
* 时间点映射:	\${logTag:time}
* 字段映射:	= value 字符串 ▼ 💼
	◆ 增加一条
* 标签映射:	=
	+ 增加一条
* 投递时间:	2020-04-0214 $\widehat{}$ :30 $\widehat{}$ :06 $\widehat{}$
	提交

### 请您参见如下参数信息表进行配置:

参数	说明
投递名称	自定义一个投递名称,便于后续管理。
投递描述	为投递任务填写有意义的描述,便于后续管理。
TSDB实例	当前Region中已购TSDB列表,选中需要导出数据的实例。
度量名称	导出到哪一个TSDB度量(metric),度量不存在时会自动创建。
AccessKeyID	当前登录账号的AK ID(供T SDB访问)。
AccessKeySecret	当前登录账号的AK Secret(供TSDB访问)。

参数	说明
	将日志记录中的时间戳映射到TSDB时间戳,默认将日志发生时间作为TSDB时间戳( \$(logTag:time}),该字段支持用表达式(\$(logTag:xx})与(\$(logColumn:xx})做数值替 换,表达式替换失败的情况下,时间戳会采用SLS日志记录时间(\$(logTag:time})。表达式 所引用的字段必须可以转换为数值(Epoch seonc/milli-second)。由于SLS日志时间戳 (time)精度只采集到秒,如果您的日志在一秒内有可能出现多条,那么您需要指定哪些SLS 日志信息作为TSDB的目标时间戳。支持使用表达式来引用一些上下文信息,表达式语法为:
	<pre>\${context:variable}</pre>
时间点映射	更多SLS日志上下文变量信息,请参见上下文查询。
	表达式中logColumn:前缀表示引用原始SLS日志中所属column的值作为映射值。
	表达式中logTag:前缀表示引用原始SLS日志中所属tag的值作为映射值。常见Tag:
	<ul> <li>topic: 整个LogGroup(一条或多条日志)的业务标记(表达式引用格式 \${logTag:topic)。</li> </ul>
	○source:产生该条日志记录的机器IP地址(表达式引用格式\${source})。
	○time:日志记录的时间戳(秒级)(表达式引用格式\${time})。
字段映射	配置哪些日志字段为TSDB字段,依次填写日志字段名称与TSDB字段名称。当只存在单个字段 映射时,导出为TSDB单值模型(TSDB字段名称只能是value),当存在多个字段映射时,导出 为TSDB多值模型。
标签映射	为导出的数据记录设置TSDB标签,依次填写标签名称与标签值,该配置项的标签名称与标签值 支持使用表达式来引用一些上下文信息,表达式语法为:
	<pre>\${context:variable}</pre>
	更多SLS日志上下文变量信息,请参见 <mark>上下文查询</mark> 。
投递时间	从什么时间开始数据投递(小于该时间范围的数据不会被投递)。

5. 当所有参数填写完毕后, 单击提交即可创建数据投递任务。

## 相关操作

- TSDB导出任务在创建完成后立即启动,如果您需要暂停或者查看服务状态,可以在导出任务创建完毕后,在SLS控制面板单击任务名称,进入任务管控面板进行操作。
- 任务创建完毕后, 您也可以修改任务配置, 其中配置参数说明请参见创建投递任务。

## 8.管理日志投递任务

创建AnalyticDB MySQL或TSDB投递任务后,您可以在日志服务控制台上管理数据投递任务,包括查看任务详情、修改投 递规则以及启动、停止、删除任务等操作。本文以AnalyticDB MySQL为例,介绍相关操作。

### 查看任务详情

- 1.
- 2. 在Project列表区域,单击目标Project。
- 3. 在日志存储 > 日志库页签中,依次展开目标Logstore下的数据处理 > 导出 > AnalyticDB,单击目标投递任务。
- 4. 查看任务详情。

ADB (分析型数据库 ) 投递微宽 (ma)									
	ADB(分析	型数据库)投递概览	(1965)				2 (4)	0 (P)	上 88更多
	基础信息								
	源日志库	oss-sis			消费进度	shard ↓†	最近消费数据时间11		
43回0 000mmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm	状态	⑤ 运行中				0	2020-11-04 13:44:18		
ADDRESSERIERS ADDRESSERIERS 2020-11-04 11:33:39 対照15152 2020-11-04 11:33:39 対照15152 2020-11-04 11:33:39 対照15152 単版素 Progect Legistore 表示。Serro test 05:455 分類現役違论語T 07.855 第週首位 外記55 07.855 07.855 第週首位 第週目示 第週首位 第週目示 第週首位 第週目示 第四目示 第二目示 第二	任時ID	0a0	D924a55			1	2020-11-04 13:39:18		
#399638999 22020-11-04 11:33:39 P148195 単元本 Project Logatore 表型、6mm 1 tot 054-09 外出行技表 第3名作型	规则修改时间	2020-11-04 11:33:39							
中国語     日前他     Project     Logitore     Logitor	规则创建时间	2020-11-04 11:33:39							
add_demo         bot         cos+d6           使計探表         〇 知道日 (15/10)         〇 知道日 (15/10)           構造作业         算造实別         算造数据源         算造目标           作业信仰:         東田田:         0.000 (10/10)         第造数据源           講協のgrave濃縮品計         15/10(11)         15/10(11)         15/10(11)           講協のgrave濃縮品計         15/10(11)         15/10(11)         15/10(11)	存儲目标	目标库	Project	Logstore					
Ch 指接表     C		adb_demo	test	oss-sis					
の IN IN R	(A)_1+P==								
受 数据投送回诊断         受 利益         受 利益         受 利益         受 利益         受 利益         受 利益         受 利益 <td>AD HEARING M</td> <td>110.107</td> <td></td> <td></td> <td></td> <td></td> <td>0.25242 - 0.802</td> <td>6 +C#5/1#6</td> <td>2000+12</td>	AD HEARING M	110.107					0.25242 - 0.802	6 +C#5/1#6	2000+12
Weight III         Press/UNI	⑤ 数据投速 結选作业	1121時1	能决定例		德洪教祝酒		(1) 11 12 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		30.000 (P)
中臣(1817): 第二日 1947年_× Logitore: 第二日 1947年_× Logitore: 第二日 1947年_× Logitore: 第二日 1947年_1941日 1947年 19	MARKET P. SEC.		300220003		PPACKAGAGAP		MAYOR CO.		
读Logstore流服总计 1天(====) : 投递成功行数总计 1天(====) : 投递大或行数总计 1天(====) : : 投递大或行数总计 1天(====) : : : : : : : : : : : : : : : : :	作业名称;		实例ID: 0a0(	1547e ×	Logstore :		美型:		
	读Logstore流	<b>通总计</b> 1天(相对)	: 19	<b>递成功行数总计</b> 1天(相对)		: 投递失时	<b>收行数总计</b> 1天(相时)		:
23.326K Bytes         34 Lines         O Lines           昨日研究         昨日研究         昨日研究         昨日研究		23.326K Bytes 昨日环比		34 昨日	· Lines 研記		O Lines 昨日环社		

参数	说明
源日志库	进行数据投递的Logstore。
状态	数据投递任务的状态,包括启动中、运行中、停止中、已终止、成功、失败。
任务ID	数据投递任务的ID。
规则修改时间	修改数据投递任务的时间。
规则创建时间	创建数据投递任务的时间。
存储目标	包括源Project、源Logstore和目标数据库信息。
消费进度	数据投递任务读取源Logstore日志数据的消费进度。
统计报表	创建投递任务后,自动生成数据投递仪表盘,包括总览指标、每秒投递行数指标、消 费延迟指标、消费速率指标、读写速率指标和详情日志。更多信息,请参见 <mark>数据投递</mark> 仪表盘。

## 停止任务

对于状态为运行中的任务,可单击页面右上角的停止,停止任务。

⑦ 说明 停止任务会保存当前的投递位置,如果再次启动任务会从停止的位置继续投递。

## 启动任务

对于状态为已终止的任务,可单击页面右上角的启动,重新执行投递。

### 修改投递规则

创建投递任务之后,您可以单击**修改投递配置**,修改投递规则。

修改投递规则后,您需要通过如下操作确认投递规则是否符合预期结果。

- 1. 检查新生成的投递任务的执行状态是否正常。
- 2. 检查新生成的投递任务投递到目标存储的数据内容是否符合预期。

### 删除任务

如果您要删除数据投递任务,可在页面右上角,选择更多 > 删除,删除当前任务。

警告 删除数据投递任务后,不可恢复,请谨慎删除。

# 9.数据投递仪表盘

创建AnalyticDB MySQL或TSDB投递任务后,日志服务自动为每个投递任务生成一个仪表盘,您可以在仪表盘中查看数据 投递任务运行指标。本文介绍投递任务运行指标详情及如何查看数据投递仪表盘。

### 查看仪表盘

- 1.
- 2. 在Project列表区域,单击目标Project。
- 3. 在日志存储 > 日志库页签中,依次展开目标Logstore下的数据处理 > 导出 > AnalyticDB,单击目标投递任务。
- 4. 在统计报表区域,查看仪表盘信息。

仪表盘信息包括总览指标(读Logstore流量总计、投递成功行数总计和投递失败行数总计)、每秒投递行数、Shard 消费延迟、Shard消费速率、读Logstore与写目标速率和日志详情。

### 总览指标

总览指标信息如下所示:

袭Logstore流量总计 1天(組对)	:	投递成功行数总计 1天(相对)	:	<b>投递失败行数总计</b> 1天(相对)	:
15.351Bil <sub>Bytes</sub> 昨日呼比		22,288,000 Lines 昨日环比		13,786,000 Lines 昨日开社	

- 读Logstore流量总计:从源Logstore的各个Shard中读取到的日志条数。
- 投递成功行数总计:从源Logstore的各个Shard中读取日志后,投递到目标云产品成功的日志条数。
- 投递失败行数总计:从源Logstore的各个Shard中读取日志后,投递到目标云产品失败的日志条数。

## 每秒投递行数

统计每分钟窗口内,数据投递任务处理的日志条数。



- dropped:从源Logstore的各个Shard中读取日志后,在投递过程中被丢弃的日志条数。
- delivered:从源Logstore的各个Shard中读取日志后,投递到目标云产品成功的日志条数。
- failed:从源Logstore的各个Shard中读取到日志后,投递到目标云产品失败的日志条数。

### Shard消费延迟和Shard消费速率

统计每分钟窗口内,投递任务读取源Logstore数据时,每个Shard的指标。



- 消费延迟:该Shard中的日志被完全消费的时间与该Shard最新采集到的日志的时间之差。
- 消费速率:在一分钟窗口内, Shard平均每秒钟读取到的日志条数。

## 读Logstore与写目标速率



统计每分钟窗口内,读取源Logstore各个Shard中的日志条数与写入目标云产品的日志条数。

## 日志详情

在**详情日志**区域,查看已投递的日志详情。您也可以在当前Project下的**internal-export-log**日志库中,通过关键词 ERROR或WARNING查看完整的错误日志。

详情日志 4小时(相对)											:
start_time :	end_time	level	\$ Q,	code	¢ 0,	message \$ 0	count			\$ C.	
20/10/10 10:07:00	20/10/10 10:07:59	ERROR		WriteDataFalled		Code=106 Message=The column id of table sistoada can not be null	275000				
20/10/10 10:06:00	20/10/10 10:06:59	ERROR		WriteDataFailed		Code=106 Message=The column id of table sistoadb can not be null	858000				
20/10/10 10:05:00	20/10/10 10:05:59	ERROR		WriteDataFalled		Code=106 Message=The column id of table sistoadb can not be null	859000				
20/10/10 10:04:00	20/10/10 10:04:59	ERROR		WriteDataFalled		Code=106 Message=The column id of table sistoadb can not be null	792000				
20/10/10 10:03:00	20/10/10 10:03:59	ERROR		WriteDataFalled		Code=106 Message=The column id of table sistoadb can not be null	744000				
20/10/10 10:02:00	20/10/10 10:02:59	ERROR		WriteDataFalled		Code=106 Message=The column id of table sistoadt can not be null	459000				
								总数:19	1 /	1 >	

## 10. 配置监控与告警

创建AnalyticDB MySQL或TSDB投递任务后,您可以在数据投递诊断仪表盘中查看投递任务的运行指标信息,还可以订阅 仪表盘,针对仪表盘中的指标信息设置告警。本文介绍为投递任务配置监控与告警的操作步骤。

## 背景信息

在数据投递诊断仪表盘中,建议您关注如下两方面的指标信息。

- 系统层面: 投递的消费延迟、是否有异常报错。
- 业务层面:处理日志行数、输出日志行数。

### 创建告警

- 1.
- 2. 在Project列表区域,单击目标Project。
- 3. 在左侧导航栏中,单击 🕑 图标。
- 4. 在仪表盘列表中, 单击数据投递诊断。
- 右数据投递诊断页面,筛选目标作业,为不同指标设置告警。 如何设置告警,请参见设置告警;如何订阅仪表盘,请参见订阅仪表盘。

### 配置投递延迟告警

- 1. 在shard消费延迟图表中, 单击: > 创建告警。
- 2. 配置告警。

例如: 触发条件设置为delay > 120, 表示阈值大于120秒时告警。更多信息,请参见设置告警。

HEHO.	export-adb		10/
* 关联图表	0 图表名称 shard消费延迟 (sec	cond)	$\checkmark$
	查询语句tag:_job_type metric   select date % 60, %y-%m-% ard ', shard) as sh, as delay group by d 000	e_; Export and type:source e_format(timetime %d %H:%!) as dt, concat('sl round(avg(fetchedDelay), 3 tt, sh order by dt asc limit 10	) D
	查询区间 ③ 15分钟(相对)		
	1) 添加		
* 频率	■定间隔 ∨ 15	分钟	
* 触发条件	delay > 120		
	支持加(+)、减(-)、乘(*)、除(/)、]	取模(%)5种基础运算符利	1大于

### 3. 配置通知方式。

此处以WebHook-钉钉机器人为例。更多信息,请参见通知方式。



4. 在钉钉群中,查看告警通知。



## 配置异常报错告警

- 1. 在详情日志图表中, 单击 > 创建告警。
- 2. 配置告警。

例如: 触发条件设置为level == 'ERROR'。更多信息,请参见设置告警。

* 告警名称	export-adb-error		1	6/6
* 关联图表	<b>0</b> 图表名称	详情日志	$\sim$	6
	查询语句 查询区间 1 — 参加	tag_:_job_type_: E ect date_format("start", 1:%S') as start_time, dat y/%m/%d %ht%s(%S') de, message, count order mit 10000 ⑤ 1小時 (借評)	<pre>kport and type:log   sel %y/%m/%d %H:% e_format("end", '% as end_time, level, co ir by start_time desc li</pre>	
* 頻率	固定间隔	∨ 1	分钟	$\sim$
· AL 113 / 11	level == 'ERROR'			

### 3. 配置通知方式。

此处以WebHook-钉钉机器人为例。更多信息,请参见<mark>通知方式</mark>。

通知列表		WebHook-钉钉机器人	~ 添加
			导入已有通知配置
✓ WebHook-1	订钉机器人		×
* 请求地址	https://oapi.dingt	alk.com/robot/sen	01008004
标题	[数据投递告警] ex	port-adb-error-log	
提醒接收者	不提醒	所有人 指定成员	
被@人列表			
	多个手机号用逗号(,)	分隔,在发送内容里要有@手机号	
* 发送内容	<pre>- [Uid] \${aliu - [Project] [\$ - [Trigger] \${ - [Condition]</pre>	id} {project}](https://sls.con AlertDisplayName) S/condition)	sole.aliyun.com
	- [Message] 通	upe	

### 4. 在钉钉群中,查看告警通知。



### 配置投递流量(绝对值)告警

- 1. 在每秒投递行数图表中, 单击; > 创建告警。
- 2. 配置告警。

```
例如:触发条件设置为accept < 2000,表示当每秒处理日志条数少于2千行时产生告警。更多信息,请参见设置告
警。
```

	告警配置	通知	
* 告警名称	export-adb-speed	1	6/64
* 关联图表	0	每秒投递行数(lines/s) V	×
	查询语句	tag_:job_type: Export and type:sink_m etric   select time_series(time '1m', '% y/%m/%d %H:%i', '0') as dt, round(sum(droppe d) / 60.0, 3) as dropped, round(sum(droppe d) / 60.0, 3) as dropped, round(sum(tailed) / 60. 0, 3) as failed group by dt order by dt asc limit 10000	
	查询区间	③ 1小时(相对)	
	· 1 添加		
频率	固定间隔	∨ 1 分钟	$\sim$
* 触发条件	delivered<2000		
	支持加(+)、减(-) (>)、大于等于(>=) (!=)、正則匹配(=~	、乘(*)、除(/)、取模(%)5种基础运算符和大 )、小于(<)、小于等于(<=)、等于(==)、不 -)、正则不匹配(I~)8种比较运算符。 <b>帮助文档</b>	;于 等于
No. 611 141 101			

3. 配置通知方式。

此处以WebHook-钉钉机器人为例。更多信息,请参见<mark>通</mark>知方式。

	古言記旦	· · · · · · · · · · · · · · · · · · ·	쁘지
通知列表		WebHook-钉钉机器人	→ 添加
✓ WebHook- <sup>4</sup>	钉钉机器人		>
* 请求地址	https://oapi.dingta	lk.com/robot/send?ac	;c
标题	[数据投递告警] exp	oort-adb-speed	
提醒接收者	不提醒所	有人 指定成员	
被@人列表			
	多个手机号用逗号(,)分	<b>计隔,在发送内容里要有@手机号</b>	
* 发送内容	- [Uid] \${aliui - [Project] [\${ - [Trigger] \${A - [Condition] \$	d} project)](https://sls.com lertDisplayName} {condition}	nsole.aliyun.cor
	- [Message] 通知	1内容	

4. 在钉钉群中, 查看告警通知。



## 配置投递流量(日同比)告警

1. 自定义监控指标。

i. 在日志存储 > 日志库页签中, 单击internal-export-log。

ii. 输入如下查询分析语句, 单击查询/分析。

该查询分析语句计算每5分钟投递的日志行数与昨日同比的指标。

```
*! select dt, today, yesterday, round((today - yesterday) * 100.0 / yesterday, 3) as inc_ration
from
(select dt, (case when diff[1] is null then 0 else diff[1] end) as today, (case when diff[2] is
null then 0 else diff[2] end) as yesterday from
(select dt, compare("delivered lines", 86400) as diff from
(select date_format(__time__ - __time__ % 300, '%H:%i') as dt, sum(delivered) as "delivered lin
es" from log group by dt order by dt asc limit 5000)
group by dt order by dt asc limit 5000))
```

⑦ **说明** 您还可以修改该查询分析语句以设置更精确的告警指标。例如只对任务ID为 5a04f1d44d44db3005fd268f39c1aebf的任务设置告警。

\_\_tag\_:\_\_schedule\_id\_: 5a04f1d44d4db3005fd268f39claebf | select dt, today, yesterday, r ound((today - yesterday) \* 100.0 / yesterday, 3) as inc\_ration from (select dt, (case when diff[1] is null then 0 else diff[1] end) as today, (case when diff[2 ] is null then 0 else diff[2] end) as yesterday from (select dt, compare("delivered lines", 86400) as diff from (select date\_format(\_\_time\_\_ - \_\_time\_\_ % 300, '%H:%i') as dt, sum(delivered) as "delivered lines" from log group by dt order by dt asc limit 5000) group by dt order by dt asc limit 5000))

iii. 在统计图表中,选择线图,并单击添加到仪表盘,添加该查询结果到仪表盘(此处以export-adb-monitor仪表盘为例)。



- 2. 在左侧导航栏中, 单击 🕒 图标。
- 3. 在仪表盘列表中,单击export-adb-monitor。
- 4. 在export-adb-monitor仪表盘中,找到目标图表,单击; > 创建告警。



5. 配置告警。

例如,**触发条件**设置为inc\_ration < (-40),表示当日志投递速率比昨日下降40%时告警。更多信息,请参见设置告警。

*告警名称 exp	oort-adb-inc-rat	tion-by-day		28/
*关联图表 0	图表名称	n log-lines-inc-ration	ı-by-day	$\sim$
	查询语句	tag_:_schedul 5fd268f39c1aebf ound((today - yest as inc_ration from is null then 0 else hen diff[2] is null th erday from (select s", 86400) as diff metime 9 elivered) as "delive dt order by dt asc by dt asc limit 500	_id: 5a04f1d4d44 select dt, today, yester arday) * 100.0 / yester (select dt, (case when iff(1) end) as today, (c. en 0 else diff[2] end) as dt, compare(*delivered rom (select date_forma 3 300, *%H:%I') as dt, i vred lines* from log gro limit 5000) group by dt 0))	1b300 day, r lay, 3) diff[1] 1se w s yest l line t(ti sum(d up by order
1	查询区间 	] ① 1小时(相对)		
频率 固)	王间開	∨ 1	分钟	
· 釉发各件 inc	ration>40			

6. 配置通知方式。

此处以WebHook-钉钉机器人为例。更多信息,请参见通知方式。



7. 在钉钉群中, 查看告警通知。



## 告警相关操作

您可以在告警列表中进行删除、修改或禁用通知操作。更多信息,请参见管理告警配置。

# 11.最佳实践 11.1.投递-对接数据仓库

日志服务日志投递(LogShipper)功能可以便捷地将日志数据投递到OSS、Table Store、MaxCompute等存储类服务, 配合E-MapReduce(Spark、Hive)、MaxCompute进行离线计算。

## 数据仓库(离线计算)

数据仓库(离线计算)是实时计算的补充,两者针对目标不同:

模式	优势	劣势	使用领域
实时计算	快速	计算较为简单	增量为主,监控、实时分析
数据仓库(离线计算)	精准、计算能力强	较慢	全量为主,BI、数据统计、比 较

目前对于数据分析类需求,同一份数据会同时做实时计算+数据仓库(离线计算)。例如对访问日志:

- 通过流计算实时显示大盘数据,例如包括当前PV、UV、各运营商信息。
- 每天晚上对全量数据进行细节分析,例如分析增长量、同比、环比和Top数据等。

互联网领域有两种经典的模式讨论:

- Lamdba Architecture: 数据进来后, 既支持流式处理、同时存入数仓。但用户发起查询时, 会根据查询需求和复杂度 从实时计算、离线计算拿结果返回。
- Kappa Architecture: kafka based Architecture。弱化离线计算部分,数据存储都在Kafka中,实时计算解决所有问题。

日志服务提供模式比较偏向Lamdba Architecture。

### LogHub/LogShipper一站式解决实时+离线场景

在创建Logstore后,可以在控制台配置日志投递(LogShipper)支持数据仓库对接。当前支持如下:

- OSS (大规模对象存储):
- TableStore (NoSQL数据存储服务): 由表格存储提供服务,从日志服务拉取数据。更多信息,请参见日志数据传送。
- MaxCompute (大数据计算服务):
|               | OSS、TableStore + E-MapReduce<br>MaxCompute                              |  |
|---------------|---|--|
| LogShipper    | Data Storage  |  |
|               | 4 3<br>5 6<br>7 8<br>9  |  |
| Logstore      |   |  |
| ····          | Spark Streaming   |  |
| 9 8 7 6 5 4 3 | <ul> <li>Storm/JStorm</li> <li>Consumer Library</li> <li>SDK</li> </ul> |  |
| LogHub 实时收集 → | → 实时计算  |  |

LogShipper提供如下功能:

- 准实时:分钟级进入数据仓库
- 数据量大:无需担心并发量
- 自动重试:遇到故障自动重试、也可以通过API手动重试
- 任务API: 通过API可以获得时间段日志投递状态
- 自动压缩: 支持数据压缩、节省存储带宽

# 典型场景 1: 日志审计

小A维护了一个论坛,需要对论坛所有访问日志进行审计和离线分析。

- G部门需要小A配合记录最近180天内用户访问情况,在有需求时,提供某个时间段的访问日志。
- 运营同学在每个季度需要对日志出一份访问报表。

小A使用日志服务收集服务器上日志数据,并且打开了日志投递(LogShipper)功能,日志服务就会自动完成日志收集、 投递、以及压缩。有审查需要时,可以将该时间段日志授权给第三方。需要离线分析时,利用E-MapReduce运行一个30 分钟离线任务,即可用最少的成本完成这两件事情。也可以使用阿里云DLA对投递到OSS中的日志数据进行数据分析。

### 典型场景 2: 日志实时+离线分析

小B是一个开源软件爱好者,喜欢利用Spark进行数据分析。需求如下:

- 移动端通过API收集日志。
- 通过Spark Streaming对日志进行实时分析,统计线上用户访问。
- 通过Hive进行T+1离线分析。
- 将日志数据开放给下游代理商,进行其他维度分析。

该场景下,通过SLS+OSS+EMR/DLA+RAM组合,可轻松实现这类需求。

# 12.常见问题 12.1. 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,由日志的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\_M\_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,计算细节参考投递日志到 MaxCompute(旧版)。

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API、SDK或控制台判断(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 表示日志服务接收到数据的时间,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或者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正斜线(/)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。

更多细节请参考使用限制。

# 12.2. 投递日志到OSS失败

接入点错误或RAM角色不存在会导致投递日志到OSS失败。

#### 问题现象

投递日志到OSS失败的可能报错如下:

● 接入点错误。

AccessDenied: The bucket you are attempting to access must be addressed using the specified endpoint. Please send all future requests to this endpoint.

• RAM角色不存在。

EntityNotExist.Role: The role not exists:

#### 可能原因

根据报错,可能的错误原因如下:

- 如果报错是接入点错误,则说明日志服务的Project和OSS的Bucket的地域不一致。
- 如果报错是RAM角色不存在,则说明填写的RAM角色的ARN错误。

# 解决方案

#### 操作步骤

- 1. 根据报错,选择解决方案:
  - 如果错误原因是地域不一致,您需要确保日志服务的Project和OSS的Bucket的地域一致。

↓ 注意 如果日志服务Project属于杭州金融云,则只支持投递到杭州金融云公网地域(oss-cn-hzfinance),不支持华东1金融云地域(oss-cn-hzjbp)。杭州金融云公网地域的OSS接入点为oss-cn-hzfinance-internal.aliyuncs.com。

• 如果错误原因是填写的RAM角色的ARN错误,您需要填写正确的ARN。

如何获取正确的ARN,请参见如何获取ARN。

# 12.3. MaxCompute投递作业(新版)常见报错 与问题

本文介绍MaxCompute投递作业(新版)的常见报错与问题。

# 出现用户不存在错误时,如何解决?

如果MaxCompute投递(新版)的执行记录中显示 ODPS-0420095: Access Denied - Authorization Failed [4002], You don't exist in project \*\*\*\*. Context ID:11111-1111-1111-1111-1111111. --->Tips:

Pricipal:INVALID\$v4\_11111111111; You don't exist in project \*\*\*\* 错误,表示您在添加RAM角色为MaxCompute 工作空间成员时,未添加目标用户或者添加用户错误,请重新添加用户并完成授权,具体操作,请参见通过默认角色写数据 到MaxCompute或通过自定义角色写数据到MaxCompute。

## 出现RAM角色无权限错误时,如何解决?

如果MaxCompute投递(新版)的执行记录中显示 ODPS-0420095: Access Denied - Authorization Failed [4019], You have NO privilege 'odps:Describe' on {acs:odps:\*:projects/\*\*\*\*/tables/\*\*\*\*}. Context ID:11111-1111-11111-1111-111111. --->Tips: Pricipal:INVALID\$v4\_11111111 错误,表示您在添加RAM角色为MaxCompute工作 空间成员时,未为RAM角色授权,请重新为RAM角色授权。具体操作,请参见通过默认角色写数据到MaxCompute或通过自定 义角色写数据到MaxCompute。

#### 出现 FAILED: mismatched input 错误时,如何解决?

如果您使用RAM用户在操作添加RAM角色为MaxCompute工作空间成员时,出现 FAILED: mismatched input 错误,表示该RAM用户没有执行 ADD USER 等语句的权限。出现该错误时,您需要先使用阿里云账号或具备如下权限策略的RAM 用户执行 ADD USER 等语句。完成此次操作后,后续您再使用RAM用户执行 ADD USER 等语句,不会出现该错误。

```
{
  "Version": "1",
  "Statement": [
    {
       "Action": "ram:CreateServiceLinkedRole",
       "Resource": "*",
       "Effect": "Allow",
       "Condition": {
        "StringEquals": {
            "ram:ServiceName": "usermgmt.odps.aliyuncs.com"
        }
     }
    }
}
```

出现操作用户无权限错误时,如何解决?

如果当前操作用户在操作添加RAM角色为MaxCompute工作空间成员时,出现 FAILED: ODPS-0130013:Authorization exception - Authorization Failed [4003], You have NO privilege to do the PROJECT SECURITY OPERATION for {acs:odps:\*:projects/\*\*\*\*/authorization/users}. Context ID:1111-1111-1111-1111-1111. 错误,表示当前的操 作用户无MaxCompute账户操作或者授权权限,您需要为该用户添加admin角色。更多信息,请参见MaxCompute和 DataWorks的权限关系。

您可以参见如下步骤,为用户添加admin角色。

- 1. 登录MaxCompute控制台。
- 2. 在页面左上角,选择地域。
- 3. 在项目管理页签中,单击目标项目对应的成员管理。
- 4. 在左侧导航栏中,单击MaxCompute高级配置,然后单击自定义用户角色。
- 5. 在角色列表中, 单击admin角色对应的成员管理。
- 6. 在成员管理对话中,为目标用户添加admin角色,然后单击确定。

成员管理			
*角色名称:	admin		
添加成员:	待添加账号		已添加账号
	i 清绘入 Q ✔ ALIYUN\$\$    @ali RAM\$2	★	请输入 Q 】
	📒 1/6 项		1项
			确定 取消

# 如何按照日志时间进行分区?

您可以在创建MaxCompute投递作业(新版)时,将MaxCompute分区列设置为\_\_partition\_time\_\_,实现数据按照 日志时间进行分区。

* MaxCompute分区列:	partition_time	Θ	http_method			
	receive_time	Θ	operation	▣		
	+					
必须是在日志服务中存在的字段						

\_\_partition\_time\_\_字段是根据日志服务中\_\_time\_\_字段的值计算得到的,结合时区配置以及分区时间格式,生成时间字 符串。为避免触发MaxCompute单表分区数目的限制,日期分区列的值按照1800秒(半小时)对齐。更多信息,请参 见\_partition\_time\_字段。

#### 如何查看MaxCompute表的分区?

您可以通过 show partitions *[table-name]*; 命令查看MaxCompute表的分区。其中, *[table-name]*为表名称,请根据 实际值替换。

### 如何投递日志时间数据?

您可以在创建MaxCompute投递作业(新版)时,在MaxCompute普通列中添加\_\_time\_\_字段。用\_\_time\_\_字段表示 日志时间,UNIX时间戳格式,精确到秒。

* MaxCompute普通列:	content_length_in	Θ	content_length_in
	_time_	Θ	c14