



数据湖分析 最佳实践

文档版本: 20220630



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。	大) 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 结果确认 页面,单击 确定 。
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.基于SLS+OSS+DLA构建海量、低成本日志分析方案	05
2.友盟数据分析	80
3.DCDN日志分析	12
4.联合查询多个MySQL实例	17
5.漏斗分析	21
6.快速搭建DataV数据大屏	25
6.1. 背景信息以及准备工作	25
6.2. 实施步骤	30
7.Quick BI	36
7.1. Quick BI查询加速	36

1.基于SLS+OSS+DLA构建海量、低 成本日志分析方案

本文档主要介绍了如何基于SLS+OSS+DLA构建海量、低成本的日志分析方案。

方案介绍

对于数据分析人员、开发人员或者运维人员而言,日志数据对分析和诊断问题以及了解系统活动等有着非常 重要的作用,日志都是其工作过程中必不可缺的数据来源。为了节约成本,通常情况下日志会被设定一定的 保存时间,此类日志称之为热日志。这种做法短期内可以满足使用需求,但从长期来看,大量的历史日志被 丢弃,在需要分析和查询数据时,会出现数据丢失。

阿里云从用户角度出发,研发了一整套小而精的历史日志数据分析方案。利用阿里云日志服务(Log Service, SLS)来投递日志和查询热日志,利用阿里云对象存储服务(Object Storage Service, OSS)来持 久存储由SLS服务实时投递的日志,利用阿里云云原生数据湖分析(Data Lake Analytics, DLA)来分析存储 在OSS的日志。方案架构如下图所示:



该方案具有以下优势:

- SLS是针对实时数据的一站式服务,在阿里集团经历大量大数据场景锤炼而成。提供日志类数据采集、智能查询分析、消费与投递等功能,全面提升海量日志处理和分析能力。SLS强大的日志投递能力,能够从源头对接各种类型的日志格式,并且稳定地将日志投递到OSS。
- OSS低廉的存储成本,能够让您的日志文件存储任意长的时间。搭载SLS构建低成本冷热分离的存储方案。
- DLA提供强大的分析能力,可以对投递到OSS上的日志按年、月、日进行多维度的分区,提高日志的命中率,降低扫描量,从而以极低的成本、极高的性能来完成大数据量历史日志分析。

实施步骤

假设您的应用服务部署在ECS集群上,该集群每台机器上的应用服务都会产生日志数据,并通过SLS投递日志数据到OSS。通过DLA的元信息发现功能,自动发现SLS投递到OSS的日志数据,然后进行数据查询和分析。本文档将以此为例,为您介绍详细的实施步骤。

在开始实施步骤之前,您需要先完成以下准备工作:

- 准备一台ECS服务器,用于产生实时日志数据。或者准备一台已有日志数据的ECS。
- 开通日志服务、创建Project和Logstore。具体请参见快速入门。假设Project为*sls-oss-dla-test*, Logstore为*sls-dla-logstore*。

- 已经在OSS服务上创建存储空间。具体请参见创建存储空间。
- 开通数据湖分析服务。
 - 1. 在ECS服务器上面模拟服务日志的产生。
 - i. 在ECS服务器的目录/root/sls-crawler/下定义一个脚本文件gen-log.sh来产生日志数据。

```
[root@iZbpld6epzrizknw2xq**** sls-crawler]# cat gen-log.sh
#/bin/sh
filename=abc_`date +%s`.txt
echo ${filename}
for i in {1..1000000}
    do
        datatimestr=`date '+%Y-%m-%d %H:%M:%S'`
        echo "111111|1|100000000|0.1|0.00000000001|true|aabb|valueadd" >>/root/s
ls-crawler/full_type_logs/${filename}
        done
[root@iZbpld6epzrizknw2xq**** sls-crawler]#
```

⑦ 说明 在实际业务场景中,用户自己的应用服务器会产生日志文件,或者直接通过SLS的 API写入日志文件。

ii. 执行命令crontab -e并在该命令中添加如下配置来运行脚本文件*gen-log.sh*, 生成日志数据到/*ro ot/sls-crawler/full_type_logs/*目录。

```
* * * * * sleep 10; sh /root/sls-crawler/gen-log.sh
```

2. 通过Logtail采集ECS服务器上的日志数据到Logstore。具体操作请参见使用分隔符模式采集日志。

其中, ECS实例选择步骤一中的ECS实例*iZbp1d6epzrizknw2xq*****, 日志路径选择步骤一中的日志路径 /root/sls-crawler/full_type_logs/。

配置完成后,请您耐心等待几分钟,您就可以在SLS的**OSS投递管理**页面,查看日志数据是否成功接 入。

<	sis-oss-dia-test	切换	☆ OSS投递(sls ×	⊘ sis-dia-logstore ×	
0	日志库	我的关注	OSS投递管理 (属于	· sis-dia-logstore)	
Ð	搜索logstore	Q +	1小时 3小时 6小时	12小时 1天 2天	
~	/ Sis-ula-logsto		任务开始时间	任务结束时间	接受日志数据时间
11			2021-01-29 15:51:42	2021-01-29 15:51:48	2021-01-29 15:51:42
B			2021-01-29 15:52:45	2021-01-29 15:52:49	2021-01-29 15:52:45
G			2021-01-29 15:56:51	2021-01-29 15:56:56	2021-01-29 15:56:51
<u>B</u>			2021-01-29 15:57:51	2021-01-29 15:57:56	2021-01-29 15:57:51
₽			2021-01-29 16:01:54	2021-01-29 16:01:59	2021-01-29 16:01:54
ш			2021-01-29 16:02:53	2021-01-29 16:03:00	2021-01-29 16:02:53
			2021-01-29 16:06:57	2021-01-29 16:07:02	2021-01-29 16:06:57
85. 1979			2021-01-29 16:07:57	2021-01-29 16:08:01	2021-01-29 16:07:57
			2021-01-29 16:12:06	2021-01-29 16:12:13	2021-01-29 16:12:06
			2021-01-29 16:13:03	2021-01-29 16:13:07	2021-01-29 16:13:03
			重试全部失败任务		

3. 将日志服务数据投递到OSS。如何投递,请参见将日志服务数据投递到OSS。

其中, OSS Bucket 配置为 dla-crawler-hangzhou, OSS Prefix 配置为 sls-daily/shipper-json-full-type1 3。

等待SLS的投递任务启动成功后,您可以在对应的OSS目录查看投递的日志数据。

4. 通过DLA的元信息发现功能,自动发现SLS投递到OSS的日志数据,然后进行数据查询和分析。具体操作 请参见SLS的OSS投递数据源。

其中,数据源配置选择手动选择,logstore选择sls-dla-logstore。

2.友盟数据分析

U-DOP数据开放平台是友盟+为开发者提供的数据开放和私域数据融合的平台,通过一键订阅分析模板、拖 拽式自助分析报表来快速完成数据分析工作。U-DOP数据开放平台不仅仅为您提供了U-App的统计明细数 据,同时包含了多主题的分析模板和可订阅的数据包,通过任务引导式教您深度挖掘数据中的业务信息。您 可以通过阿里云数据湖分析DLA的数据订阅功能对原始明细log数据进行查询和分析。

开通友盟+DLA的组合方案

1. 登录<mark>友盟+数据银行控制台</mark>,在未开通服务页面的Data Lake Analytics区域,单击立即开通。

未开通服务		
■ MaxCompute+DataWorks 弹性计算 二维表格 您可以通过MaxCompute和DataWorks订阅友 盟+辅加工的明细数据,从而进行数据查询或融 合分析,MaxCompute为您提供快速、完全托管 的PB级数据仓库解决方案,经济并高效的批量分 析海量数据能力。配合DataWorks专业高效、安 全可靠的一站式大数据智能云研发,为您提供大	Data Lake Analytics Serverless	Quick Bi + Data Lake Analytics 交互式分析 可视化图表 自助式智能的服务解决方案,简单易用的可视化爆 作和灵活高效的多维分析能力,结合友盟+多主题 数据包(通过数据库DLA订阅数据),为您搭建了 丰富的行业惯板,帮助您实现用户增长。
了解产品立即开通	了解产品 立即开通	了解产品 立即开通
□ 日志服务(SLS) 实时查询 采集处理 您可以通过SLS订阅友盟+楠加工的明细数据,从 而进行实时查询或融合分析。与此同时,您可以享 用实时数据一站式服务,海量日志的数据采集、消 费、投递及查询分析能		
了解产品立即开通		

2. 输入需要订阅数据的友盟账号和密码,单击确认授权,进行阿里云账号三方登录的确认授权。

请登录友盟账号	阿里云 将获得以下权限:
名 友盟用户名	: 读取你的登陆状态信息及关联手机号
▲ 请输入登录密码	授权后表明您已同意 授权须知
确认授权	
忘记密码 还没有友盟账号, 点击注册	

3. 绑定友盟账号和阿里云账号。

如果没有阿里云账号可以快速注册阿里云账号,并和友盟账号进行绑定;如果已经有阿里云账号可以直 接输入阿里云账号和密码进行绑定。

当前科学篇次整局印度云、需要研一些编单设置、以便忽而期登录	首次登录设置	
快速注册 绑定已有阿里云账号 注册后,可继终使用用三方意思到调度示解联号 绑定后,忽可以通过第二方账号登录到减金的调里云账号 途也可以在账号中心设置登录密制后使用新账号 请输入阿里云账号	当前账号首次登陆阿里云,需要做一些简单设置,以便您后期登录	
请输入阿里云账号 =2	快速注册 注册后,可继续使用第三方登录到阿里云新账号 忽也可以在账句中心设置登录虚视后使用新账号	绑定已有阿里云账号 博迦斯,您可以通过第三方账号登录到博迪的问里云账号
2	请输入阿	卫 云姚号
		2
登录并绑定		型录并确定
☑ 同意(阿里云拆运协议)	() 3月 2	<u>显示规定协议</u>)

? 说明

友盟账号和阿里云账号绑定后,可以通过友盟账号登录到阿里云账号(第三方登录),请谨慎绑定,绑定后不支持更换账号操作。

绑定成功后,系统会提示您绑定电子邮箱,您可以根据实际情况选择立即绑定或者下次再说。

4. 在云资源访问授权页面,单击同意授权,授权U-DOP使用子账号角色来访问数据湖分析服务DLA。

云资源访问授权 如需修改角色权限,请前往 RAM 控制台角色管理中设置,需要注意的是,错误的配置可能导致 CloudMonitor 无法获取到必要的权限。
✓ MpuApp 请求获取访问您云资源的权限。
下方是系统创建的可供 MpuApp 使用的角色,授权后,MpuApp 拥有对您云资源相应的访问权限。
权限描述:用于移动统计的授权策略
同意授权 取消

5. 您需要购买数据湖分析流量包并完成支付,通过阿里云数据湖分析DLA产品来进行数据的查询和投递。 支付完成后请您不要关闭浏览器,等待开通成功的提示窗口。

Data Lake Analy	tics产品10TB+100G存储版原	价520元起,限时免存储费280g	元起!		配置清单	
基础信息 购买时长:	1年				数据湖分析流量包 1 元 节省 ¥279 倍 1元试用 购买时长: 1年	
数据湖分析流量 资源包类型:	包标准流量包					
地域:	华北 2(北京)				○元 节省 ¥279 立即购买	
流量包规格:	10TB	50TB	200TB	500TB		

查询数据

服务开通成功之后,等待一段时间您就可以看到友盟自动为您的阿里云账号创建了一个名为 integration_biz_ump_databank的数据库,这个数据库下面会有各种不同的数据明细表。您可以在DLA 的SQL执行页面进行数据查询了。如果您要查询某个App在2021年1月18日的相关日志明细,请参考以下 SQL:

select
umid, install_channel, app_version, app_key, is_new_install, install_datetime, app_chan
nel,
<pre>install_app_version, carrier, os, os_version, resolution, network_access, device_ip, country, province, city, session_start, session_end, session_id, session_duration, activity_cnt, is_new_upgrd, pre_channel, pre_app_version, is_new_device, imei, idfa, mac, android_id, idfv, open_udid, reg_app_name, device_model, std_device_model, device_</pre>
brand,
std_device_brand, provider, puid, is_new_reg, reg_channel, sdk_version, umid_from, cli_timestamp, svr_timestamp, log_date_time, oaid
from
(select
umid, install_channel, app_version, app_key, is_new_install, install_datetime, app_chan
nel,
<pre>install_app_version, carrier, os, os_version, resolution, network_access, device_ip, country, province, city, session_start, session_end, session_id, session_duration, activity_cnt, is_new_upgrd, pre_channel, pre_app_version, is_new_device, imei, idfa, mac, android_id, idfv, open_udid, reg_app_name, device_model, std_device_model, device_</pre>
brand,
<pre>std_device_brand, provider, puid, is_new_reg, reg_channel, sdk_version, umid_from, cli_timestamp, svr_timestamp, log_date_time, oaid from dwd_ump_log_uapp_launch_di where ds = '20210118' and app_key = 'xxxxxx' and is_new_install = 1 order by cli_timestamp desc) a group by install_channel,install_app_version,umid;</pre>

3.DCDN日志分析

全站加速DCDN日志分析是结合对象存储OSS和全站加速DCDN为一体的日志数据分析和查询服务。您可以在 全站加速DCDN控制台开通离线日志转存服务,将使用全站加速DCDN过程中产生的日志自动转存到您的OSS 上进行长期管理,通过DLA Serverless Presto对日志进行分析。本文档主要介绍了如何通过DLA Serverless Presto来进行全站加速DCDN日志的数据分析和查询。

前提条件

进行全站加速DCDN日志分析前,需确保您已经开通了对象存储OSS和全站加速DCDN离线日志转存服务。开 通详情如下:

- 开通对象存储OSS,请前往对象存储OSS产品详情页进行开通。
- 开通全站加速DCDN离线日志转存服务,请登录全站加速控制台进行开通。具体请参见通过数据湖转存和 分析离线日志

背景信息

您在开通全站加速DCDN离线日志转存服务之后,系统会自动开通离线日志转存区域的DLA服务,并且DLA会 自动创建好离线日志转存表。

操作步骤

本文档中假设开通了全站加速DCDN离线日志转存服务的日志转存到了上海。

- 1. 登录Data Lake Analytics管理控制台。
- 2. 在页面左上角,下拉选择区域(此处选择上海)。如下图所示:

∃ (-)阿里云	华东1(杭州) ▲	
云原生数据湖分析	亚太	欧洲与美洲
	💴 华东1 (杭州)	📕 德国 (法兰克福)
概览	- 华东2 (上海)	🚟 英国 (伦敦)
账号管理	▲ 华北1 (青岛)	📕 美国 (硅谷)
虚拟集群管理	华北2 (北京)	🗮 美国 (弗吉尼亚)
数据湖管理 HOT	华北3 (张家口)	
元信息发现	▲ 华北5 (呼和浩特)	中东与印度
元数据管理	416 (乌兰察布)	🔽 印度 (孟买)
数据入湖	华南1 (深圳)	🔚 阿联酋 (迪拜)
实时数据湖	华南2 (河源)	
Secondara SOI	华南3 (广州)	
Serveriess SQL	西南1 (成都)	
SQL访问点		
SQL执行	◎ 新加坡	
SQL监控	🎦 澳大利亚 (悉尼)	
Serverless Spark	🛄 马来西亚 (吉隆坡)	
作业管理	━ 印度尼西亚 (雅加达)	
优惠套餐	• 日本 (东京)	

3. 在左侧导航栏,单击Serverless Presto > SQL执行。

云原生数据湖分析		SQL 执行				
概览						
账号管理		搜索 Schema (9	SQL执行集群	public(共享集群)	~
虚拟集群管理		"双击"切换Schema		1		
数据湖管理 HOT	^	🔰 🚍 alicdn_offline_log (curre	nt)			
元信息发现		> 🚍 public_dataset_tpch_1x_p	arquet			
元数据管理						
数据入湖						
实时数据湖						
Serverless SQL	^					
SQL访问点						
SQL执行		1				
SQL监控						
Serverless Spark	^			执行历史	执行结果	SQL监控
作业管理				0	Os	show full
优惠赛餐					1s	show full
系统设置						

4. 在SQL执行页面,执行SQL语句来进行DCDN日志数据的查询和分析。示例如下:

○ urlPath访问错误top10

//先通过HTTP状态码(httpcode)进行过滤,一般认为大于400的状态码都是有问题的。然后根据请求的URL, 不含域名信息 (urlPath) 聚合,获取各个请求urlPath对应的错误数量,根据数量倒序排列,取前十个。 SELECT urlPath,count(*) as c from alicdn_offline_log.alicdn_offline_log_dcdn WHERE ht tpcode > 400 GROUP BY urlPath ORDER BY c DESC LIMIT 10;

搜索 Schema								
	C so	LL执行集群 public(共享集群)	✓ 同步执行(F8)	异步执行(F9) 格式化(F10)	主题 >		新建SQL虚拟集群	() 登录DMS执行SQL
"双击"切换Schema		1 SELECT urlPath, count(*) :	as c from alicdn_offline_1	og.alicdn_offline_log_dcdn WHERE	httpcode $>$ 400 GROUP BY urlPath	ORDER BY & DESC LINIT 10:		
🗠 🚍 alicdn_offline_log								
✓ ■ alicdn_offline_log	g_dcdn							
contenttype								
🔳 domain								
i hitinfo								
http2								
📰 httpcode								
respsize								
reqsize								
i remoteip								
🔳 refer	<u>.</u>							
i method	1							
🔳 rt								
schema 🖬								
ua ua								
🔳 unixtime								
📰 urlpath								
urlrawquery								
🔳 ds								
batch_id								
> gublic_dataset_tpch;	_1x_parquet							
> E customer		执行历史 执行结果	SOL ##					法果事 く 協志
> 🖬 lineitem			•					
> mation		序号	urlPath		c			detail
> orders		1	/test.html		11			详情

○ 访问来源占比分析

```
//首先通过域名信息(domain)过滤条件去掉访问来源为空的日志。通过聚合计算各个访问源的请求次数,后
面再计算出单个域名信息(domain)请求次数在所有请求中的占比,并保留两位小数。
select
 refer domain as "来源",
 c as "次数",
 round(c * 100.0 / (sum(c) over()), 2) as "百分比%"
from
 (
   select
    domain as refer domain,
    count(*) as c
   from
    alicdn_offline_log.alicdn_offline_log_dcdn
   group by
    refer_domain
   order by
    c desc
   limit
    100
 )
 where refer_domain != "";
```

○ 访问次数Top域名

//用group by根据Domain分组,用count计算每个域名出现的次数,用order by按次数进行排序,用desc表示从大到小的顺序,limit 100表示取前100个,用饼图进行展示,分类设置为return_code,数值列设置为c。 select

```
Domain,
count(*) as c
FROM
alicdn_offline_log.alicdn_offline_log_dcdn
group by
Domain
order by
c desc
limit
100
```

SQL执行									语法手册
搜索 Schema C	SQL执行集群 public[共享集团	判 > 同步执行(8) 异步执行(F9)	格式化(F10)	主題 イ		[新建SQL虚拟集群	1 登录DMS执行SQL
Y24: 1085/chema > ■ alich, etfine, log, don ■ alich, etfine, log, don ■ donain ■ hindo ■ resplate ■ resplate ■ netbod ■ unistime ■ unistime ■ unistime ■ unistime ■ unistime ■ lanch, d > ■ public, datast, tph, th, parguet > ■ public, datast, tph, th, parguet	<pre>1 relect 2 refer_domin as 7 3 e eas 7,830, 4 round/e = 10.0 / 0 5 from 0 (7 select as re 0 domin as re 0 for would) as 11 aliech.off1 12 group by 13 refer_domin 14 erder by 15 e dere 16 limit 19 where refer_domin</pre>	語。 man (c) over (), 2) as "百分 fer_demain, c n_lor, aliceh_offline_lor_d n != **;	tts″						
> 📷 lineitem	执行历史 执行结果	SQL监控						Ş	出结果集 > 隐蔵
> 🖬 nation > 📷 orders > 📼 nart	序号 1	來源 zuoshang1.zs.test.com		次数 11		百分 100.0	9 1		detail 详情

○ 数据类型流量分布

//用sum函数计算respsize的总和作为sum_res,用group by按content_type分组,用order by按sum_res s排序,取最大的前10个,用饼图展示,分类设置为content_type,数值列设置为sum_res。

```
select
  contenttype,
  sum(respsize) as sum_res
FROM alicdn_offline_log.alicdn_offline_log_dcdn
group by
  contenttype
order by
  sum_res desc
limit
  10;
```

SQL执行					潘法手册 函数手册
搜索 Schema C	SQL执行集群 public(共享集群)	✓ 同歩执行(F8) 局歩执行(F9)	格式化(F10) 主题 >	象标题SQL	籃奴集群 (j) 登录DMS执行SQL
ULT UBS-hema > ■ sinde_Uffice_Log = sinde_Uffice_Log = contentype = domain = hiningo = hiningo = respise = respi	1 select 2 contentype, s pug_res 3 mon temple) s pug_res 5 promo by guffling_log_alcd 5 contentype 8 swgreedes 9 limit 10 l0:	n_offline_log_doth			
S public_dataset_tpch_lx_parquet S customer	执行历史 执行结果 SQL				
> main nation	序号	contenttype		sum_res	detail
> = out	1	text/plain; charset=UTF-8		8681	详情

4.联合查询多个MySQL实例

本文档将以两个云数据库RDS MySQL版(简称MySQL)为例,介绍如何通过DLA联合查询多个MySQL实例数据。

背景信息

Data Lake Analytics(简称DLA)作为云上数据处理的枢纽,支持通过标准JDBC对单个实例RDS、Table Store、OSS、MongoDB中的数据进行查询和分析。随着业务的扩展或者数据的划分,您可能会创建多个 RDS、Table Store、OSS等实例存储数据。在这种多数据源场景下,仍然可以通过DLA实现多个相同类型数 据源的联合查询。

前提条件

通过DLA联合查询两个MySQL实例数据前,您需要通过以下操作在两个MySQL实例中准备好测试数据

↓ 注意

- DLA和两个MySQL所属地域必须相同,否则无法进行本文档操作。
- 由于DLA将通过MySQL的VPC连接MySQL数据库,建议您创建MySQL实例时,网络类型选择VPC。
 同时,MySQL支持通过切换为专有网络将经典网络切换为VPC。
- 1. 创建两个MySQL实例
- 2. 设置白名单
- 3. 连接实例
- 4. 写入测试数据
 - MySQL实例1中创建orders_db数据库和order_item表,写入测试数据:

```
create table orders_db.order_item (
    id bigint not null primary key auto_increment,
    prod_id bigint comment '商品数量'
);
insert into orders_db.order_item values
  (1, 1, 2),
  (2, 2, 3),
  (3, 3, 4),
  (4, 2, 5),
  (5, 1, 6);
```

○ MySQL实例2中创建prod_db数据库和prod表,写入测试数据:

```
create table prod_db.prod (
   id bigint not null primary key auto_increment,
   prod_name varchar(31) comment '商品名称'
);
insert into prod_db.prod values
  (1, '键盘'),
  (2, '鼠标'),
  (3, '显示器');
```

实施步骤

注意事项

在DLA中创建MySQL数据库连接前,需要将IP地址段 100.104.0.0/16 分别加入到MySQL白名单列表中。

由于您的MySQL实例位于VPC内,默认情况下DLA无法访问该VPC中的资源。为了让DLA访问MySQL,需要利用VPC反向访问技术,即在MySQL白名单中添加 100.104.0.0/16 IP地址段。

? 说明

权限声明:当您在MySQL白名单中添加了 100.104.0.0/16 IP地址段,即视为您同意我们利用VPC反向 访问技术读取MySQL数据库数据。

步骤一:在DLA中创建MySQL数据库连接

登录DLA控制台,单击左侧导航栏的SQL访问点,单击登录到DMS,分别在DLA中创建两个底层映射到 MySQL的数据库连接。

```
CREATE SCHEMA dla orders db WITH DBPROPERTIES (
 CATALOG = 'mysql',
 LOCATION = 'jdbc:mysql://rm-*****.mysql.rds.aliyuncs.com:3306/orders db',
 USER = 'mysql_db_user_name',
 PASSWORD = 'mysql db password',
 VPC ID = 'mysql vpc id',
 INSTANCE ID = 'mysql instance id'
);
CREATE SCHEMA dla prod db WITH DBPROPERTIES (
 CATALOG = 'mysql',
 LOCATION = 'jdbc:mysql://rm-*****.mysql.rds.aliyuncs.com:3306/prod db',
 USER = 'mysql_db_user_name',
 PASSWORD = 'mysql db password',
 VPC_ID = 'mysql_vpc_id',
 INSTANCE ID = 'mysql instance id'
);
```

参数说明

• LOCATION: 由 jdbc:mysql://MySQL内网连接地址:端口号/MySQL数据库名 构成。

基本信息	·	
账号管理	基本信息	设置白名单 迁移可用区
数据库管理	实例ID: m9n3ee	名称: 1
数据库连接	地域可用区: 华东1(杭州)可用区H	类型及系列:常规实例 (高可用版)
数据库代理	内网地址: rm- mysql.rds.aliyuncs.com	内网端口: 3306

- USER: 连接MySQL数据库所使用的用户名。
- PASSWORD: 连接MySQL数据库所使用的用户名对应的密码。
- VPC_ID: MySQL实例所属VPC ID。

● INSTANCE_ID: MySQL实例ID。

基本信息	▲ 数据库连接
账号管理	
数据库管理	实例连接 读写分离
数据库连接	数据库连接
数据库代理	网络类型: 专有网络 (VPC:vpc-izd6rwt) 2
监控与报警	内网地址: rm-b

步骤二:在DLA中创建MySQL外表

针对MySQL中order_item表,本示例在DMS for Data Lake Analytics中创建同名外表order_item:

```
create external table order_item (
    id bigint,
    prod_id bigint,
    prod_cnt int,
```

);

针对MySQL中prod表,本示例在DMS for Data Lake Analytics中创建同名外表prod:

```
create external table prod (
    id bigint,
    prod_name varchar(31)
);
```

步骤三:通过DLA联合查询多个MySQL实例数据

MySQL数据库连接和外表创建成功后,接下来您可以通过MySQL客户端或者MySQL命令行工具连接DLA,使用标准SQL语句操作MySQL数据库数据。

也可以直接在DMS for Data Lake Analytics中操作MySQL数据库数据。

以下示例通过MySQL命令行工具连接DLA,联合查询order_item表和prod表数据:

```
mysql> select * from dla_orders_db.order_item;
+----+
| id | prod id | prod cnt |
+----+
| 1 | 1 | 2 |
| 2 |
         2 |
                 3 |
| 3 |
         3 |
                 4 |
| 4 | 2 | 5 |
| 5 | 1 | 6 |
+----+
5 rows in set (0.22 sec)
mysql> select * from dla_prod_db.prod;
+----+
| id | prod name |
+----+
| 1 | 键盘 |
| 2 | 鼠标 |
| 3 | 显示器 |
+----+
3 rows in set (0.18 sec)
select p.prod name, sum(o.prod cnt) cnt
 from dla_prod_db.prod p inner join dla_orders_db.order_item o
  on p.id = o.prod id
  group by p.prod_name
  order by cnt desc;
+----+
| prod name | cnt |
+----+
|显示器| 4|
+----+
3 rows in set (0.55 sec)
```

后续操作

我们通过DLA联合查询多个MySQL实例数据后,可以通过DLA将查询到的数据存入阿里云对象存储服 务(Object Storage Service,简称 OSS)。OSS是阿里云提供的海量、安全、低成本、高可靠的云存储服 务。后续需要这部分数据时,DLA可以在不移动OSS文件的情况下,直接读取数据。

5.漏斗分析

DLA漏斗分析旨在帮助运营人员分析一个多步骤操作过程中每一步的转化与流失情况。假设购买商品过程 中,需要触发一系列事件:启动>登录>搜索商品>查看商品>生成订单等。运营人员分析某段时间内(例 如,2017年01月05号~2017年02月05号)所有依次触发启动>登录>搜索商品>查看商品>生成订单事件 的人群的转化流失情况,即计算所有用户中触发登录事件的总人数A、A中触发搜索商品事件的总人数B、B 中触发查看商品事件的总人数C以及C中触发生成订单事件的总人数D。



漏斗分析中包含时间窗口的概念,即需要保证所有事件在同一个窗口期内触发。例如时间窗口为1天,用户 001触发搜索商品事件的时间和触发登录事件的时间间隔在一天内,搜索商品事件才有效,否则视为无效。 同理,用户001触发查看商品事件的时间和触发登录事件的时间间隔也必须在一天内。时间窗口可以为1天、 3天、7天或者1小时、6小时等任意长时间段。

前提条件

本文示例中的测试数据存储在OSS中,您可以按照以下步骤将您的测试数据存储在OSS中。

- 1. 开通OSS服务。
- 2. 创建存储空间。
- 3. 上传文件。

本文示例中将以下测试数据上传到OSS的 funnel_data 目录。

1000002	1483258815538	收藏商品 {} 20170101
1000002	1483274981790	启动 {} 20170101
1000002	1483223087508	搜索商品 {"content": "computer", "page_num": 1} 20170101
1000002	1483232016805	搜索商品 {"content": "Apple", "page_num": 2} 20170101
1000002	1483200895552	订单付款 {} 20170101
1000003	1483218002826	搜索商品 {"content": "computer", "page_num": 2} 20170101
1000003	1483206471681	加入购物车 {} 20170101
1000003	1483284553531	加入购物车 {} 20170101
1000003	1483279891663	浏览商品 {"brand": "Apple", "price": 9500} 20170101
1000003	1483259182702	加入购物车 {} 20170101
1000004	1483260505099	登录 {} 20170101
1000004	1483231995064	收藏商品 {} 20170101
1000004	1483206318588	加入购物车 {} 20170101
1000004	1483241973408	登录 {} 20170101
1000004	1483202785557	收藏商品 {} 20170101
1000005	1483214265416	收藏商品 {} 20170101
1000005	1483206343383	收藏商品 {} 20170101
1000005	1483229265488	浏览商品 {"brand": "HuaW", "price": 9500} 20170101
1000005	1483244294552	生成订单 {"price_all": 5000} 20170101
1000005	1483246988534	登录 {} 20170101
1000006	1483207838307	收藏商品 {} 20170101

```
      文件名 (Object Name )
      文件大小 存储类型

      5
      / funnel_data/

      1
      funnel.txt

      1.231KB
      标准存储
```

开始使用漏斗分析之前,您需要在DLA中完成以下准备工作。

1. 开通云原生数据湖分析服务。

2. 重置数据库账号密码。

步骤一: 创建OSS Schema

```
CREATE Schema funnel_test_schema
WITH DBPROPERTIES (
    catalog = 'oss',
    location = 'oss://bucket-name/'
)
```

步骤二: 创建表

```
CREATE EXTERNAL TABLE IF NOT EXISTS funnel_test (
    user_id bigint NOT NULL COMMENT '',
    event_time bigint NOT NULL COMMENT '',
    event_id int NOT NULL COMMENT '',
    event_attr varchar NOT NULL COMMENT '',
    event_date date NOT NULL COMMENT '',
    event_date date NOT NULL COMMENT '',
    event_date date NOT NULL COMMENT '',
    EVENT_ATT VARCHAR NOT NULL COMMENT '',
    EVENT_ATT DELIMITED FIELDS TERMINATED BY '\t'
STORED AS TEXTFILE
LOCATION 'oss://bucket-name/funnel data/';
```

步骤三:使用漏斗分析函数进行数据分析

按照启动 > 登录 > 搜索商品 > 查看商品 > 生成订单的事件触发顺序,对应的事件ID为10001 > 10002 > 10003 > 10004 > 10007。

● 以下示例通过 funnel_count 函数,查询1月1号到20号一共20天、时间窗口为7天(单位毫秒)、事件 个数为5个的漏斗,同时能够保存路径各节点的人群。

```
SELECT user_id,
funnel_count(event_time,7 * 86400000,event_id,
'10001,10002,10003,10004,10007') AS xwho_state
FROM funnel_test
WHERE event_id IN (10001, 10002, 10003, 10004, 10007)
AND event_date BETWEEN '2017-01-01' AND '2017-01-20'
GROUP BY user_id;
```

• 以下示例在 funnel_count 函数的基础上使用 funnel_sum 函数,得出定义转化路径的总体转化率。

```
SELECT funnel_sum(xwho_state, 5)
FROM (
    SELECT user_id,funnel_count(event_time,7 * 86400000, event_id,
    '10001,10002,10003,10004,10007') AS xwho_state
    FROM funnel_test
    WHERE event_id IN (10001, 10002, 10003, 10004, 10007)
        AND event_date BETWEEN '2017-01-01' AND '2017-01-20'
    GROUP BY user_id
);
```

● 如果需要对事件对应的JSON属性列 <u>event_attr</u>的某个属性进行过滤,例如对10004类事件中price在 3500到5000之间的数据进行过滤。

```
SELECT funnel_sum(xwho_state, 5)
FROM (
    SELECT user_id,funnel_count(event_time, 7 * 86400000,event_id,
    '10001,10002,10003,10004,10007') AS xwho_state
    FROM funnel_test
    WHERE event_id IN (10001, 10002, 10003, 10004, 10007)
        AND event_date BETWEEN '2017-01-01' AND '2017-01-20'
        OR (event_id = 10004 AND
            json_extract_scalar(event_attr, '$.price') BETWEEN 3500 AND 5000)
    GROUP BY user_id
);
```

上述路径和事件通过ID进行表征,实际场景中可能只有事件的名称,而没有事件对应的ID,目前DLA也支持 直接使用事件名称(本例中的 event_name 列)进行路径事件的表征。

```
SELECT funnel_sum(xwho_state, 7) AS funnel
FROM (SELECT user_id,funnel_count(event_time, 7 * 86400000,event_name,'启动,登录,搜索商品,浏览
商品,生成订单,订单付款,评价商品') AS xwho_state FROM funnel_test
WHERE event_name IN
('启动', '登录', '搜索商品', '浏览商品', '生成订单', '订单付款', '评价商品')
GROUP BY user id);
```

6.快速搭建DataV数据大屏 6.1.背景信息以及准备工作

DataV数据可视化(简称DataV)旨在让更多的人看到数据可视化的魅力,并帮助非专业的工程师通过图形 化的界面轻松搭建具有专业水准的可视化应用。DataV提供了丰富的可视化模板,极大程度满足您会议展 览、业务监控、风险预警、地理信息分析等多种业务的展示需求。与使用传统数据库加DataV构建大屏相 比,在绝大部分频率较低的数据刷新业务大屏场景下,DataV+DLA+OSS的方式远比DataV+传统数据库的成 本低。

实施方案

实际应用中DataV+DLA+OSS的搭配使用有以下两种方案。

DLA+OSS+DataV

业务应用程序产出的增量数据上传到OSS,通过定时查询,刷新到业务数据大屏中。

• LOG+OSS+DLA+DataV

将业务应用程序产出的业务日志采集到日志服务中,通过定时日志投递(最小5分钟延迟)到OSS中,再 通过定时查询,刷新到业务数据大屏中。



上述方案有以下优势:

- 日志服务(Log Service,简称LOG/原SLS)是针对实时数据一站式服务,在阿里集团经历大量大数据场景 锤炼而成。提供日志类数据采集、智能查询分析、消费与投递等功能,全面提升海量日志处理/分析能 力。LOG强大的日志投递能力,能够从源头对接各种类型的日志格式,并且稳定地将日志投递到指定的位 置。
- 阿里云对象存储服务(Object Storage Service,简称 OSS)低廉的存储成本,能够让您的数据文件存储 任意长的时间。
- Data Lake Analytics (简称DLA)强大的分析能力,Serverless的架构,按扫描量收费。DLA可以对投递到 OSS上的数据按年、按月、按日进行多维度的分区,提高数据的命中率,降低扫描量,从而以极低的成 本、极高的性能来完成大数据量历史数据分析。
- DataV中有多种场景模板,解决您的设计难题。提供多种业务模块级别而非图表组件的工具,所见即所得的配置方式,无需编程能力,您只需要通过拖拽,即可创造出专业的可视化应用。创建的可视化应用能够发布分享,没有购买DataV产品的用户也可以访问到应用,作为您对外数据业务展示的窗口。

本教程以DLA+OSS+DataV方案为例,介绍如何制作DataV数据大屏。

前提条件

使用DLA+OSS+DataV制作数据大屏时,您需要先通过以下操作在OSS中准备测试数据、在DLA中创建OSS Schema和表。

步骤一:在OSS中准备测试数据

- 1. 开通OSS服务
- 2. 创建存储空间
- 3. 上传文件

单击下载TPC-H100MB的数据集,数据集中共有8个数据文件,在OSS中为每个数据文件创建对应的文件 夹,并上传数据文件。

⑦ 说明 示例使用的数据集引用自TPC。	更多信息,	请参见TPC。
----------------------	-------	---------

ille se fil	int .	
概览	文件管理 基础设置 域名管理 图片处理 事件通知 函数计算	智能媒体
上传文件	新建目录 碎片管理 授权 批量操作 > 刷新	
	文件名 (Object Name)	文件大小
5	/ TPC-H/	
• 🦻	customer/	
• 📄	lineitem/	
• 🦻	nation/	
	orders/	
• 🦻	part/	
• 📄	partsupp/	
• 🫅	region/	
	supplier/	

http://bucket-name.oss-cn-hangzhou.aliyuncs.com/TPC-H/customer/customer.tbl
http://bucket-name.oss-cn-hangzhou.aliyuncs.com/TPC-H/lineitem/lineitem.tbl
http://bucket-name.oss-cn-hangzhou.aliyuncs.com/TPC-H/nation/nation.tbl
http://bucket-name.oss-cn-hangzhou.aliyuncs.com/TPC-H/part/part.tbl
http://bucket-name.oss-cn-hangzhou.aliyuncs.com/TPC-H/partsupp/partsupp.tbl
http://bucket-name.oss-cn-hangzhou.aliyuncs.com/TPC-H/region/region.tbl
http://bucket-name.oss-cn-hangzhou.aliyuncs.com/TPC-H/partsupp/partsupp.tbl
http://bucket-name.oss-cn-hangzhou.aliyuncs.com/TPC-H/partsupp/partsupp.tbl
http://bucket-name.oss-cn-hangzhou.aliyuncs.com/TPC-H/partsupp/partsupp.tbl
http://bucket-name.oss-cn-hangzhou.aliyuncs.com/TPC-H/region/region.tbl

步骤二: 创建OSS Schema

```
CREATE SCHEMA tpch with DBPROPERTIES(
  LOCATION = 'oss://bucket-name/TPC-H/',
  catalog='oss'
);
```

步骤三: 创建OSS表

● nation表

```
CREATE EXTERNAL TABLE nation (

N_NATIONKEY INT,

N_NAME STRING,

N_ID STRING,

N_REGIONKEY INT,

N_COMMENT STRING

)

ROW FORMAT DELIMITED FIELDS TERMINATED BY '|'

STORED AS TEXTFILE

LOCATION 'oss://bucket-name/TPC-H/nation';
```

• lineitem表

```
CREATE EXTERNAL TABLE lineitem (
   L ORDERKEY INT,
   L PARTKEY INT,
   L SUPPKEY INT,
   L_LINENUMBER INT,
   L QUANTITY DOUBLE,
   L EXTENDEDPRICE DOUBLE,
   L DISCOUNT DOUBLE,
   L TAX DOUBLE,
   L RETURNFLAG STRING,
   L LINESTATUS STRING,
   L SHIPDATE DATE,
   L COMMITDATE DATE,
   L RECEIPTDATE DATE,
   L SHIPINSTRUCT STRING,
   L SHIPMODE STRING,
   L COMMENT STRING
)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '|'
STORED AS TEXTFILE
LOCATION 'oss://bucket-name/TPC-H/lineitem';
```

● orders表

```
数据湖分析
```

```
CREATE EXTERNAL TABLE orders (

O_ORDERKEY INT,

O_CUSTKEY INT,

O_ORDERSTATUS STRING,

O_TOTALPRICE DOUBLE,

O_ORDERDATE DATE,

O_ORDERPRIORITY STRING,

O_CLERK STRING,

O_CLERK STRING,

O_COMMENT STRING

)

ROW FORMAT DELIMITED FIELDS TERMINATED BY '|'

STORED AS TEXTFILE

LOCATION 'oss://bucket-name/TPC-H/orders';
```

• supplier表

```
CREATE EXTERNAL TABLE supplier (

S_SUPPKEY INT,

S_NAME STRING,

S_ADDRESS STRING,

S_NATIONKEY INT,

S_PHONE STRING,

S_ACCTBAL DOUBLE,

S_COMMENT STRING

)

ROW FORMAT DELIMITED FIELDS TERMINATED BY '|'

STORED AS TEXTFILE

LOCATION 'oss://bucket-name/TPC-H/supplier';
```

• part supp表

```
CREATE EXTERNAL TABLE partsupp (

PS_PARTKEY INT,

PS_SUPPKEY INT,

PS_AVAILQTY INT,

PS_SUPPLYCOST DOUBLE,

PS_COMMENT STRING

)

ROW FORMAT DELIMITED FIELDS TERMINATED BY '|'

STORED AS TEXTFILE

LOCATION 'oss://bucket-name/TPC-H/partsupp';
```

```
    customer表
```

```
CREATE EXTERNAL TABLE customer (

C_CUSTKEY INT,

C_NAME STRING,

C_ADDRESS STRING,

C_NATIONKEY INT,

C_PHONE STRING,

C_ACCTBAL DOUBLE,

C_MKTSEGMENT STRING,

C_COMMENT STRING
)

ROW FORMAT DELIMITED FIELDS TERMINATED BY '|'

STORED AS TEXTFILE

LOCATION 'oss://bucket-name/TPC-H/customer';
```

● part表

```
CREATE EXTERNAL TABLE part (

P_PARTKEY INT,

P_NAME STRING,

P_MFGR STRING,

P_BRAND STRING,

P_TYPE STRING,

P_SIZE INT,

P_CONTAINER STRING,

P_RETAILPRICE DOUBLE,

P_COMMENT STRING
)

ROW FORMAT DELIMITED FIELDS TERMINATED BY '|'

STORED AS TEXTFILE

LOCATION 'oss://bucket-name/TPC-H/part';
```

• region表

```
CREATE EXTERNAL TABLE region (

R_REGIONKEY INT,

R_NAME STRING,

R_COMMENT STRING
)

ROW FORMAT DELIMITED FIELDS TERMINATED BY '|'

STORED AS TEXTFILE

LOCATION 'oss://bucket-name/TPC-H/region';
```

建表完毕后,刷新页面,可以查看Schema下的8张表。

对象列表 刷新 🕜
🕂 🚞 dataworks_demo
🖶 🛅 dla_orders_db (1)
🕂 🚞 dla_oss_db
🕂 🚞 dla_oss_demo
🕂 🚞 dla_prod_db
🕂 🚞 hangzhou_ots_test
🖶 🚞 hello_mysql_vpc_rds
🕂 🚞 my_schema_name
🕂 🛅 oss_schema
🕂 🛅 oss_test_schema
🖶 🚞 porlardb_test
🕂 🛅 public_dataset_tpch_1x_text
🖶 🚞 redis_test
🕂 🚞 test_pivot_unpivot
😑 🚞 tpch (8)
💼 customer
📺 lineitem
- 📺 nation
- 💼 orders
📷 part
📰 partsupp
- 📺 region
supplier 📷

下一步

实施步骤

6.2. 实施步骤

本文主要介绍搭建DataV数据大屏的操作步骤。

步骤一:购买DataV基础版

登录DataV管理控制台,本教程使用兼容MySQL方式连接DLA服务,所以购买DataV基础版即可满足要求。

步骤二:添加DLA数据源

1. 在添加数据源之前,您必须先为您的DLA配置白名单,以确保DataV能正常访问您的DLA。

• DLA外网公共白名单: DLA外网所有region都需要配置以下公共白名单。

139.224.92.81/24,139.224.92.22/24,139.224.92.35/24,139.224.4.30/24,139.224.92.102/24, 139.224.4.48/24,139.224.4.104/24,139.224.92.11/24,139.224.4.60/24,139.224.92.52/24,13 9.224.4.26/24,139.224.92.57/24,112.74.156.111/24,120.76.104.101/24,139.224.4.69/24,11 4.55.195.74/24,47.99.11.181/24,47.94.185.180/24,182.92.144.171/24 ○ 在外网公共白名单的基础上,根据DLA所属region添加以下对应的白名单。

华北3(张家口): 47.92.22.210/24,47.92.22.68/24 华北1(青岛): 118.190.212.44/24 华北5(呼和浩特): 39.104.29.35/24 关于DataV白名单详情,请参见配置数据库白名单。

- 2. 登录DataV管理控制台,单击我的数据。
- 3. 在我的数据页面,单击添加数据。

我的可视化 总 我的数据 总 我的组件	由优秀案例&教程
+ 添加数据	
G SDK datav_proxy	2018/12/5 上午10:44:38
MySQL test	2018/10/18 下午5:40:34
MySQL rdsformysql_datav	2018/9/20 上午9:53:38

4. 在添加数据页面,填写DLA数据库信息。

添加数据		×
*类型		查看数据源文档
兼容MySQL数据库		-
*名称		
dla_hangzhou_tpch		
*域名		
.cn-hangzhou	u.datalakeanalytics.aliyuncs	.com
*用户名		
an and the second second		
· *密码		
*端口		
10000		
*数据库		
获取数据列表	tpch	
		从列表中选择数据库

- 类型: DLA使用的数据源类型是兼容MySQL数据库。
- 名称: 数据源的显示名称, 您可以自由命名。
- **域名**: DLA的外网连接地址(经典网络地址),如何获取外网连接地址,请参见创建服务访问点。
- **用户名**:登录DLA所使用的账号。
- 密码: 登录DLA所使用的账号对应的密码。
- 端口: DLA的外网端口为10000。
- 数据库:手动输入DLA Schema的名称,本教程使用的Schema名为tpch。
- 5. 单击**确定**,系统自动进行测试连接,验证数据库是否能连通正常。数据源添加成功后,新添加的数据源 自动显示在数据源列表中。

● 我的可视化	回优秀案例&教程
+ 添加数据	
MySQL	2019/4/18 下午2:52:22
MySQL	2018/10/24 下午1:40:48
MySQL	2018/9/21 上午10:59:18

步骤三: 创建可视化应用

- 1. 登录DataV管理控制台。
- 2. 单击我的可视化。
- 3. 此实例选择PC端创建方式进行创建。

\$ 我的可视化	Ø 1	我的数据	皋 我的资产	合 教程					
我的分组	+	选择下面的方式	达进行创建	_			Al Vislab		
全部应用			PC端创建		移动端创建	AI	识图创建		
未分组									
		全部应用 0	个/还可创建20个 🕜					۹	按修改时间排)

- 4. 选择销售实时监控,单击创建项目。
- 5. 输入数据大屏名称,单击创建。

步骤四:修改大屏样式

1. 本教程大屏的目标是显示各个国家的销售数据情况,需要把下面红框内的组件删除,并调整大屏。



- 为了显示各个国家的销售数据情况,需要使用世界地图。删除现有的中国地图组件,然后在导航栏的地图中添加3D平面世界地图。
- 3. 将销售额分省排名组件的名称修改为销售额分国家排名。

步骤五:为大屏中的地图组件配置数据

- 1. 选中步骤四中添加的世界地图组件,在数据页面进行参数配置。
 - 数据源类型:选择数据库。
 - 数据源:选择步骤二中添加的DLA数据源(dla_hangzhou_tpch)。
 - SQL: 输入以下SQL, 按国家销售额降序排序。

```
select sum(o_totalprice) total_price, n_nationkey, n_name, n_id
from orders
join customer on o_custkey = c_custkey
join nation on c_nationkey = n_nationkey
group by n_nationkey, n_name, n_id
order by total price desc;
```

- 。字段id和value:分别填入上述SQL中返回的n_id列和total_price列。
- 根据业务数据更新需求,选择大屏数据自动更新请求,比如60秒一次。
- 2. 完成上述参数配置后,单击刷新数据。

步骤六:为大屏中的销售总额组件配置数据

- 1. 选中销售总额组件,在数据页面进行参数配置。
 - 数据源类型:选择数据库。
 - 数据源:选择步骤二中添加的DLA数据源(dla_hangzhou_tpch)。
 - SQL: 输入以下SQL, 查询销售总额。

```
select sum(o_totalprice) total_price
from orders;
```

- 。字段value,填入上述SQL返回的total_price列。
- 根据业务数据更新需求,选择大屏数据自动更新请求,比如60秒一次。

2. 完成上述参数配置后,单击刷新数据。

步骤七:为大屏中的国家销售额排名组件配置数据

- 1. 选中销售额国家排名组件,在数据页面进行参数配置。
 - 数据源类型:选择数据库。
 - 数据源:选择步骤二中添加的DLA数据源(dla_hangzhou_tpch)。
 - SQL: 输入以下SQL, 按国家销售额降序排序。

```
select sum(o_totalprice) total_price, n_nationkey, n_name
from orders
join customer on o_custkey = c_custkey
join nation on c_nationkey = n_nationkey
group by n_nationkey, n_name
order by total price desc;
```

- 字段value和content,填入上述SQL返回的total_price列和n_name列。
- 根据业务数据更新需求,选择大屏数据自动更新请求,比如60秒一次。
- 2. 完成上述参数配置后,单击刷新数据。

步骤八:预览和发布大屏

- 1. 单击右上角的预览, 预览大屏发布后的效果。
- 2. 确认大屏无误后,单击右上角的**发布**,打开发布开关,单击分享链接进行复制,在浏览器中粘贴该链接 即可访问大屏。

7.Quick BI 7.1. Quick BI查询加速

本文档主要介绍了配置Quick Bl加速查询的操作步骤。

背景信息

Quick BI是专为云上用户量身打造的新一代智能BI服务平台。提供海量数据实时在线分析服务,支持拖拽式操作,提供了丰富的可视化效果。Quick BI提供了通过DLA加速MaxCompute查询的功能。当您的报表基于 MaxCompute数据源创建时,您可以在Quick BI上配置DLA加速引擎,通过DLA Presto为MaxCompute数据源 进行加速,从而节约报表的查询时间,提升用户体验。

操作步骤

1. 开通DLA Presto引擎CU版。

DLA Presto只在CU版提供了MaxCompute数据源加速查询功能。需要您开通Presto CU版本,具体操作 请参考<mark>创建虚拟集群</mark>。

? 说明

- DLA Presto引擎与需要加速的MaxCompute项目空间要在同一个地域。
- 推荐您使用32核128 GB及以上规格的CU。

2. 配置加速引擎。

在Quick BI上配置加速引擎。具体操作请参见配置Quick引擎。

⑦ 说明 配置加速引擎之前,需要将IP地址 10.152.69.0/24,10.152.163.0/24,139.224.4.0/24 加入到DLA的白名单。否则,Quick BI无法成功连接DLA,起不到加速查询的效果。