阿里云

云原生数据仓库AnalyticDB MySQL版 用户指南

文档版本: 20220712

(一) 阿里云

I

法律声明

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

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

通用约定

格式	说明	样例
⚠ 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。	
△)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。	② 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 Instance_ID
[] 或者 [a b]	表示可选项,至多选择一个。	ipconfig [-all -t]
{} 或者 {a b}	表示必选项,至多选择一个。	switch {active stand}

目录

1.新用户必读	06
2.一站式数据管理	07
2.1. 登录数据库	07
2.2. 数据资产管理	08
2.3. 任务开发与调度	10
3.物化视图	15
3.1. 概述	15
3.2. 创建物化视图	15
3.3. 刷新物化视图	18
3.4. 管理物化视图	20
3.5. 查询物化视图	22
3.6. 物化视图客户案例	23
3.6.1. 案例一	23
3.6.2. 案例二	24
3.6.3. 案例三	24
4.一键构建数据仓库	26
4.1. 什么是一键建仓	26
4.2. 一键创建RDS数据仓库	26
4.3. 一键创建PolarDB MySQL数据仓库	28
4.4. 一键创建日志仓库	30
4.5. 查看或变更同步任务	32
4.6. 常见问题	34
5.标签	35
5.1. 创建标签	35
5.2. 删除标签	36
5.3. 根据标签筛选集群	37

6.诊断与优化	39
6.1. 数据建模诊断	39
6.2. 库表结构优化	40
6.3. SQL Pattern	43
6.4. 通过SQL诊断功能分析执行计划	46
6.4.1. SQL诊断功能介绍	46
6.4.2. 查询监控图和查询列表介绍	48
6.4.3. 使用执行计划分析查询	51
7.每日健康报告	58

1.新用户必读

如果您是首次接触云原生数据仓库AnalyticDB MySQL版的用户,建议您先阅读以下部分。

- **服务亮点和定价**: 产品详情页提供产品介绍、核心能力和产品定价。
- <mark>快速入门</mark>:引导您完成创建实例、设置白名单、创建数据库账号、创建数据库、连接集群、创建表以及同步数据的操作。

完成入门操作后,您可以通过下列资源快速使用其他各项功能:

集群版本	资源
3.0	 连接数据库:介绍在应用开发中连接AnalyticDB MySQL的方法。 数据接入:如果您需要将已有数据源中的数据迁移到AnalyticDB MySQL,阿里云为此提供了完善的数据同步解决方案。 数据可视化:您可以通过AnalyticDB MySQL兼容的BI工具,对AnalyticDB MySQL中的数据进行有效整合,快速准确地提供报表并提出决策依据。 SQL手册:详细介绍AnalyticDB MySQL支持的数据类型、SQL语法以及示例。 系统函数:介绍AnalyticDB MySQL支持的函数和示例。
2.0	 连接数据库:详细介绍在应用开发过程中连接AnalyticDB MySQL的方法。 同步数据:如果您需要将传统数据仓库中的数据迁移到AnalyticDB MySQL,阿里云为此提供了完善的迁移解决方案。 数据可视化:您可以通过AnalyticDB MySQL兼容的BI工具,对AnalyticDB MySQL中的数据进行有效整合,快速准确的提供报表并提出决策依据。 SQL开发规范:介绍常见的SQL优化方法。 向量:介绍向量的使用场景、应用案例以及详细的使用手册。 SQL手册:详细介绍支持的数据类型、SQL语法以及示例。 系统函数:介绍支持的函数以及示例。 最佳实践:介绍使用场景和表设计最佳实践。

2.一站式数据管理

2.1. 登录数据库

云原生数据仓库AnalyticDB MySQL版集群支持通过DMS进行一站式数据管理。本文介绍在使用一站式数据管理功能前,如何通过DMS(新版)登录AnalyticDB MySQL版数据库。

前提条件

已在目标AnalyticDB MySQL版集群中创建数据库账号。创建数据库账号的方法,请参见创建数据库账号。

操作步骤

- 1. 登录云原生数据仓库AnalyticDB MySQL控制台。
- 2. 在页面左上角,选择集群所在地域。
- 3. 在左侧导航栏,单击集群列表。
- 4. 在数仓版(3.0)页签下,单击目标集群ID。
- 5. 在左侧导航栏,选择一站式数据管理 > 数据资产管理进入登录页面。
 - ⑦ 说明 您也可以在左侧导航栏中,选择一**站式数据管理 > 任务开发与调度**进入登录页面。
- 6. 在弹出的对话框中,填写登录信息。

参数	说明
数据库类型	默认为ADB3.0-MySQL,无需选择。
	默认为当前实例所在地域,无需选择。
实例地区	② 说明 若您需要登录其他地域下的AnalyticDB MySQL集群,从下拉列表中选择目标集群的所在地域即可。
	默认为当前集群的集群ID,无需选择。
实例ID	② 说明 若您需要登录其他AnalyticDB MySQL集群,从下拉列表中选择目标集群ID即可。
数据库账号	集群的账号名称。

参数	说明
	账号名对应的密码。
数据库密码	② 说明 您可以选中记住密码,方便之后再次登录当前AnalyticDB MySQL集群时,无需输入数据库账号和密码即可自动登录。

? 说明

- 首次通过DMS登录AnalyticDB MySQL集群时,管控模式默认为自由操作。登录成功后,您还可以通过编辑实例功能来修改管控模式。更多信息,请参见编辑实例和管控模式。
- 配置完登录参数后,您可以单击左下角**测试连接**。如果测试连接失败,请按照报错提示检查录入的集群信息,如账号或密码是否正确。
- 系统会自动尝试往云数据库的白名单中添加DMS的服务器访问地址,若自动添加失败请手动添加。详情信息,请参见设置白名单和DMS白名单列表。

7. 单击登录即可。

后续步骤

通过DMS(新版)控制台成功登录AnalyticDB MySQL集群后,您就可以管理集群的数据资产或进行任务开发与调度操作,详情请参见数据资产管理和任务开发与调度。

2.2. 数据资产管理

本文介绍如何通过DMS的数据资产管理功能对云原生数据仓库AnalyticDB MySQL版集群中的元数据、数据血缘、数据质量进行统一管理。

背景信息

AnalyticDB MySQL版中包含了大量对内对外极具价值或核心的数据(这里不仅指数字,还包含业务逻辑、业务定义等可以勾画整个企业经营的信息),这类数据统称为企业数据资产。数据资产是企业在数字世界的投影,对企业经营或者数字化转型至关重要,因此需要对数据资产进行全面管理。通过DMS的数据资产管理功能不仅能使数据资产可视、可用,同时可以确保数据资产的安全性,进而在企业经营过程中帮助企业挖掘数据价值,产生更多数据资产,形成正向循环。

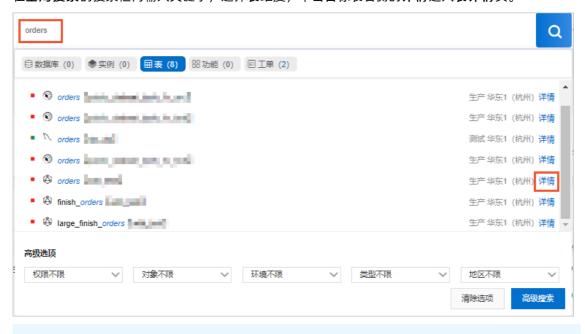
操作步骤

- 1. 登录云原生数据仓库AnalyticDB MySQL控制台。
- 2. 在页面左上角,选择集群所在地域。
- 3. 在左侧导航栏,单击集群列表。
- 4. 在数仓版(3.0)页签下,单击目标集群ID。
- 5. 在左侧导航栏,选择一站式数据管理>数据资产管理。

? 说明

- 若首次通过DMS(新版)登录AnalyticDB MySQL版集群数据库,您需要先完成数据库登录信息录入,登录数据库方法,请参见<mark>登录数据库</mark>。
- 。 若之前已登录过AnalyticDB MySQL版集群数据库且集群管控模式为自由操作或稳定变更,但在录入数据库登录信息时未选中记住密码,进入DMS控制台后,需要先在弹出的对话框中输入目标数据库账号的密码才可进行后续操作。关于管控模式的更多详情,请参见管控模式。
- 6. 在数据资产页签下,您可以进行如下操作:
 - 查看表元数据

在全局搜索的搜索框内输入关键字,选择表维度,单击目标表右侧的详情进入表详情页。



? 说明

- 仅当目标集群的管控模式为**安全协同**时,才支持查看表维度的搜索结果。更多详情,请参见<mark>管控模式</mark>。
- 您可以在下方高级选项区域选择更多维度来筛选搜索结果。

在表**详情**页,您可以查看详细的表信息,例如使用的字符集、所属数据库,当前账号拥有的目标表权限等。您还可以查看目标表的数据血缘、数据质量等详情。



○ 资产大盘

单击左侧导航中的**资产大盘**,即可查看目标账号下所有数据资产的各项统计信息,例如**实例分布占比、数据库分布占比、表分布占比等。**

○ 数据类目

单击左侧导航中的数据类目,即可根据分门别类地对数据资产进行管理。您可以在目标类目下创建多级子类目,在每个类目下均可将添加实例、库或表至对应的类目中。例如您可以将 order 、 fini sh_order 等添加至 订单 类目中。

○ 实例管理

单击左侧导航中的**实例管理**,即可目标账号下所有的实例或者数据库进行批量管理。 例如,您可以选中多个实例进行批量编辑、禁用或删除。

2.3. 任务开发与调度

云原生数据仓库AnalyticDB MySQL版集群支持使用DMS的任务编排功能来编排、调度、管理和监控AnalyticDB MySQL任务。本文介绍如何通过DMS来进行任务开发与调度。

背景信息

● 挑战和诉求: 事件调度

传统的数据库事件调度(如MySQL的事件调度器Event Scheduler)虽然功能强大,但在使用过程中仍发现一些不足:

- 门槛高:用户需要掌握事件定义的特殊SQL语法(CREATE EVENT 、 ALTER EVENT) , 无法通过简单的界面完成配置。
- 强依赖数据库内核:内核需要支持并开启事件调度器。
- 受限单个数据库:只能针对单个数据库进行调度,无法与其他数据库或工具进行联动。
- 难追踪:无法查看调度事件的执行状态、执行历史、运行时长等信息。
- 难运维:无法对事件执行暂停、重跑等操作,事件执行失败也无法恢复。
- 无通知:事件执行状态(成功或失败)无法主动通知到用户(短信、Email等)。

● 解决方案: DMS任务编排

DMS的任务编排功能即可解决上述问题,它并不依赖数据库内核的事件调度能力,而是一个外部独立的任务编排与调度系统。任务编排功能的特点如下:

- 支持十几种数据库引擎(MySQL、Oracle、PostgreSQL、SQL Server等)和多种生态工具(数据迁移、 备份、跨库分析等),可以轻松实现跨库、跨产品功能组合联动。
- 支持可视化的编辑界面,可以通过拖拽和简单的配置即可创建出十分复杂的调度任务。
- 支持多种消息通知机制(钉钉、短信、邮件等)。
- 支持暂停、终止、重跑等多种运维操作。

流程介绍

本文示例中,需要通过DMS的任务编排功能在AnalyticDB MySQL版数据库的 orders 订单表中筛选出已完成且订单金额大于10000元的订单,主要流程如下。

- ? 说明 本文示例中所用的数据库信息如下:
 - 库名: adb test
 - 已在数据库中创建了 orders 、 finish_orders 和 large_finish_orders 三张表,建表语 句如下:

```
create table orders (
 order id bigint not null comment'order ID',
  order status varchar not null comment'order state',
  total price decimal(15,2) not null comment'total price',
  order date date not null comment'orderdate',
  PRIMARY KEY (order id)
  );
 create table finish orders (
  order id bigint not null comment'order ID',
  total price decimal(15,2) not null comment'order state',
  order date date not null comment'total price',
  PRIMARY KEY (order id)
  create table large finish orders (
  order id bigint not null comment'order ID',
  total price decimal(15,2) not null comment'order state',
  order date date not null comment'total price',
  PRIMARY KEY (order id)
```

步骤	说明
步骤一:新增任务流	创建一个任务流。
步骤二: 创建任务节点	在此步骤中,您需要在目标任务流中创建如下2个任务节点: • 订单清洗: 从 orders 订单表中清洗出已完成(即 order_status=F)的订单,并将结果写入 finish_orders 表。 • 大额订单生成: 从 finish_orders 表中搜索订单金额大于10000(即 total_price>10000)元的订单,并将结果写入 large_finish_orders)表。
步骤三: 配置任务流调度	开启目标任务并使其周期性地运行。

步骤一:新增任务流

- 1. 登录云原生数据仓库AnalyticDB MySQL控制台。
- 2. 在页面左上角,选择集群所在地域。
- 3. 在左侧导航栏,单击集群列表。
- 4. 在数仓版(3.0)页签下,单击目标集群ID。
- 5. 在左侧导航栏,选择一站式数据管理 > 任务开发与调度。

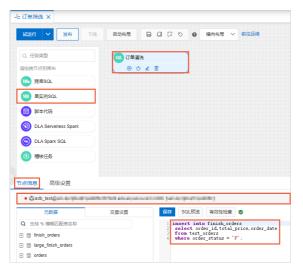
? 说明

- 若首次通过DMS(新版)登录AnalyticDB MySQL版集群数据库,您需要先完成数据库登录信息录入,登录数据库方法,请参见<mark>登录数据库</mark>。
- o 若之前已登录过AnalyticDB MySQL版集群数据库且集群管控模式为自由操作或稳定变更,但在录入数据库登录信息时未选中记住密码,进入DMS控制台后,需要先在弹出的对话框中输入目标数据库账号的密码才可进行后续操作。关于管控模式的更多详情,请参见管控模式。
- 6. 在顶部菜单栏中,选择集成与开发(DTS)>数据开发>任务编排。
- 7. 新增任务流。本文将任务流命名为订单筛选。
 - i. 单击新增任务流。
 - ② 说明 如果您使用的是旧版数据管理DMS,单击页面左侧的 如按钮。在开发空间页面中,单击新建任务流。
 - ii. 在新建任务流对话框中,输入任务流名称和描述,单击确认。

步骤二: 创建任务节点

- 1. 在订单筛选的任务编排页面,创建并配置以下2个任务节点:
 - 订单清洗
 - a. 在页面左侧选择**单实例SQL**,并将其拖拽到右侧空白画布中。
 - b. 选中新加入的任务节点, 单击∠图标将其重新命名为 订单清洗 。
 - c. 在页面下方的**节点信息**页签中,从数据库下拉列表中选择目标数据库。
 - d. 在目标数据库下方的对话框中,输入任务语句,单击**保存**。语句如下:

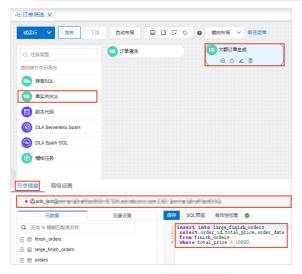
```
insert into finish_orders
  select order_id,total_price,order_date
  from orders
  where order_status = 'F';
```



○ 大额订单生成

- a. 在页面左侧选择单实例SQL,并将其拖拽到右侧空白画布中。
- b. 选中新加入的任务节点,单击 _ 图标将其重新命名为 大额订单生成 。
- c. 在页面下方的**节点信息**页签中,从数据库下拉列表中选择目标数据库。
- d. 在目标数据库下方的对话框中,输入任务语句,单击**保存**。语句如下:

```
insert into large_finish_orders
select order_id,total_price,order_date
from finish_orders
where total_price > 10000;
```



2. 在任务流画布中,将鼠标放在 订单清洗 节点上,单击节点右侧的小圆圈并将拉出连接线,连接至 大额订单生成节点 节点,即可在两个任务节点间连接形成依赖任务流。



步骤三:配置任务流调度

1. 在任务流画布下方的调度配置区域,打开开启调度开关,并配置相关参数。



- ② 说明 本文示例中将调度任务设置为在2021-06-01至2021-06-30期间的每日01:00周期性运行,您也可以根据业务需要自定义调度任务。更多关于调度配置的参数说明,请参见配置调度。
- 2. 发布任务流。
 - i. 单击画布左上角的**发布**。
 - ii. 在**发布**对话框中输入**备注**信息,并单击**确认**,发布任务流。
 - ② 说明 如果您不再需要使用该任务流,可以下线任务流:
 - i. 单击画布左上角的下线。
 - ii. 在确认对话框中单击**确认**,下线任务流。
- 3. 确认发布状态。
 - i. 单击画布上方的**前往运维**。
 - ii. 在界面右侧,查看任务流是否发布的状态。
 - **已发布**:该任务流处于已发布状态。 ■ **未发布**:该任务流处于未发布状态。

3.物化视图

3.1. 概述

物化视图是数仓领域的核心特性之一。不同于逻辑视图(view),物化视图(materialized view)会持久化视图的查询结果。

② 说明 请先升级内核到最新版本(最低支持的版本为3.1.3.4)再使用物化视图。如果版本较低请联系官方支持(钉钉号:x5v_rm8wqzuqf)进行升级。关于如何查看实例内核版本,请参见如何查看实例版本信息。

您可以对物化视图执行以下操作:

- 创建物化视图
- 刷新物化视图
- 管理物化视图
- 查询物化视图

物化视图可用于加速分析,并能简化ETL,适用于多种场景,例如报表类业务,大屏展示需求,来自BI工具的 查询等等。

关于使用物化视图的客户案例, 请参见案例一。

3.2. 创建物化视图

本文介绍如何创建物化视图。

权限要求

- 需要有数据库或表级别的CREATE权限。
- 需要有数据库或表级别的INSERT权限。
- 需要有物化视图所涉及的所有表的相关列(或整个表)的SELECT权限。
- 如果要启用自动刷新功能,需要创建者具备通过服务器本地(即 127.0.0.1)或者任意IP(即 '%') 刷新视图的权限。

语法

```
CREATE [OR REPLACE] MATERIALIZED VIEW <mv_name>
[MV DEFINITION]
[REFRESH COMPLETE [ON <DEMAND|OVERWRITE>] [START WITH date] [NEXT date]]
AS
<QUERY BODY>;
```

关键字说明

关键字	说明
-----	----

关键字		说明
MV DEFIN	ITION	定义物化视图中表相关的特性。物化视图使用一张普通的表结构来存储数据,您可以使用该关键字指定物化视图的主键,分区键,索引等。 ② 说明 与创建表的语法一致,创建物化视图时也不支持定义查询中没有输出的列。
	COMPLE	定义物化视图的刷新模式,当前仅支持 COMPLETE 全量刷新模式。
REFRES H STA WITH	ON	定义全量刷新模式的触发方式,支持如下取值: ON DEMAND :表示按需触发全量刷新,即您可以在需要刷新物化视图时手动触发刷新,或根据刷新需求配置自动触发刷新。 ON OVERWRITE :表示物化视图引用的基表在执行 INSERT OVERWRITE 语句导致数据被覆盖后,会触发全量刷新,适合通过batchload批量导入数据的场景。
		② 说明 ● 仅当刷新模式 REFRESH 参数设置为 COMPLETE 时,该参数生效。 ● 未指定该参数时,默认全量刷新触发模式为 ON DEMAND 。
	START WITH 和 NEXT	定义自动进行全量刷新物化视图时的首次刷新时间(即 NEXT)。 ② 说明 ■ START WITH 参数选填,若不填,默认首次刷新时间为当前时间点。 ■ 若需要设置自动刷新, NEXT 参数必填。 ■ 支持使用时间函数,但只支持精确到秒,毫秒部分会被截断。 更多关于刷新物化视图的详情,请参见刷新物化视图。
QUERY BOI	DY	定义物化视图的查询主体,其中: • 需要为查询中输出的表达式列定义别名,建议使用具有实际意义的别名。例如 (sum(price) AS total_price) ,表示设置表达式列 sum(price) 的别名为 total_price。 • 物化视图查询所涉及的基表不可被删除,且基表中的列不可被删除或修改。 • 可以使用 WITH 语法进行查询,查询主体可以基于逻辑视图或者物化视图。

使用限制

- 不支持对物化视图执行 INSERT 、 DELETE 或 UPDATE 操作。
- 不支持删除或重命名物化视图中引用的基表或基表中的列。如需修改基表需要先删除物化视图。
- 默认情况下,一个AnalyticDB MySQL版集群支持创建的物化视图上限如下:
 - 集群版本小于3.1.4.7: 最多支持创建8个物化视图。
 - 集群版本等于或大于3.1.4.7: 最多支持创建64个物化视图。
 - ② 说明 若您需要创建更多物化视图,请提交工单联系技术支持。

示例

```
# 每五分钟刷新一次
 CREATE MATERIALIZED VIEW myview
REFRESH NEXT now() + interval 5 minute
 SELECT count(*) as cnt FROM base;
 # 每周一晚上2点刷新
 CREATE MATERIALIZED VIEW myview
 START WITH DATE FORMAT(now() + interval 7 - weekday(now()) day, '%Y-%m-%d 02:00:00')
 NEXT DATE FORMAT(now() + interval 7 - weekday(now()) day, '%Y-%m-%d 02:00:00')
 SELECT count(*) as cnt FROM base;
 # 每天凌晨2点刷新
 CREATE MATERIALIZED VIEW myview
 START WITH DATE FORMAT(now() + interval 1 day, '%Y-%m-%d 02:00:00')
 NEXT DATE_FORMAT(now() + interval 1 day, '%Y-%m-%d 02:00:00')
 SELECT count(*) as cnt FROM base;
 # 每个月第一天凌晨2点刷新
 CREATE MATERIALIZED VIEW myview
 REFRESH NEXT DATE FORMAT(last day(now()) + interval 1 day, '%Y-%m-%d 02:00:00')
 SELECT count(*) as cnt FROM base;
 # 只刷一次
 CREATE MATERIALIZED VIEW myview
 REFRESH START WITH now() + interval 1 day
 SELECT count(*) as cnt FROM base;
 # 不自动刷新,完全依靠手动刷新
 CREATE MATERIALIZED VIEW myview (
 PRIMARY KEY (id)
 ) DISTRIBUTED BY HASH (id)
 SELECT id, name FROM base;
 # 指定列建立索引,默认全部列建立索引
CREATE MATERIALIZED VIEW myview (
  INDEX (name),
  PRIMARY KEY (id)
) DISTRIBUTED BY HASH (id)
```

```
AS
SELECT id, name, age FROM base;
# 指定分区键和注释
CREATE MATERIALIZED VIEW c (
    name varchar(10),
    value double,
    KEY INDEX_ID(id) COMMENT 'id',
    CLUSTERED KEY INDEX(name, value),
    PRIMARY KEY(id)
)
DISTRIBUTED BY hash(id)
PARTITION BY value(date_format(dat, "%Y%m%d")) LIFECYCLE 30
COMMENT 'MATERIALIZED VIEW c'
AS
SELECT * FROM base;
```

使用建议

- 创建物化视图时支持使用普通表定义中所有的参数,如分区键、二级分区键和冷热数据存储策略等。创建物化视图时建议定义分区键和主键,以提高后续查询性能。关于创建普通表支持的参数说明,请参见CREATE TABLE。
- 若集群规格较小,尽量避免在同一时刻刷新集群内的所有物化视图,以免影响集群稳定性。
- 物化视图和普通表一样,默认全列建立索引。为减少存储空间和写入I/O,如果不需要全列建立索引,您可以指定INDEX关键字对特定列建立索引。在物化视图中建索引的方式与普通表的语法一致,建索引的语句,请参见CREATE TABLE。

3.3. 刷新物化视图

您可以在创建物化视图时指定刷新模式或者手动刷新物化视图。

权限要求

- 需要物化视图INSERT权限。
- 需要有物化视图所涉及的所有基表相关列(或整个表)的SELECT权限。
- 如果要启用自动刷新功能,需要创建者在127.0.0.1上有刷新视图的权限。因为系统会代理创建者身份,从服务器本地(127.0.0.1)登录来执行自动刷新。这样当创建者没有权限之后,系统也无法刷新,避免越权问题。如果用户权限是在所有IP上('%'),那么就不会产生这个问题,因为所有IP也包括了127.0.0.1。

创建时指定刷新模式

物化视图刷新分为全量刷新(COMPLETE)和增量刷新(FAST)两种刷新模式。如果未指定刷新模式, 默认为全量刷新。

每种刷新模式适用于不同的触发方式。

```
CREATE MATERIALIZED VIEW mv_name
[REFRESH [FAST|COMPLETE] [ON <DEMAND|OVERWRITE>] [START WITH date] [NEXT date]]
```

⑦ 说明 目前仅支持全量刷新 (COMPLETE)。

手动刷新

手动触发一次刷新

REFRESH MATERIALIZED VIEW <mv name>

全量刷新模式

创建物化视图时默认为全量刷新模式。全量刷新会计算刷新时刻的查询结果,并且用覆盖的方式替换原来的结果。

全量刷新触发机制

ON DEMAND:按需刷新。按需刷新要么用户手动触发刷新,要么自动刷新。全量刷新默认采用按需刷新(ON DEMAND)。

```
# 手动刷新
CREATE MATERIALIZED VIEW my_mv
REFRESH ON DEMAND
AS
SELECT * FROM base
# 自动刷新 (5分钟后第一次刷新,此后每分钟1次)
CREATE MATERIALIZED VIEW my_mv
REFRESH ON DEMAND START WITH now() + interval 5 minute NEXT now() + interval 1 minute
AS
SELECT * FROM base
```

ON OVERWRITE: 当基表被INSERT OVERWRITE覆写后,会触发对应物化视图的刷新。适合T+1场景,在数据清洗覆盖后自动更新物化视图。

```
# 创建
CREATE MATERIALIZED VIEW my_mv
REFRESH ON OVERWRITE
AS
SELECT * FROM base
# 触发 my_mv 刷新
INSERT OVERWRITE base SELECT * FROM t0;
```

自动刷新

自动刷新支持使用时间函数(START WITH ... NEXT)来设置刷新规律。 START WITH 是可选的,默认从当前时刻开始。

```
# 每周一晚上2点刷新
CREATE MATERIALIZED VIEW myview
START WITH DATE FORMAT(now() + interval 7 - weekday(now()) day, '%Y-%m-%d 02:00:00')
NEXT DATE FORMAT(now() + interval 7 - weekday(now()) day, '%Y-%m-%d 02:00:00')
SELECT count(*) as cnt FROM base
# 每天晚上夜里2点刷新
CREATE MATERIALIZED VIEW myview
REFRESH
START WITH DATE FORMAT(now() + interval 1 day, '%Y-%m-%d 02:00:00')
NEXT DATE FORMAT(now() + interval 1 day, '%Y-%m-%d 02:00:00')
SELECT count(*) as cnt FROM base
# 每个月第一天凌晨2点刷新
CREATE MATERIALIZED VIEW myview
REFRESH NEXT DATE FORMAT(last day(now()) + interval 1 day, '%Y-%m-%d 02:00:00')
SELECT count(*) as cnt FROM base
# 只刷一次
CREATE MATERIALIZED VIEW myview
REFRESH START WITH now() + interval 1 day
SELECT count(*) as cnt FROM base
```

如果某次刷新所用的时间超过自动刷新指定的间隔,会自动跳过超过的时间点,等到下次最近的时间点再刷新。

假设自动刷新设置是在00:05:00,00:10:00,00:15:00这3个时间点,间隔5分钟。如果00:05:00时候开始了某次刷新,一直到00:12:00刷新才结束,那么设置的00:10:00刷新会被跳过,直到00:15:00再开始刷新。

3.4. 管理物化视图

本文介绍如何管理物化视图。

查看物化视图定义

权限要求

需要有物化视图的SELECT权限。

语法

```
SHOW CREATE MATERIALIZED VIEW <mv_name>
```

功能说明

返回物化视图创建DDL语句详细信息。

查找物化视图

权限要求

需要有物化视图的SELECT权限。

语法

SHOW MATERIALIZED VIEWS [LIKE 'pattern']

功能说明

返回当前库里面所有具有SELECT权限的物化视图列表。如果有指定pattern,则返回pattern匹配上的所有物化视图。

删除物化视图

权限要求

需要有物化视图的DROP权限。

语法

DROP MATERIALIZED VIEW <mv name>

功能说明

删除物化视图。物化视图不能使用DROP TABLE来删除。

物化视图信息表

information_schema.mv_info

所有物化视图的元信息。

常用字段的解释如下表所示:

字段	描述
MV_SCHEMA	物化视图的所在库名。
MV_NAME	物化视图的名称。
MV_DEFINITION	物化视图表属性的DDL。
MV_QUERY	物化视图查询部分的SQL。
FIRST_REFRESH_TIME	如果已配置自动刷新,则表示首次刷新的时间。
NEXT_REFRESH_TIME_FUNC	如果已配置自动刷新,则表示下次刷新的时间推导函数。
OWNER	物化视图的创建者。

如果只需查看单张物化视图定义,建议使用SHOW CREATE MATERIALIZED VIEW命令。

使用示例如下:

统计整个实例有多少张物化视图

SELECT count(*) FROM information schema.mv info

information schema.mv auto refresh jobs

自动刷新任务记录。只有配置了自动刷新功能的物化视图才会出现在这张表里,最多保留1024条记录。

常用字段的解释如下表所示:

字段	描述
MV_SCHEMA	物化视图的所在库名。
MV_NAME	物化视图的名称。
SCHEDULED_START_TIME	计划启动刷新的时间。
START_TIME	实际启动刷新的时间。
END_TIME	刷新结束的时间。
STATE	刷新状态。 PENDING:等待触发刷新。 RUNNING:刷新中,如果很多物化视图同时开始刷新,状态虽然是RUNNING,但是可能在等计算资源。 FINISHED:刷新完成。 FAILED:刷新失败。刷新失败会自动重试,直到重试次数耗尽,那么会跳过这次刷新任务,等待下次刷新。 TIMEOUT:刷新任务进入未知状态,迟迟没有成功。

使用示例如下:

```
# 查看 mymv 的自动刷新记录
SELECT * FROM information_schema.mv_auto_refresh_jobs where mv_name = 'mymv'
```

3.5. 查询物化视图

本文介绍如何在任意查询语法中(例如 SELECT , INSERT INTO SELECT WITH 等)使用物化视图进行查询。

权限要求

仅需具备物化视图的SELECT权限,无需具备物化视图所引用基表的SELECT权限。

```
② 说明 您可以使用 GRANT 命令为用户添加物化视图的SELECT权限,语句如下:

GRANT SELECT ON <mv_name> TO '<user>'@'%';
```

使用方法

查询物化视图与查询其他表或视图的语句一样,您可以在任何查询语句中使用物化视图。语句示例如下:

```
SELECT * FROM adbview
WHERE device = 'PC'
AND city = 'Beijing';
```

② 说明 在查询中使用物化视图时,可能存在基表数据已更新,但物化视图还未更新的情况。此时的查询结果可能并不准确。因此建议您在基表数据更新后,手动刷新一次物化视图再进行查询。手动刷新视图的方法,请参见刷新物化视图。

3.6. 物化视图客户案例

3.6.1. 案例一

本文介绍使用物化视图降低客户查询延迟时间的案例。

业务背景

生意参谋是阿里巴巴旗下为千万商家提供的一项重要产品服务,帮助商家及时分析店铺运营情况。尤其是在 大促期间,面对突发的流量和海量的数据,数据分析尤为重要。

生意参谋基于AnalyticDB构建了一套大促期间小时粒度的全店诊断能力。多维度(展示门店+商品+渠道)下支付金额、支付件数、支付买家数、访客数、加购人数五大指标的整体诊断情况;展示行业实时排名,商家所处的同行同层排名;展示本店今日排名趋势,本店全天的排名变化趋势图等等。

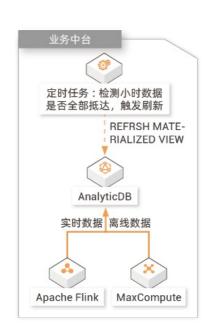
业务效果

如果前端应用每次对小时级别的数据进行查询,AnalyticDB每次都对小时级别的数据做实时计算。这些计算会包含JOIN,GROUP BY等相对复杂运算,时间最少也要1秒。向商家展示当天24小时的数据就需要24秒,这个时间延迟比较久。

利用物化视图,可以大幅降低延迟时间。将每小时展示信息结果存储到物化视图中,每次查询只需要查询物化视图即可,平均每次查询时间降低至100毫秒。







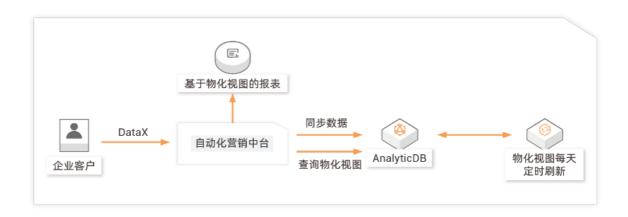
3.6.2. 案例二

业务背景

某客户为鞋服类企业客户提供自动化营销服务。通常需要结合会员标签,库存信息等多种维度来运算报表,运算量较大且耗时较久。这样不利于营销人员及时查询,影响报表快速展示。同时,并发的大查询也会影响后台数据库稳定性。

业务效果

客户使用物化视图定时自动刷新功能,每天晚上定时刷新视图。视图中已有运算好的数据,利于快速查询, 生成报表。同时利用定时刷新功能,客户可以节省维护额外的任务调度系统开销。



3.6.3. 案例三

业务背景

菜鸟物流是阿里巴巴旗下的重要业务,管理着成千上万的包裹。当包裹进入仓库后,很重要的一个环节是实时抽检,然后相关人员会定期查看近7天的抽检报表。报表会展示商品,抽检率,合格率,包裹数等各种指标。计算报表需要消耗大量计算资源,双十一大促时,不仅报表生成的速度会很慢,当多人查看时,不时发送大查询到数据库也容易导致系统不稳定。

业务效果

使用物化视图将近7天的抽检报表数据保存下来,每次查看报表时只需要查询物化视图即可。物化视图按小时自动刷新,带上时间过滤条件即可,SQL如下所示:

```
CREATE MATERIALIZED VIEW xxx_report

REFRESH NEXT now + interval 1 hour

AS

SELECT ...

WHERE create_time >= select current_date() - interval 7 day
```



4.一键构建数据仓库

4.1. 什么是一键建仓

云原生数据仓库AnalyticDB MySQL版(简称ADB,原分析型数据库MySQL版)支持一键建仓功能,可以将RDS、PolarDB MySQL、或者日志服务中某个日志库中的数据快速同步到ADB集群中。

应用场景

- 一键建仓主要解决云上企业构建数据仓库时配置数据同步链路繁杂的问题,同时提供了高性价比的ADB存储计算资源方案。一键建仓适用于以下场景:
- 全库同步,通过一键建仓一键生成同步链路,将一个或多个RDS实例、PolarDB MySQL集群中的数据同步到同一个ADB集群中。
 - ② 说明 一键建仓支持跨区域同步数据,例如您可以将杭州区域的RDS、PolarDB MySQL数据同步到上海ADB集群中。
- 多源汇集,通过一键建仓,可以将RDS、PolarDB MySQL、日志数据便捷地汇集到同一个ADB集群。

使用方式

- 一键创建RDS数据仓库
- 一键创建PolarDB MySQL数据仓库
- 一键创建日志仓库

4.2. 一键创建RDS数据仓库

您可以在AnalyticDB MySQL版控制台上通过一键建仓将一个或多个RDS实例中的数据同步到一个AnalyticDB MySQL集群中,快速创建RDS数据仓库。

前提条件

RDS数据库中存在表数据。

计费

- 一键创建RDS数据仓库时,系统自动通过DTS同步数据,DTS计费如下:
- 若同步链路数量小于或等于10条, DTS免费, 限期2年。

? 说明

- 。 一个RDS实例的数据同步到AnalyticDB MySQL集群即为一条链路。
- o DTS两年免费的链路规格为micro和small,一键建仓创建的同步任务的链路规格为small。
- 若已使用10条同步链路,即使释放链路重新购买,也属于新购链路,按照11条以上同步链路 价格标准收费。
- 若同步链路数量大于10条,6个月内DTS免费;6个月之后需要续费才能继续使用,按照同步链路价格的6 折收费,DTS费用请参见产品定价。

注意事项

- 数据同步过程中,若在源实例RDS中新增库,新增库数据不会自动同步,需要手动配置同步链路,请参见手动配置。
- 不要在数据同步过程中增加无主键表,将导致同步链路中断。
- 同步链路中断7天后不再保证数据正确性,请及时处理链路中断问题。
- 一键建仓链路不支持修改同步的库表对象,例如不支持手动移除或添加同步表等。

支持同步的SQL操作

操作类型	SQL操作语句
DML	INSERT \ UPDATE \ DELETE
DDL	CREATE TABLE、DROP TABLE、RENAME TABLE、TRUNCATE TABLE、ADD COLUMN、DROP COLUMN

○ 警告 如果在数据同步过程中变更了源表的字段类型,同步任务将报错并中断。您可以<mark>提交工单</mark>处理或参照以下方法手动修复。

- 1. 在同步至目标库AnalyticDB MySQL时,源表(例如customer)因字段类型变更而导致同步任务失败。
- 2. 在AnalyticDB MySQL 3.0中创建一个新表(customer_new),表结构与customer表保持一致。
- 3. 通过INSERT INTO SELECT命令,将customer表的数据复制并插入到新创建的customer_new表中,确保两张表的数据保持一致。
- 4. 重命名或删除同步失败的表customer, 然后将customer new表的名称修改为customer。
- 5. 在DTS控制台,重新启动数据同步任务。

操作步骤

- 1. 登录AnalyticDB for MySQL控制台。
- 2. 在左侧导航栏,单击集群列表,然后单击一键建仓 > 数据库一键建仓。
- 3. 在一键建仓页面,按照页面提示进行参数配置。

类别	参数	说明
数据源配置	源实例地域	从源实例地域下拉列表,选择RDS实例所在地域,然后从左侧实例列表中选择一个或多个实例,单击 为将其添加到右侧的已选择实例中。 ② 说明 一键建仓支持跨地域同步数据,例如您可以将杭州地域的RDS数据同步到上海AnalyticDB MySQL集群中,跨地域详情以具体页面为准。
	集群地域	从 集群地域 下拉列表,选择AnalyticDB MySQL集群所在地域。

类别	参数	说明
云原生数据仓库	集群名称	从集群名称下拉列表,选择一个AnalyticDB MySQL集群。 AnalyticDB MySQL将根据您选择的RDS实例中的数据存储量,自动进行存储空间计算,只在集群名称中显示符合条件的AnalyticDB MySQL集群。 您也可以单击立即购买新建AnalyticDB MySQL集群,将RDS数据同步到新集群中,请参见创建集群。

4. 完成上述参数配置后,单击创建,AnalyticDB MySQL系统自动创建DTS同步链路并进行数据同步。

相关操作

- 杳看或变更同步任务
- 常见问题

4.3. 一键创建PolarDB MySQL数据仓库

您可以在AnalyticDB MySQL版控制台上通过一键建仓将一个或多个PolarDB MySQL实例中的数据同步到一个 AnalyticDB MySQL集群中,快速创建PolarDB MySQL数据仓库。

前提条件

PolarDB MySQL数据库中存在表数据。

计费

- 一键创建PolarDB MvSQL数据仓库时,系统自动通过DTS同步数据,DTS计费如下:
- 若同步链路数量小于或等于10条, DTS免费, 限期2年。

? 说明

- 一个PolarDB MySQL实例的数据同步到AnalyticDB MySQL集群即为一条链路。
- o DTS两年免费的链路规格为micro和small,一键建仓创建的同步任务的链路规格为small。
- 若已使用10条同步链路,即使释放链路重新购买,也属于新购链路,按照11条以上同步链路 价格标准收费。
- 若同步链路数量大于10条,6个月内DTS免费;6个月之后需要续费才能继续使用,按照同步链路价格的6 折收费,DTS费用请参见产品定价。

注意事项

- 数据同步过程中,若在源实例PolarDB MySQL中新增库,新增库数据不会自动同步,手动配置同步链路, 请参见手动配置。
- 不要在数据同步过程中增加无主键表,将导致同步链路中断。
- 同步链路中断7天后不再保证数据正确性,请及时处理链路中断问题。
- 一键建仓链路不支持修改同步的库表对象,例如不支持手动移除或添加同步表等。

支持同步的SQL操作

操作类型	SQL操作语句
DML	INSERT , UPDATE , DELETE
DDL	CREATE TABLE、DROP TABLE、RENAME TABLE、TRUNCATE TABLE、ADD COLUMN、DROP COLUMN

○ 警告 如果在数据同步过程中变更了源表的字段类型,同步任务将报错并中断。您可以<mark>提交工单</mark>处理或参照以下方法手动修复。

- 1. 在同步至目标库AnalyticDB MySQL时,源表(例如customer)因字段类型变更而导致同步任务失败。
- 2. 在AnalyticDB MySQL 3.0中创建一个新表(customer_new),表结构与customer表保持一致。
- 3. 通过INSERT INTO SELECT命令,将customer表的数据复制并插入到新创建的customer_new表中,确保两张表的数据保持一致。
- 4. 重命名或删除同步失败的表customer, 然后将customer new表的名称修改为customer。
- 5. 在DTS控制台,重新启动数据同步任务。

操作步骤

- 1. 登录云原生数据仓库AnalyticDB MySQL控制台。
- 2. 在左侧导航栏,单击集群列表,然后单击一键建仓 > 数据库一键建仓。
- 3. 在数据库一键建仓页面,按照页面提示进行参数配置。

类别	参数	说明
数据源配置	源实例地域	从 源实例地域 下拉列表,选择PolarDB MySQL实例所在地域,然后从左侧实例列表中选择一个或多个实例,单击 > 将 其添加到右侧的已选择实例中。
		? 说明 一键建仓支持跨地域同步数据,例如您可以将杭州地域的PolarDB MySQL数据同步到上海AnalyticDB MySQL集群中,跨地域详情以具体页面为准。
云原生数据仓库	集群地域	从 集群地域 下拉列表,选择AnalyticDB MySQL集群所在地域。
	集群名称	从集群名称下拉列表,选择一个AnalyticDB MySQL集群。 AnalyticDB MySQL将根据您选择的PolarDB MySQL实例中的数据存储量,自动进行存储空间计算,只在集群名称中显示符合条件的AnalyticDB MySQL集群。 您也可以单击立即购买新建AnalyticDB MySQL集群,将PolarDB MySQL数据同步到新集群中,请参见创建集群。

4. 完成上述参数配置后,单击**创建**,AnalyticDB MySQL自动创建DTS同步链路并进行数据同步。

相关操作

- 支持的数据源
- 查看或变更同步任务
- 常见问题

4.4. 一键创建日志仓库

您可以在ADB控制台上通过一键建仓将一个日志服务项目(Project)中的日志投递到一个ADB集群中,快速创建日志仓库。

前提条件

日志库 (Logstore) 中存在日志。

操作步骤

- 1. 登录AnalyticDB for MySQL控制台。
- 2. 在左侧导航栏,单击集群列表,然后单击一键建仓 > 日志一键建仓。
- 3. 在日志一键建仓页面,按照页面提示进行参数配置。



参数	说明
	从 Project名称 下拉列表,选择日志服务。
Project名称	⑦ 说明 日志一键建仓不支持跨区域同步日 志。
Logstore名称	从 Logstore名称 下拉列表,选择日志库。
投递名称	为投递任务取一个名字,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。
集群名称	从下拉列表中选择目标ADB集群。
	从下拉列表中选择目标ADB数据库名。
表名	从下拉列表中选择目标ADB表名。

参数	说明
账号名称	目标ADB中创建的 <mark>数据库账号</mark> 。
账号密码	填写账号名称对应的密码。
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ADB字段。 单击增加一条新增字段。
投递开始时间	设置投递开始时间。 当数据写入日志服务后,日志服务实时将数据投递到 ADB表中。
是否过滤脏数据	设置是否过滤脏数据。

4. 完成上述参数配置后,单击**确定**,日志服务将在您设置的投递时间将日志实时投递到ADB表中。 将日志投递到ADB后,您就可以通过SELECT对日志数据进行自由灵活地分析。 更多日志服务信息,请参见日志服务。

相关操作

- 支持的数据源
- 常见问题

4.5. 查看或变更同步任务

使用一键建仓功能成功创建数据库仓库或日志仓库后,系统会自动通过DTS将建仓数据同步至AnalyticDB MySQL版集群,同时控制台出现一键建仓的同步任务。用户可以查看一键建仓同步任务的状态和进度,也可以修改、暂停或删除同步任务。本文介绍如何通过AnalyticDB MySQL版控制台或DTS控制台查看、变更同步任务。

前提条件

已存在一键建仓的同步任务。

通过AnalyticDB MySQL控制台查看或变更同步任务

- 1. 登录云原生数据仓库AnalyticDB MySQL控制台。
- 2. 在页面左上角,选择集群所在地域。
- 3. 在左侧导航栏,单击集群列表。
- 4. 在数仓版 (3.0) 页签下, 单击目标集群ID。
- 5. 在左侧导航栏,单击**同步概览**即可查看该集群下的同步任务详情。您也可以修改同步的数据源,暂停、 启动、删除同步任务。



② 说明 您可以单击同步任务名称栏中的任务ID跳转至DTS控制台,来查看目标同步任务的详细信息。关于DTS控制台上的同步任务信息介绍,请参见查看同步链路状态和性能。

通过DTS控制台查看同步任务

你可以通过以下任意方式进入DTS控制台查看同步任务。

● 方式一

- i. 登录云原生数据仓库AnalyticDB MySQL控制台。
- ii. 在页面左上角,选择集群所在地域。
- iii. 在左侧导航栏,单击集群列表。
- iv. 在数仓版 (3.0) 页签下, 单击目标集群ID。
- v. 在左侧导航栏,单击**同步概览**。
- vi. 单击同步任务名称,可自动跳转至DTS控制台的作业配置页。

• 方式二

- i. 登录数据传输控制台。
- ii. 在左侧导航栏,单击数据同步。
- iii. 在同步作业列表页面顶部,选择数据同步实例所属地域。
 - ⑦ 说明 数据同步实例的所属地域与目标AnalyticDB MySQL版集群所属地域一致。
- iv. 在**同步作业列表**页的搜索框中,输入**同步作业名称为ADB一键建仓同步作业**,即可查看当前地域下所有AnalyticDB MySQL版一键建仓的同步任务详情。



● 方式三

- i. 登录云原生数据仓库AnalyticDB MySQL控制台。
- ii. 在页面左上角,选择集群所在地域。
- iii. 在左侧导航栏,单击集群列表。
- iv. 在页面右上角,单击数据同步即可自动跳转至DTS控制台的同步作业列表页。



- v. 在**同步作业列表**页面顶部,选择数据同步实例所属地域。
 - ② 说明 数据同步实例的所属地域与目标AnalyticDB MySQL版集群所属地域一致。
- vi. 在同步作业列表页的搜索框中,输入同步作业名称为ADB一键建仓同步作业,即可查看当前地域

下所有AnalyticDB MySQL版一键建仓的同步任务详情。



4.6. 常见问题

● 数据同步过程中如果源实例中的数据库名、表名或者列名中含有划线(-)会报错吗?

由于ADB中不允许数据库名、表名或者列名中含有划线(-),为保证数据同步成功,系统会将中划线(-)映射为下划线(_)。

如果在数据同步过程中遇到其他因数据库名、表名或者列名造成的同步失败(例如表名中含有空格、中文等),可以向ADB提工单解决该问题。

- ② 说明 更多ADB使用限制请参见约束和限制。
- 数据同步过程中,是否支持修改源表中的字段类型?

数据同步过程中,可以修改源表中的字段类型,目前仅支持整型数据类型之间、浮点数据类型之间的列类型更改,并且只能将取值范围小的数据类型更改为取值范围大的数据类型,或者将单精度数据类型更改为双精度数据类型。

- 整型数据类型:支持Tinyint、Smallint、Int、Bigint间,小类型到大类型的更改,例如支持将Tinyint更改为Bigint,不支持将Bigint更改为Tinyint。
- 浮点数据类型:支持将Float更改为Double类型,不支持将Double更改为Float类型。
- 如何排查ADB集群中的数据延迟问题?
 - DTS同步链路规格默认为medium模式,源库的数据写入量过大时,若想达到规格同步性能上线,需要升级实例配置。
 - 无主键表的主键选择可能会导致热点行更新,而热点行更新速度很慢,可以向ADB提工单解决该问题。
 - 当ADB集群的写入性能已达到瓶颈,需要升级ADB规格。
- 若源库中存在无主键表,同步过程会报错吗?

针对源库中的无主键表,同步时会把表中第一列作为主键同步到ADB中。

- ② 说明 ADB要求只有主键表才可以进行INSERT和UPDATE操作。
- 若源库中存在ADB不支持的数据类型,同步过程会报错吗?

若源库中存在ADB不支持的数据类型(例如地理位置数据等类型),ADB在进行结构初始化时会丢弃掉不支持数据类型的列,ADB常见数据类型请参见基础数据类型。

5.标签

5.1. 创建标签

如果您有大量集群,可以通过给集群绑定标签,对集群进行分类管理。每个标签由一对键值组成,您可以通过键值,对集群进行二级分类。

限制说明

- 每个集群最多可以绑定20个标签,且标签键必须唯一。相同的标签键会被覆盖。
- 每次绑定/解绑标签的数量分别不能超过 20个。
- 不同地域的标签信息是独立的。
- 任一标签在解绑后,如果没有绑定任何集群,则该标签会被删除。

操作步骤

- 1. 登录 AnalyticDB 控制台。
- 2. 在页面左上角,选择集群所在地域。
- 3. 在左侧导航栏,单击集群列表。
- 4. 在数仓版(3.0)中,单击目标集群ID。
- 5. 在集群信息页面的集群属性区域,单击编辑标签。
- 6. 在编辑标签页面,单击新建标签,输入标签的标签键和标签值,单击确定。
 - ⑦ 说明 如果您已经新建了标签,可以单击已有标签,选择历史标签。





7. 填写完所有要绑定的标签后,单击确定完成绑定。

5.2. 删除标签

如果集群调整或者不再需要标签,您可以删除该集群的标签。

限制说明

- 每次解绑的标签数量不能超过20个。
- 任一标签在解绑后,如果没有绑定任何集群,则该标签会被删除。

操作步骤

- 1. 登录 AnalyticDB 控制台。
- 2. 在页面左上角,选择集群所在地域。
- 3. 在左侧导航栏,单击集群列表。
- 4. 在数仓版(3.0)中,单击目标集群ID。
- 5. 在集群信息页面的集群属性区域,单击编辑标签。
- 6. 单击要删除的标签后的X删除标签。



7. 单击确定,删除标签。

5.3. 根据标签筛选集群

集群绑定标签后, 您可以根据标签筛选集群。

- 1. 登录 AnalyticDB 控制台。
- 2. 在页面左上角,选择集群所在地域。
- 3. 在左侧导航栏,单击集群列表。
- 4. 单击标签,选择标签的标签键和标签值筛选集群。
 - ⑦ **说明** 按标签筛选集群后,如果您需要取消筛选,可以**清除**筛选条件。





6.诊断与优化

6.1. 数据建模诊断

本文介绍如何通过控制台进行分区字段合理性诊断、分布字段倾斜诊断和复制表合理性诊断。

分区字段合理性诊断

AnalyticDB MySQL版在创建表时可以指定分区键来提高数据查询的性能,但是因为数据特征存在未知性或者随着业务特征变化发生变化,同时分区的合理性也与实例的规模相关,所以可能存在分区字段选择不合理的情况。分区字段合理性诊断可以帮助用户监控分区的变化,当分区不合理时及时提醒用户进行表结构的优化,保持或者提高实例的查询性能。

• 分区字段不合理判定规则

如果表中有10%以上的分区记录条数不在合理区间则视为不合理,具体的判定规则见下表:

节点数	二级分区记录数
1	1600 w~16000 w
2	6400 w~64000 w
4~6	12800 w~128000 w
8~14	19200 w~192000 w
16~30	25600 w~256000 w
32~62	51200 w~512000 w
大于64	102400 w~1024000 w

如果分区记录数过少,建议加大分区粒度(如由按天分区改为按月分区),如果分区记录数过多,建议降低分区粒度(如由按月分区改为按天分区)。具体方法,请参见表结构设计。

• 优化方法

- i. 登录云原生数据仓库AnalyticDB MySQL控制台。
- ii. 在页面左上角,选择集群所在地域。在左侧导航栏,单击**集群列表**。
- iii. 在数仓版 (3.0) 页签下, 单击目标集群ID。
- iv. 在左侧导航栏, 单击诊断与优化。
- v. 单击**数据建模诊断**页签,即可在**分区字段合理性诊断**区域查看分区字段合理性诊断信息。



您可以通过该表判断哪些表存在分区不合理的情况,并且根据**不合理分区**的信息来判断该表的哪些分区不合理,从而针对这些表或者分区进行相应的优化。具体方法,请参见表结构设计。

分布字段合理性诊断

AnalyticDB MySQL版在创建表(创建方法,请参见CREATE TABLE)时可以通过 DISTRIBUTED BY HASH 指定分布键,来完成数据在各个存储节点的均匀分布。但是因为数据特征存在未知性或者随着业务特征变化会发生变化,所以可能存在分布字段选择不合理的情况,最终会导致数据倾斜问题,造成各个节点处理数据时资源使用不均衡,子任务出现长尾,最终影响查询性能。分布字段合理性诊断会帮助用户及时观察到不合理的分布字段的使用,减少数据倾斜。

• 分布字段不合理判定规则

数据倾斜计算公式:

- i. 去掉一个最大的shard大小,再计算平均shard大小。
- ii. 大于平均shard大小×threshold倍数的shard,或者小于shard分区/threshold的shard都属于倾斜的shard。threshold表示阈值,可以调整配置,当前默认值是3。

• 优化方法

- i. 登录云原生数据仓库AnalyticDB MySQL控制台。
- ii. 在页面左上角,选择集群所在地域。在左侧导航栏,单击**集群列表**。
- iii. 在数仓版(3.0)页签下,单击目标集群ID。
- iv. 在左侧导航栏, 单击诊断与优化。
- v. 单击**数据建模诊断**页签,即可在**分布字段倾斜诊断**区域查看分布字段倾斜诊断信息。

在倾斜表中,您可以查看某个数据库中已经存在倾斜的表,重新进行分布字段的选择。具体方法,请参见表结构设计。

复制表合理性诊断

AnalyticDB MySQL版在创建表时可以指定数据的分布方式为复制方式(DIST RIBUT ED BY BROADCAST),复制表会在实例的每个存储节点保存一份完整的表数据,您在使用其他表和复制表JOIN时,不需要在网络间对复制表进行重分发,从而提升查询的性能,但是复制表在写入数据时会存在写入放大的问题,影响AnalyticDB MySQL版整体写入性能。复制表合理性诊断可以帮助用户检测到这些不合理的复制表。

● 复制表不合理判定规则

复制表单表大于2万条记录即视为不合理。

• 优化方法

- i. 登录云原生数据仓库AnalyticDB MySQL控制台。
- ii. 在页面左上角,选择集群所在地域。在左侧导航栏,单击**集群列表**。
- iii. 在数仓版 (3.0) 页签下, 单击目标集群ID。
- iv. 在左侧导航栏, 单击诊断与优化。
- v. 单击数据建模诊断页签,即可在复制表合理性诊断区域查看复制表合理性诊断信息。

您可以通过页面直接查看诊断结果,或使用COUNT函数查看复制表中的记录数,然后和建议值进行比较。

6.2. 库表结构优化

AnalyticDB MySQL版提供的库表结构优化功能,为用户提供降低集群使用成本,提高集群使用效率的调优建议。本功能持续收集SQL查询的性能指标及其使用到的数据表、索引等信息,并进行算法统计分析,自动给出调优建议,减少用户手动调优的负担。

背景信息

随着各类数据分析业务的丰富和发展,数据库所承载的查询数量和复杂度持续增加。库表结构的设计和优化对数据库整体使用成本和查询性能的影响尤其显著。要做好库表结构设计和优化,用户通常需要关注以下信息:

● 数据库引擎架构

用户需要了解数据库引擎的存储和计算架构特点,才能结合业务的数据分布特征与业务场景特征,来进行数据建模,从而设计出符合引擎架构特点的数据表结构。

● SOL特征差异

即席查询的SQL变化较大,包括参与JOIN的表个数、JOIN条件、分组聚合的字段个数以及过滤条件等。

● 数据特征差异

用户的数据分布和查询特征会随着业务特征的变化而变化,若一直按照最初的建模方式和SQL语句,将无法保障SQL引擎发挥出最大优势,而数据特征或业务模型的变化,都会导致SQL性能回退。

基于此,AnalyticDB MySQL版为用户提供了高效且智能的库表结构优化功能,直观地为用户提供调优建议,来降低集群使用成本,提高集群使用效率。

注意事项

- 3.1.4以上版本的AnalyticDB MySQL版支持库表结构优化功能。
- 库表结构优化功能的调优建议是根据用户数据和查询特征的历史数据分析得到的。当用户数据和查询特征 稳定,则相关建议的有效性可持续;当用户数据和查询特征发生剧烈变化,则由历史数据分析得到的调优 建议的参考价值明显下降。在使用该功能前,请用户根据自身业务特征的变化情况考虑是否采用相应建 议。

功能说明

库表结构优化功能提供冷热数据优化、索引优化和分布键优化三种类型的优化建议。

冷热数据优化

分析数据表的使用情况,对长期未使用的数据表,建议将其迁移至冷盘存储,以减低数据表存储成本。此类型的优化建议针对数据表级别。迁移数据表到冷盘存储的具体操作,请参见数据存储冷热分离。

② 说明 数据表迁移至冷盘后,该数据表的查询时间会变长。

索引优化

分析数据索引的使用情况,对长期未使用的数据索引,建议将其删除,以降低数据索引的存储成本。数据库利用索引进行查询数据过滤的原理,请参见过滤条件不下推。

② 说明 数据索引删除后,如果查询再次使用该数据列进行过滤操作,查询时间会变长。

分布键优化

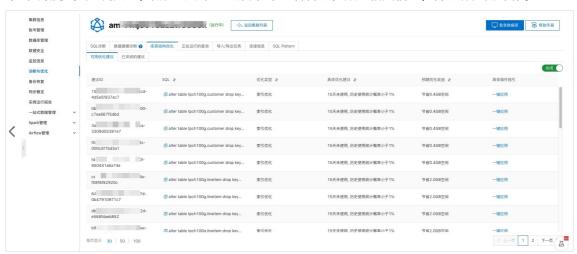
分布键优化通过算子级别的查询分析和智能优化算法,综合考虑了分布键选取原则和注意事项,自动提出合理的分布键优化建议,同时还提供了收益预估,为用户进行数据库分布键设计和优化提供了简单易用的指引。此类型的优化建议针对数据表级别。

分布键优化会基于建议的分布键为用户提供建表语句,但建议的分布键需要保证数据是均匀分布的,目前系统对建议分布键的数据分布统计只是估算值,用户需要自行查询相关列的数据分布情况,以确保不存在大量重复值导致的数据分布不均匀。选择分布键的规则,请参见选择分布键。

AnalyticDB MySQL版目前暂不支持修改分布键并自动重分布数据,因此用户需要重新建表并导入数据,具体方法,请参见更改分区键/分布键。如果有增量数据,还需要考虑如何保证增量数据导入的正确性。

操作步骤

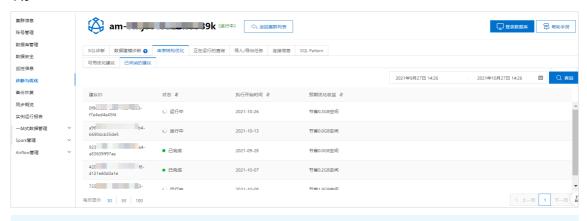
- 1. 登录云原生数据仓库AnalyticDB MySQL控制台。
- 2. 在页面左上角,选择集群所在地域。
- 3. 在左侧导航栏,单击集群列表。
- 4. 在3.0集群列表页签下,单击目标集群ID。
- 5. 在左侧导航栏,单击诊断与优化。
- 6. 单击库表结构优化页签。
- 7. 在**可用优化建议**页签下,单击右上角的**开启**,打开库表结构优化功能。若当前实例已开启库表结构优化功能,可跳过本步骤。
- 8. 单击可用优化建议和已采纳的建议,用户可以查看可用优化建议和已采纳的建议。
 - 单击可用优化建议页签,用户可以在查询列表中查看各项参数的相关信息,并执行应用操作。



参数	说明
建议ID	优化建议的ID。
SQL	优化建议需要变更的表和相应定义等。
优化类型	优化类型分为:冷热数据优化、索引优化和分布键优化。
具体优化建议	针对优化类型,给出的具体优化建议。



○ 在**已采纳的建议**页签下,选择查询时间后单击右上角**查询**,按采纳建议日期和时间查询已采纳的建议。



② 说明 建议SQL下发后,需要数据表完成Build方可完成应用,Build操作是数据库系统按一定规则自动触发的,未触发前,相应建议处于"运行中"的状态,触发后变为"已完成"状态。

6.3. SQL Pattern

AnalyticDB MySQL版的实时统计分析功能提供了SQL Pattern指标特性,有效提升智能诊断效率,能够对相似的SQL进行聚合,并呈现出聚合后的特性。

背景信息

SQL Pattern是依托于全量且实时的SQL而产生的功能,通过对SQL进行分类诊断和分析,聚合相似SQL为一类SQL,有效提升智能诊断的效率。同时,SQL Pattern的聚合结果能作为数据库优化的有效依据。针对导致实例压力过大的Bad SQL,用户还可以通过拦截功能,对其进行拦截,快速恢复实例,详情请参见Persist plan and Query-Blocker。

功能说明

SOL Pattern具备以下功能。

- Pattern聚合:根据SQL文本,将相似的SQL进行分组。
- Pattern宏观统计:建立均值与最值比较。
- 异常SQL排查:排查出异于历史的Pattern并提供钻取功能,直达问题SQL,获取诊断结果。

操作步骤

- 1. 登录云原生数据仓库AnalyticDB MySQL控制台。
- 2. 在页面左上角,选择集群所在地域。
- 3. 在左侧导航栏,单击集群列表。
- 4. 在3.0集群列表页签下,单击目标集群ID。
- 5. 在左侧导航栏,单击诊断与优化。
- 6. 单击SQL Pattern页签。

默认展示最近30分钟的SQL Pattern。您可以通过SQL关键字或时间范围搜索SQL Pattern。



SQL Pattern列表各项参数说明如下:

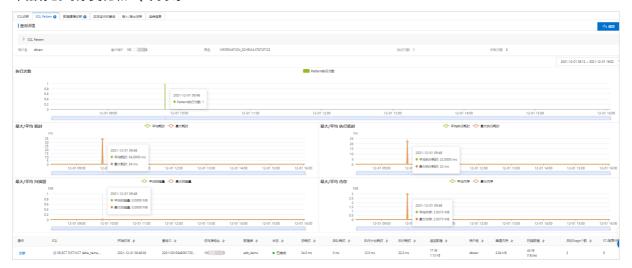
参数	说明
操作	单击 查看详情 ,查看特定SQL Pattern的详细信息。更多信息,请参见 <mark>查询详</mark> 情。
SQL Pattern	SQL Pattern的具体语句。您可以单击
用户名	提交Pattern相关SQL的数据库用户名。
客户端IP	提交Pattern相关SQL的客户端IP地址。
表名	SQL Pattern扫描的数据库表。
创建时间	在查询时间范围内,Pattern相关SQL的最早提交时间。

参数	说明
执行次数	在查询时间范围内,Pattern相关SQL的执行次数。如果实例出现资源水位突增的情况,用户可以通过执行次数排序,查看实例运行异常时段哪些SQL执行次数较多,并进一步分析这些SQL执行次数突增是否正常。
失败次数	在查询时间范围内,Pattern相关SQL的执行失败次数。
平均总耗时	在查询时间范围内,Pattern相关SQL的平均查询耗时,单位毫秒(ms)。
最大总耗时	在查询时间范围内,Pattern相关SQL的最大查询耗时,单位毫秒(ms)。通过比较最大总耗时和平均查询时间,可以了解指定Pattern的耗时增加是否受到其他SQL的影响。如果在实例运行正常时段内最大总耗时和平均查询时间接近,那么在实例运行异常时段当前pattern的耗时增加可能是受到了其他SQL的影响。
平均执行耗时	在查询时间范围内,Pattern相关SQL的平均执行耗时,单位毫秒(ms)。
最大执行耗时	在查询时间范围内,Pattern相关SQL的最大执行耗时,单位毫秒(ms)。执行耗时的定义,详见 <mark>查询监控图和查询列表介绍</mark> 。
平均峰值内存	在查询时间范围内,Pattern相关SQL的平均峰值内存,单位Byte。
最大峰值内存	在查询时间范围内,Pattern相关SQL的最大峰值内存,单位Byte。通过比较最大峰值内存和平均峰值内存,您可以判断指定SQL使用的内存资源的稳定性。如果最大峰值内存远大于平均峰值内存,说明当前Pattern的SQL可能存在扫描数据量增加或者执行计划改变的问题,可以单击查看详情,在查询详情页面,查看SQL列表进一步定位原因。
平均扫描量	在查询时间范围内,Pattern相关SQL的平均扫描量,单位MB。
最大扫描量	在查询时间范围内,Pattern相关SQL的最大扫描量,单位MB。通过比较最大扫描量和平均扫描量,可以判断同类Pattern的SQL的数据读取量的稳定性。如果出现最大扫描量远大于平均扫描量的情况,说明数据读取量不稳定,则需判断是否为正常情况。

查询详情

查询详情页面以图形化的方式展示SQL在时间维度上的执行次数、查询耗时、执行耗时、扫描量以及峰值内存等重要指标,其中查询耗时、执行耗时、扫描量和峰值内存指标分别提供了最大值和平均值,方便用户进行比较和分析。SQL列表提供了当前Pattern在分析时段内的所有相关的SQL,您可以单击诊断,查看SQL的诊断结果和执行计划,具体操作,参见使用执行计划分析查询。

在**SQL Pattern**页签的SQL Pattern列表中,单击**查看详情**,进入**查询详情**页面,您可以查看SQL Pattern多个指标的时序变化和SQL列表。



相关API

- DescribeSQLPatterns
- DescribePatternPerformance

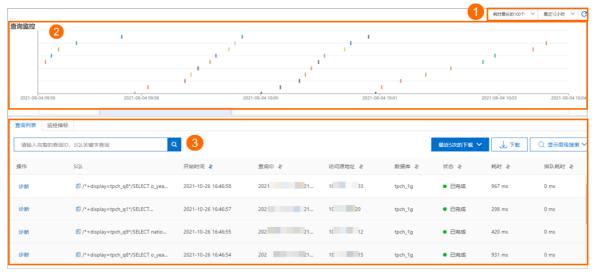
6.4. 通过SQL诊断功能分析执行计划

6.4.1. SQL诊断功能介绍

AnalyticDB MySQL版集群提供了SQL诊断功能,支持通过多种条件检索出符合条件的SQL查询(如慢查询),并将检索结果以图形化的方式展示,您还可以将检索结果下载保存到本地进行查看。本文介绍如何进入SQL诊断页面以及支持的检索方式。

进入SQL诊断页签

- 1. 登录云原生数据仓库AnalyticDB MySQL控制台。
- 2. 在页面左上角,选择集群所在地域。
- 3. 在左侧导航栏,单击集群列表。
- 4. 在数仓版(3.0)页签下,单击目标集群ID。
- 5. 在左侧导航栏,单击**诊断与优化**即可进入**SQL诊断**页签。



序号	介绍
0	SQL查询的检索条件。支持选择不同维度和时间范围来检索出符合条件的SQL查询信息。 更多详情,请参见 <mark>检索方式。</mark>
2	以图像化的方式展示SQL查询的检索结果。更多关于查询监控图的详情,请参见 <mark>查询监控</mark> 图。
3	以查询列表展示SQL查询的检索结果。更多关于查询列表的详情,请参见 <mark>查询列表</mark> 。

检索方式

AnalyticDB MySQL版集群支持通过整体检索和高级搜索两种方式来获取SQL查询详情:

● 整体检索

- 检索条件: 支持检索出在指定时间范围(如最近5分钟)内,不同SQL查询状态(例如已完成的查询) 或查询耗时(例如长耗时查询(>1min))的SQL查询详情。
- **结果展示**:整体检索的结果会在**查询监控**和SQL**列表**中展示。更多**查询监控**和SQL**列表**的详情,请参见查询监控图和查询列表介绍。

? 说明

- 默认展示最近5分钟内耗时最长的100个SQL查询(不包含正在执行中的查询)。
- AnalyticDB MySQL版支持自定义检索最近2周内的SQL查询详情。自定义检索的结束时间需晚于开始时间,且开始和结束时间间隔不能超过24小时。

● 高级搜索

○ **搜索条件**: 高级搜索功能可以根据内存、扫描量、用户名、数据库名、资源组等条件对整体检索的结果进行筛选。其中用户名、数据库名、资源组的可选取值是整体检索结果中已有的取值,而不是当前AnalyticDB MySQL版集群中的所有取值。

例如,AnalyticDB MySQL版集群中有3个数据库: db1 、 db2 、 db3 ,而符合整体检索条件的查询仅涉及 db1 和 db2 。此时查询列表右上角高级搜索中的数据库选择范围仅为 db1 和 db2 。

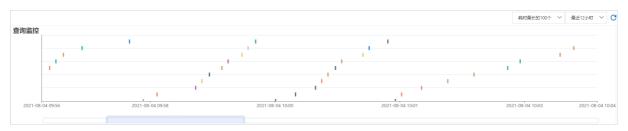
○ 结果展示: 高级检索的筛选结果仅在SQL列表中展示, 而不会影响查询监控中的结果。

6.4.2. 查询监控图和查询列表介绍

AnalyticDB MySQL版集群提供了SQL诊断功能,支持通过多种维度检索出符合条件的SQL查询(如慢查询),并将检索结果以图像化的方式展示,您还可以将检索结果下载保存到本地进行查看。本文介绍如何使用SQL诊断中的查询监控图和查询列表。

杳询监控图

您可以在SQL诊断中的**查询监控**区域查看指定时间范围内,查询执行耗时的分布情况,方便快速定位执行耗时长的查询。



? 说明

- 图中每个色块即代表一次查询。
- 矩形的颜色没有特殊含义,仅用来区分不同查询。色块长度越长,表示对应查询的执行耗时越 长。
- 将鼠标放在色块上即可查看对应查询的相关信息(例如,查询开始或结束时间、扫描量等),单 击详情即可进入目标查询的详情页来查看查询属性、查询语句和执行计划等信息。更多详情, 请参见使用执行计划分析查询。
- **查询监控**中最多展示1万个查询,每个查询的具体信息会在监控图下方的**SQL列表**中展示。更多 详情,请参见查询列表。
- 查询监控仅展示整体检索(如检索最近5分钟内耗时最长的100个查询)的结果,而SQL列表右上角的高级搜索不会影响查询监控的结果展示。更多详情,请参见检索方式。

查询列表

相较于查询监控图,查询列表能够展示SQL查询的更多信息,例如具体的SQL语句、访问源地址等。同时您还可以通过查询列表右上角的高级搜索功能对列表中的结果进行进一步筛选(这里的筛选结果并不会影响查询监控图的结果展示),帮助您更精准地定位到有问题的查询。



字段名	字段说明	使用说明
SQL	SQL语句。	用户可以单击。即复制完整SQL。
开始时间	查询提交的时间。	无
访问源地址	与客户端建立连接的IP地址。	在 诊断与优化 页面,单击 连接信息 ,查看访问源 地址。
数据库	客户端建立连接的数据库名。	用户可以使用高级搜索功能搜索出数据库变更的 SQL。
状态	SQL语句的执行状态,包括: 已完成运行中失败	无
总耗时	SQL语句从提交到最终执行完 成的总耗时。	总耗时=排队耗时+执行计划耗时+执行耗时。
排队耗时	SQL语句从提交到AnalyticDB MySQL版接入节点后到开始执 行前的排队时间。	AnalyticDB MySQL版在接入节点进行了流量控制,当并发较大时,超过流量控制队列大小的查询会进行排队,且排队时间较长。如果此时计算资源水位较低,可以尝试调大队列,具体操作,请参见查询流量控制;如果此时计算资源水位较高,则需要进行Bad SQL查找并定位原因。关于接入节点的详情,请参见整体架构。
执行计划耗时	生成执行计划的耗时。	SQl语句提交到接入节点后会生成执行计划,执行计划耗时较长的可能原因包括: • SQL语句比较复杂; • 接入节点压力较大。 如果SQL语句较复杂,可以尝试降低SQL复杂度,如减少Join的个数;如果接入节点压力较大,可以查看是否有较多数据返回到客户端或查询并发是否较大,具体详见Query级别诊断结果。
执行耗时	SQL语句的执行耗时。	可以根据执行耗时字段排序,排除排队耗时和执行 计划耗时的影响,查找到消耗资源的SQL。
返回数据	Select语句返回到客户端的数 据量。	返回到客户端的数据量不宜过大,数据量过大会导致查询占用前端队列资源,影响其他查询的提交和执行。用户可以根据返回数据大小进行排序,找到返回数据量较大的查询。
用户名	客户端建立连接时使用的用户名。	在 诊断与优化 页面,单击 连接信息 ,查看用户名 和用户连接数。

字段名	字段说明	使用说明
峰值内存	查询消耗的峰值内存。	AnalyticDB MySQL版在执行SQL的过程中,是分阶段(Stage)执行的,有依赖关系的阶段会串行执行,无依赖关系的阶段会并行执行,因此查询消耗的内存存在一个峰值,峰值内存可以体现查询对于内存资源的占用量。
扫描数据	查询从存储层返回到计算层的 数据量。	扫描数据量的大小在一定程度上能体现查询对存储层的压力,读取的数据量越大,消耗的磁盘IO资源也越大,在计算层的处理也会消耗更多资源,影响查询速度。
总的Stage个数	查询生成的总的Stage个数。	Stage的个数多少,在一定程度上能体现一条查询语句的复杂度,Stage个数越多,AnalyticDB MySQL版在执行过程中需进行的网络交互就越多,对系统整体的压力也越大。如果某些查询的Stage个数较多,需要进行相应的调优。关于Stage的概念,请参见查询流程和执行计划。
ETL的写表行数	ETL相关SQL写出到目标表的 行数。	无

您可以在SQL列表页签下执行如下操作:

支持的操作	说明	
	单击 下载 即可将当前的查询结果保存为Excel表格,查询结果表格会自动保存在 最近5次的下载 的下拉列表中。	
下载	说明 当前单次最多支持下载10万条记录。	
最近5次的下载	从 最近5次的下载 下拉列表中选择需要下载的SQL查询搜索结果即可完成下载。	
高级搜索	单击 显示高级搜索 ,根据业务需要选择查询维度,例如按照 资源组 或 访问源地址 维度对查询监控的搜索结果进行进一步的筛查。	
诊断	单击操作栏中的 诊断 ,即可进入目标查询的详情页来查看 查询属性、查询语句 和 执行 计划等信息。更多详情,请参见使用执行计划分析查询。	

? 说明

- SQL列表右上角的高级搜索支持对整体检索的结果进行进一步筛选,筛选条件的可选取值是整体检索结果中已有的取值,而不是当前AnalyticDB MySQL版集群中的所有取值。更多详情,请参见检索方式。
- 当整体检索条件选择**正在运行的查询**时,检索结果仅显示执行耗时超过10s的SQL查询详情,且在SQL**列表**中会展示**资源消耗排名**列,资**源消耗排名**的数值越小,说明该查询消耗的CPU、内存等资源越多。
- **SQL列表的SQL**列,最长支持显示5120个字符,超出字符限制的SQL语句会被截断显示。您可以在下载的Excel文件中,或目标查询详情页的**查询语句**页签下查看完整的SQL。

6.4.3. 使用执行计划分析查询

AnalyticDB MySQL版的SQL诊断功能支持以树形图的形式展现SQL查询的执行计划。执行计划树分为两层:第一层是Stage层,第二层是算子(Operator)层。本文介绍如何使用Stage层和算子层执行计划树来分析查询。

Stage层执行计划树

Stage层执行计划树由多个Stage节点组成,数据流向自下而上,先由具有扫描算子的Stage进行数据扫描,再经过中间Stage节点的层层处理后,再由最上层的根节点将查询结果返回客户端。



Stage层执行计划树中主要包含如下信息:

● 基本信息

图中的每个矩形框代表一个Stage, 框里会包含Stage ID、数据输出类型、耗时或内存(选择**按内存**排序时展示)等信息。

② 说明 当Stage层执行计划树上出现红色警示号,表示该Stage被诊断出存在可优化点。

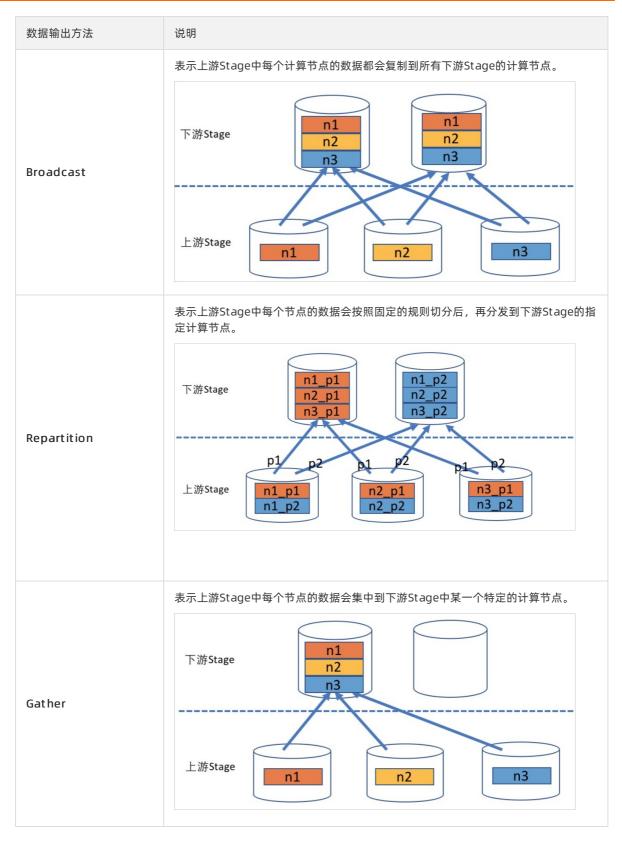
● 数据输出行数

两个相邻Stage间连线上的数字表示上游Stage向下游Stage输出的数据行数。数据输出行数越多,Stage间的连线越粗。

● 数据输出方法

表示在两个相邻的Stage间,上游向下游Stage传输数据时所用的方法。AnalyticDB MySQL版支持以下数据输出方法。

数据输出方法 说明



● 查看内存使用率或执行耗时Top 10的Stage详情

执行计划树右侧的**Top10耗时/内存倒序**页签下,会展示执行耗时占总查询耗时比例最大,或使用内存占总查询使用内存比例最大的前10个Stage的ID和对应的比例。

? 说明

- 默认**按耗时**排序,您也可以选择执行计划树右上角的**按内存**排序。
- Top10耗时/内存倒序页签下不会显示内存使用量或执行耗时占比小于1%的Stage数据。
- 。 由于统计方式的差异,可能存在一个查询中的所有Stage耗时或内存的百分比之和不等于 100%的情况。

● 诊断结果

单击执行计划树中某个Stage(如Stage[1]),即可在右侧查看对应Stage的**诊断结果**详情,包括如下两类诊断:

- **Stage诊断**: 这类诊断结果包含了对目标Stage诊断结果的详细说明,包括诊断出的问题(如存在较大的数据量被广播或数据倾斜)以及对应的调优方案。
- **算子诊断**: 这类诊断仅展示当前Stage中有问题的算子名称以及对应问题的概述,详细的说明和调优方案需要到算子执行计划树中才能看到。更多详情,请参见<mark>算子层计划执行树</mark>。

关于Stage的诊断结果的说明,请参见Stage级别诊断结果。

● 统计信息

诊断结果下方展示了目标Stage中各项指标的统计信息。

指标项	说明
峰值内存	目标Stage的峰值内存消耗量。系统会根据实际内存消耗量选择Bytes、 KB、MB、GB或TB为单位。
累积耗时	目标Stage内存中所有算子多节点多线程执行时耗时的累加,系统会根据实际耗时选择ms、s、m或h为单位。
	⑦ 说明 该累积耗时不能和当前查询的总耗时进行比较。
输出行数	目标Stage输出的行数。
输出大小	目标Stage输出的数据量。系统会根据实际数据量选择Bytes、KB、MB、GB 或TB为单位。
输入行数	目标Stage输入的行数。
输入大小	目标Stage输入的数据量。系统会根据实际数据量选择Bytes、KB、MB、GB或TB为单位。
	目标Stage扫描的行数。
扫描行数	? 说明 仅当目标Stage中存在数据扫描算子时支持该参数。

目标Stage扫描的数据量。系统会根据实际数据量选择Bytes、KB、	
或TB为单位。	ИВ、GB
扫描量 ② 说明 仅当目标Stage中存在数据扫描算子时支持该参数。	

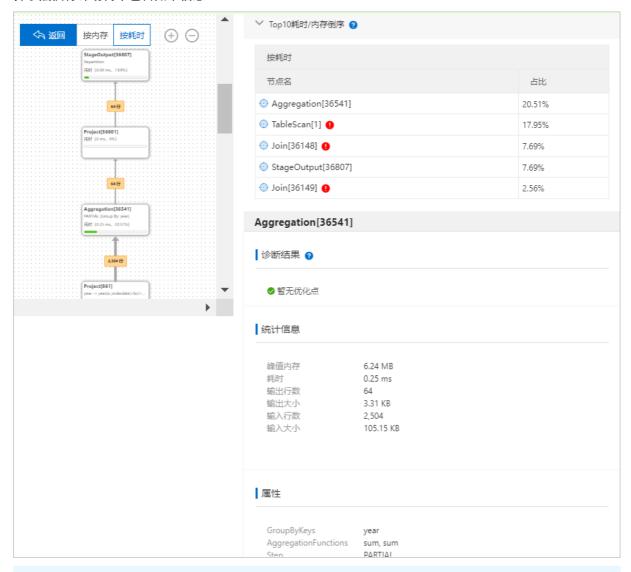
算子层计划执行树

算子层执行计划由多个算子组成,图中的每个矩形框代表一个算子,数据流向自下而上,扫描数据过程或接收网络数据由最上游的算子(TableScan和RemoteSource)完成,扫描到的数据和接收到的网络数据经过中间算子层层处理后,再由最上层的根节点算子(StageOutput或Output)将数据输出到下游Stage或将查询结果返回给客户端。

您可以将鼠标移动到目标Stage上,在弹出的信息框中单击**查看Stage计划**,即可进入对应Stage的计划详情页来查看算子层计划执行树。



算子层执行计划树中包含如下信息:



② 说明 当算子层执行计划树上出现红色警示号,表示该算子被诊断出存在可优化点。

● 基本信息

图中的每个矩形框代表一个算子,框里会包含算子名称及ID、算子属性(如Join算子的Join条件、算法等)、耗时或内存(选择按内存排序时展示)等信息。

● 数据输出行数

两个相邻算子间连线上的数字表示上游算子向下游算子输出的数据行数。数据输出行数越多,算子间的连线越粗。

● 查看内存使用率或执行耗时Top 10的算子详情

执行计划树右侧的Top10耗时/内存倒序页签下,会展示执行耗时占总查询耗时比例最大,或使用内存占总查询使用内存比例最大的前10个算子的ID和对应的比例。

? 说明

- 默认**按耗时**排序,您也可以选择执行计划树右上角的**按内存**排序。
- Top10耗时/内存倒序页签下不会显示内存使用量或执行耗时占比小于1%的算子数据。
- 。由于统计方式的差异,可能存在一个Stage中的所有算子耗时或内存的百分比之和不等于 100%的情况。

● 诊断结果

单击执行计划树中某个算子(如**Join**[36148]),即可在右侧查看对应算子的**诊断结果**详情,包括诊断出的问题(如存在数据膨胀或Join的右表过大)以及对应的调优方案。更多关于算子诊断的详情,请参见算子级别诊断结果。



● 统计信息

诊断结果下方展示了目标算子中各项指标的统计信息。

指标项	说明	
峰值内存	目标算子的峰值内存消耗量。系统会根据实际内存消耗量选择Bytes、KB、MB、GB或TB为单位。	
耗时	目标算子在一定并发度上的平均耗时,系统会根据实际耗时选择ms、s、m或h为单位。	
	② 说明 该耗时可以和当前查询的耗时进行比较。	
输出行数	目标算子输出的行数。	
输出大小	目标算子输出的数据量。系统会根据实际数据量选择Bytes、KB、MB、GB或TB为单位。	
输入行数	目标算子输入的行数。	

指标项	说明
输入大小	目标算子输入的数据量。系统会根据实际数据量选择Bytes、KB、MB、GB或TB为单位。
Builder统计信息	包括Builder的类型、峰值内存、耗时、输出输入行数和数据量等信息。主要有如下Builder类型: HashBuiler: 用于构建Hash表来完成Hash Join计算。 SetBuilder: 用于构造Set结构来完成Semi Join计算。 NestLoopBuilder: 用于完成Nest Loop Join计算。 ② 说明 仅Join算子支持该统计指标项。
属性	不同算子的属性信息不同,例如Join算子的属性信息中包括了Join的类型、方式等。更多详情,请参见 <mark>算子</mark> 。

7.每日健康报告

本文介绍了如何查看每日健康报告及报告中数据的含义。

查看每日健康报告

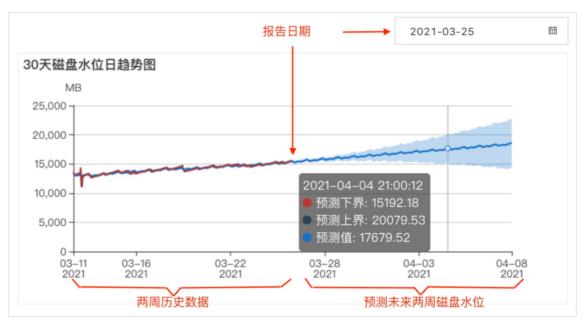
- 1. 使用阿里云账号登录云原生数据仓库AnalyticDB MySQL控制台。
- 2. 在页面左上角,选择集群所在地域。在左侧导航栏,单击集群列表。
- 3. 在3.0集群列表中,单击目标集群ID。
- 4. 单击实例运行报告 > 每日健康报告。

AnalyticDB MySQL版支持按天统计所选日期的主要指标的平均值和最大值,指标包括:每秒查询量(QPS)、每秒写入量(TPS)、CPU使用率、磁盘存储IO指标、磁盘总体水位。



同时提供磁盘水位、QPS和TPS的趋势预测图。统计过去一段时间的磁盘水位、QPS和TPS,利用预测算法预测:预测算法基于历史数据规律,建立数学模型,进行最优近似,因此不能预警异常或者突发情况,请参考相关告警设置来进行异常报警。

○ 按天粒度聚合的未来两周磁盘水位趋势图,可以根据这个预测趋势判断未来磁盘使用量,便于提前规划扩容磁盘、增加worker数量或者删除无用数据。



- 按小时粒度聚合的QPS未来一天趋势图,可以根据查询量趋势提前规划业务、升降配或者弹性扩容。
- 按小时粒度聚合的TPS未来一天趋势图,可以根据写入量趋势提前规划业务或者扩容磁盘。

