

ALIBABA CLOUD

阿里云

云原生数据仓库AnalyticDB
MySQL版
快速入门

文档版本：20201214

 阿里云

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。	 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。
 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。	 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 <code>cd /d C:/window</code> 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	<code>bae log list --instanceid</code> <i>Instance_ID</i>
[] 或者 [a b]	表示可选项，至多选择一个。	<code>ipconfig [-all -t]</code>
{ } 或者 {a b}	表示必选项，至多选择一个。	<code>switch {active stand}</code>

目录

1.使用限制	05
2.AnalyticDB MySQL使用流程	08
3.准备阿里云账号	09
4.创建集群	10
5.创建数据库账号	13
6.设置白名单	15
7.连接集群	16
8.创建数据库	18
9.导入数据并查询	20

1.使用限制

为保障集群的稳定及安全，AnalyticDB for MySQL有部分使用上的约束，详情如下表所示。

对象	命名规则	限制
数据库名	以小写字母开头，可包含字母、数字以及下划线（_），但不能包含连续两个及以上的下划线（_），长度不超过64个字符。	数据库名不能是analyticdb，analyticdb是内置数据库。
表名	以字母或下划线（_）开头，可包含字母、数字以及下划线（_），长度为1到127个字符。	<ul style="list-style-type: none"> 表名不包含引号、感叹号（!）和空格。 表名不能是SQL保留关键字。
列名	以字母或下划线（_）开头，可包含字母、数字以及下划线（_），长度为1到127个字符。	<ul style="list-style-type: none"> 列名不包含引号、感叹号（!）和空格。 列名不能是SQL保留关键字。
账号名	以小写字母开头，小写字母或数字结尾，可包含小写字母、数字以及下划线（_），长度为2到16个字符。	无
密码限制	包含大写字母、小写字母、数字以及特殊字符（!）、（@）、（#）、（\$）、（%）、（^）、（&）、（*）（（））、（_）、（+）、（-）、（=），每个密码至少包含其中三项（大写字母、小写字母、数字以及特殊字符），长度为8到32个字符。	无
表中COMMENT	无	最大长度为1024个字符。
列中COMMENT	无	最大长度为1024个字符。
索引名长度	无	最大长度为64个字符。
列默认值	无	最大长度为127个字符。

购买及配置限制

限制	描述	其他
每个阿里云账号可购买集群数量的最大值	64	无
每个集群可购买节点组的最大值（C4规格除外）	128	C4规格除外
C4规格可购买节点组的最大值	4	无
C4规格可购买磁盘空间范围	100GB~200GB	无
每个集群可购买节点数的最小值	2	无
每个集群可创建数据库的最大值	256	无

限制	描述	其他
每天集群扩容/缩容/变配的总次数	12	无
节点磁盘空间阈值	80%	锁定集群后数据无法写入。
每个集群可创建普通账号的最大值	256	无
每个集群最大连接数	5000	无
每个集群下表数量的最大值	<ul style="list-style-type: none"> • 预留模式 <ul style="list-style-type: none"> ◦ 集群版：1-20个节点组，5000张表，可以通过扩容节点数量增加表数量。 ◦ 基础版： <ul style="list-style-type: none"> ▪ T8：1000 ▪ T16/T32：3000 ▪ T52：5000 • 弹性模式 <p>计算资源1440核以内，5000张表，可以通过扩容节点数量增加表数量。</p> 	无
每张表中列数的最大值	1024	无
每张表中分区数的最大值	102400	无
每个集群中IP白名单个数的最大值	25000	每个集群中最多可创建50个白名单分组，每个分组最多可添加1000个IP。
WHERE条件中ITEM最大值	256	无
IN中条件数的最大值	2000	无
每行写入数据的最大长度	16MB	无
每个字段的最大长度	16MB	无
<code>decimal(m,d)</code> 中m和d的取值范围	m是数值的最大精度，取值范围为1~1000；d是小数点右侧数字的位数，要求d≤m。	无

操作限制

操作	时间限制 (ms)	超时说明
创建集群	600000	无-
删除集群	300000	无-
扩容集群	600000	无
创建数据库	10000	无
删除数据库	10000	无
创建表/视图	10000	无
删除表/视图	10000	无
新增列	10000	无
删除列	60000	无
更新列	60000	无
TRUNCATE TABLE	60000	无
单次查询时间	1800000	<p>超时后，系统自动终止查询。您可以通过更改以下配置，修改单次查询的最大执行时间。</p> <ul style="list-style-type: none"> 配置名：queryTimeout Hint名：query_timeout
写入时间	300000	无
INSERT SELECT FROM	24*3600000	<p>执行超时且任务结束后，您可以更改以下配置，修改INSERT SELECT FROM的最大执行时间。</p> <ul style="list-style-type: none"> 配置名：insertSelectTimeout Hint名：insert_select_timeout
INSERT OVERWRITE	24*3600000	<p>执行超时且任务结束后，您可以更改以下配置，修改INSERT OVERWRITE的最大执行时间。</p> <ul style="list-style-type: none"> 配置名：insertSelectTimeout Hint名：insert_select_timeout

2.AnalyticDB MySQL使用流程

欢迎使用云原生数据仓库AnalyticDB MySQL版入门指南。云原生数据仓库AnalyticDB MySQL版（简称ADB，原分析型数据库MySQL版）是云端托管的PB级高并发实时数据仓库，是专注于服务OLAP领域的数据库。本指南将指引您完成一次AnalyticDB MySQL集群创建及使用。

如果您是首次使用AnalyticDB MySQL的用户，我们建议您先阅读以下部分：

- **产品简介**-本内容概述了AnalyticDB MySQL的产品概念、产品优势及应用场景等内容。
- **产品定价**-本内容介绍了AnalyticDB MySQL的产品定价、计费方式等信息。
- **AnalyticDB MySQL入门指南（本指南）**-本指南提供了有关使用AnalyticDB MySQL创建示例集群并使用示例数据的教程。

在本教程中，操作流程概览如下：



1. **准备阿里云账号**
2. **创建集群**
3. **创建数据库账号**
4. **设置白名单**
5. **连接集群**
6. **创建数据库**
7. **导入数据并查询**

完成本教程后，您可以查通过用户指南查看了解后续步骤等更多信息。[了解更多](#)。

3. 准备阿里云账号

使用云原生数据仓库AnalyticDB MySQL版前，您需要创建一个阿里云账号。本文为您介绍如何创建阿里云账号。

背景信息

云原生数据仓库AnalyticDB MySQL版的账号登录体系与阿里云保持一致，统一采用RAM主子账号登录的方式。

- 阿里云账号（即主账号）是阿里云资源的归属及使用计量计费的基本主体，负责生成本企业组织下的子账号，并对子账号进行管理、授权等操作。
- RAM子账号从属于阿里云主账号，由主账号在RAM系统中创建并进行管理。这些RAM子账号本身不拥有实际的任何资源，也没有独立的计量计费，由所属主账号统一控制和付费。

注册阿里云账号

如果您还没有阿里云账号，请单击[立即注册](#)，进入阿里云账号注册页面完成阿里云账号注册。

阿里云账号实名认证

登录阿里云账号，单击[实名认证](#)，按步骤填写信息完成账号实名认证。

② 说明

- 阿里云账号需要进行实名制认证后，才能购买和使用阿里云上的各种产品。为保证后续操作顺利进行，请务必完成实名认证操作。
- 如果您是 enterprise 级用户，建议进行企业级认证，以获取更多的便利。

4. 创建集群

您可以使用阿里云账号在云原生数据仓库AnalyticDB MySQL版（简称ADB，原分析型数据库MySQL版）控制台上创建集群，集群创建成功后需要为集群设置白名单和创建数据库账号，然后就可以将您的数据同步到集群中，实现毫秒级针对千亿级数据进行即时的多维分析透视和业务探索。

前提条件

- 已注册阿里云账号。具体操作请参见[注册阿里云账号](#)。
- 若您要创建按量付费的集群，请确保您的阿里云账号的余额大于等于100元。

操作步骤

1. 使用阿里云账号登录[ADB控制台](#)。
2. 在左侧导航栏，单击[集群列表](#)。
3. 在[3.0集群列表](#)页面，单击[创建集群](#)。



4. 选择商品类型。

- **按量付费**：属于后付费，即按小时扣费。适合短期需求，用完可以立即释放集群，节省费用。
- **包年包月**：属于预付费，即在新建集群时需要支付费用。适合长期需求，价格比按量付费更实惠，且购买时长越长，折扣越多。

5. 根据页面提示，进行参数配置。



模式	<input checked="" type="radio"/> 预留模式	<input type="radio"/> 弹性模式					
系列	<input checked="" type="radio"/> 集群版	<input type="radio"/> 基础版					
规格	<input checked="" type="radio"/> C8	<input type="radio"/> C24	<input type="radio"/> C32	<input type="radio"/> S8			
节点组数量	<p>8核64GB，计算密集型</p> <p>节点组数量: <input type="text" value="2"/></p> <p>【集群版】 每个节点组包含三个在线节点（副本），相比于主备节点或者双副本而言，不仅增加了可靠性，查询开发度也大幅提升。 计费方面，只要花费两个节点的費用即可买到一个节点组。 参与计算总CPU核数计算方式：规格核数 × 节点组个数 × 3。 【基础版】 与集群版不同，基础版没有节点组概念，只有单节点提供服务。</p>						
存储空间	<input checked="" type="radio"/> 100	<input type="radio"/> 200	<input type="radio"/> 300	<input type="radio"/> 400	<input type="radio"/> 500	<input type="radio"/> 600	<input type="radio"/> 700
	<input type="radio"/> 800	<input type="radio"/> 900	<input type="radio"/> 1000				
	<p>【集群版】 此处选择的是每个节点组的空间大小，实际总空间需要乘以节点组个数。 C8、C24规格节点组磁盘空间可选范围为100GB~1000GB。 C32规格节点组磁盘空间可选范围为100GB~4000GB。 S8规格节点组磁盘空间可选范围为1000GB~12000GB。 【基础版】 与集群版不同，基础版只有一个节点，所买空间即为用户总空间。</p>						

实例参数

参数	说明
版本	选择3.0。
地域	集群所在的地理位置，购买后无法更换地域。一般建议申请离客户业务最近的地域，从而提升用户访问速度。
可用区	可用区是地域中的一个独立物理区域，不同可用区之间没有实质性区别。
网络类型（VPC）	专有网络，也称为VPC（Virtual Private Cloud）。VPC是一种隔离的网络环境，安全性较高。
专有网络交换机	选择专有网络下的交换机。
模式	ADB中产品模式包含 弹性模式 和 预留模式 ，其中预留模式又分为 基础版 和 集群版 ；弹性模式中只包含 集群版 。

参数	说明
系列	<p>ADB中产品系列包括基础版、集群版和集群版（新版）。</p> <ul style="list-style-type: none"> 基础版：单节点提供服务，计算与存储分离，极大降低用户的使用门槛。 集群版：采用三副本高可用架构，提供金融级别的数据可靠性保证。支持节点和存储空间秒级弹性扩展，可承载更大吞吐数据实时写入和读取能力。 弹性模式集群版（新版）：通过存储分离架构，提供存储与计算资源独立、自由弹性伸缩、多用户隔离、存储冷热分离、异构化存储格式以及用户可编程的能力，可以通过较低的资源成本获得更高性能的灵活化实时分析体验。 <div style="border: 1px solid #ccc; background-color: #e6f2ff; padding: 5px; margin-top: 10px;"> <p> 说明 不同地域支持的产品系列不同，请以实际界面为准。各个系列的详细介绍，请参见概述。</p> </div>
规格	<p>预留模式下才有的概念。</p> <p>ECU类型，详情请参见规格详情。</p>
节点组数量	<p>预留模式下才有的概念。</p> <p>节点组的数量，每个节点组默认三副本。</p>
存储空间	<p>预留模式下才有的概念。</p> <p>单个节点组的存储空间。</p>
计算资源	<p>弹性模式才有的概念。</p> <p>计算资源用于数据计算，增加计算资源可以提高数据查询速度，建议计算资源CPU核数多于数据资源CPU核数。</p>
数据资源	<p>弹性模式才有的概念。</p> <p>数据资源用于数据读写，增加数据资源可以提高集群读写速度。</p>
数据存储空间	<p>弹性模式才有的概念。</p> <p>弹性模式下，购买集群时无需选择空间，根据使用量按小时计费。</p>

6. 设置购买时长（仅针对包年包月集群），然后单击右侧的**立即购买**。

7. 在**确认订单**页面，勾选云原生数据仓库AnalyticDB MySQL版服务协议并阅读，根据提示完成支付。

5. 创建数据库账号

ADB包含高权限账号和普通账号两种类型的数据库账号，不同的账号类型对应不同的操作权限。

数据库账号类型

数据库账号类型	说明
高权限账号	<ul style="list-style-type: none"> 只能通过控制台创建和管理高权限账号。 一个集群中只能创建一个高权限账号，高权限账号可以管理所有普通账号和数据库。 使用高权限账号可以断开任意普通账号的连接。 开放了更多权限，可满足个性化和精细化的权限管理需求，例如可按用户分配不同表的查询权限等。 ADB中的高权限账号相当于MySQL中的root账号。
普通账号（子账号）	<ul style="list-style-type: none"> 只能通过数据链路的SQL语句创建和管理普通账号。 一个集群最多可以创建256个普通账号。 需要手动为普通账号授予指定数据库的权限。 普通账号不能断开其他普通账号的数据库连接。

创建高权限账号

1. 使用阿里云账号登录[ADB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集群列表中，单击目标[集群ID](#)。
5. 在左侧导航栏单击[账号管理](#)，然后在[账号管理](#)页面单击[创建账号](#)。
6. 在[创建账号](#)页面，按照页面提示进行参数设置。

参数	说明
数据库账号	高权限账号的账号名。 命名规则：以小写字母开头，小写字母或数字结尾，可包含小写字母、数字以及下划线（_），长度为2个字符~16个字符。
账号类型	高权限账号，不可更改。
密码	高权限账号的密码，密码设置规则如下： <ul style="list-style-type: none"> 包含大写字母、小写字母、数字以及特殊字符。 密码至少包含其中三项（大写字母、小写字母、数字以及特殊字符），长度为8~32个字符。 支持的特殊字符有（!）、（@）、（#）、（\$）、（%）、（^）、（&）、（*）、（()）、（_）、（+）、（-）、（=）。

参数	说明
确认密码	确认高权限账号的密码。
备注说明	备注信息，可选项。

7. 单击**确定**，创建高权限账号。

创建普通账号

创建及授权普通账号，请参见[CREATE USER](#)、[GRANT](#)。

6. 设置白名单

创建ADB集群后，您需要为集群设置白名单，以允许外部设备访问该集群。

背景信息

- 集群默认在白名单只包含IP地址127.0.0.1，表示任何设备均无法访问该集群。
- 白名单可以让ADB集群得到高级别的访问安全保护，建议您定期维护白名单。
- 设置白名单不会影响ADB集群的正常运行。

操作步骤

1. 使用阿里云账号登录[ADB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**中，单击目标**集群ID**。
5. 在左侧导航栏单击**数据安全**。
6. 在白名单设置页面，单击**default**白名单分组右侧的**修改**。

 **说明** 您也可以单击**创建白名单分组**创建自定义分组。



7. 在**修改白名单分组**对话框中，删除默认IP 127.0.0.1，填写需要访问该集群的IP地址或IP段，然后单击**确定**。
 - 若填写IP段，例如10.10.10.0/24，表示10.10.10.X的IP地址都可以访问该集群。
 - 若您需要添加多个IP地址或IP段，请用英文逗号(,) 隔开（逗号前后都不能有空格），例如192.168.0.1,172.16.213.9。
 - 设置白名单后，新的白名单将于1分钟后生效。

7.连接集群

ADB支持通过DMS（Data Management Service）、MySQL客户端（Navicat for MySQL、DBeaver、DBVisualizer、SQL WorkBench/J）、BI可视化工具、或者MySQL命令行工具连接ADB集群。您也可以在应用程序中通过配置集群连接地址、端口、数据库账号等信息连接ADB集群。

背景信息

DMS是阿里云提供的图形化数据管理工具，可用于管理关系型数据库和NoSQL数据库，支持数据管理、SQL操作、数据方案（数据导入/导出、数据库克隆等）、性能与优化、安全审计等功能。

使用DMS连接ADB

1. 使用阿里云账号登录[ADB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集群列表中，单击目标集群ID。
5. 在集群信息页面，单击右上角登录数据库。
6. 填写ADB集群的登录信息。

登录实例

* 数据库类型: ADB3.0-MySQL

* 实例地区: 华东1 (杭州)

* 实例ID: am-...

* 数据库账号: 请输入数据库账户

* 数据库密码: [输入框]

记住密码 ?

测试连接 登录 取消

- 数据库类型：ADB3.0-MySQL。
 - 实例地区：选择集群所属地区。
 - 实例ID：选择集群的集群ID。
 - 数据库账号：集群的账号名称。
 - 数据库密码：账号名对应的密码。
7. 单击测试连接测试能否连通集群，或直接单击登录。

说明 若您希望浏览器记住该账号的密码，可以先勾选记住密码，再单击登录。

应用开发中通过代码连接ADB

- [Java](#)
- [Druid连接池配置](#)
- [Python](#)
- [PHP](#)
- [C# \(Mac\)](#)
- [Golang](#)

通过MySQL命令行工具连接ADB

[MySQL命令行连接AnalyticDB for MySQL](#)

通过客户端连接ADB

- [DBeaver](#)
- [DBVisualizer](#)
- [Navicat](#)
- [SQL WorkBench/J](#)

通过数据可视化工具连接ADB

- [FineBI](#)
- [Quick BI](#)
- [永洪BI](#)
- [DataV](#)
- [Tableau](#)
- [QlikView](#)
- [FineReport](#)

8. 创建数据库

您可以使用DMS（Data Management Service）、客户端（Navicat for MySQL、DBeaver、DBVisualizer、SQL WorkBench/J）、业务系统中的程序代码或者MySQL命令行工具连接ADB集群，然后通过CREATE DATABASE语句创建数据库。

背景信息

本文以DMS为例，介绍如何创建数据库。

 **说明** 每个集群最多可以创建256个数据库。

操作步骤

1. 在SQL INFORMATION_SCHEMA页签下，在SQL Console中输入CREATE DATABASE语句创建数据库。



- 语法： CREATE DATABASE [IF NOT EXISTS] \$db_name
- 参数说明： db_name ：数据库名。以小写字母开头，可包含字母、数字以及下划线（_），但不能包含连续两个及以上的下划线（_），长度不超过64个字符。

 **说明** 数据库名不能是analyticdb，analyticdb是内置数据库。

- 示例：

```
create database adb_demo;
```

```
create database if not exists adb_demo2;
```

2. 单击左上角的执行，数据库创建成功。



9.导入数据并查询

AnalyticDB for MySQL提供多种数据同步方案，可满足不同场景下的数据同步需求。本文以数据文件存储在OSS中为例，介绍如何将OSS中的数据文件导入AnalyticDB for MySQL的 `adb_demo` 数据库中并进行查询。更多导入数据方式请参见[同步数据](#)。

前提条件

- 通过以下步骤在[对象存储](#)（Object Storage Service，简称OSS）中创建存储AnalyticDB for MySQL数据的目录。

- i. [开通OSS服务](#)

 **说明** OSS与AnalyticDB for MySQL所属Region相同。

- ii. [创建存储空间](#)

- iii. [新建目录](#)

- iv. [上传测试数据文件](#)

本示例将 `oss_import_test_data.txt` 文件上传至OSS中的 `bucket-name.oss-cn-hangzhou.aliyuncs.com/adb/` 目录，数据行分隔符为换行符，列分隔符为 `;`，文件示例数据如下所示。

```
number;note
0001;hello_world_1
0002;hello_world_2
0003;hello_world_3
0004;hello_world_4
0005;hello_world_5
0006;hello_world_6
...
```

- 根据AnalyticDB MySQL入门指南，完成创建集群、设置白名单、创建账号和数据库等准备工作。

操作步骤

- 通过**CREATE TABLE**，在 `adb_demo` 数据库中创建外表。创建CSV、Parquet或TEXT格式OSS外表的建表语法请参见[创建OSS外表语法](#)。
- 查询OSS数据。

查询外表映射表和查询AnalyticDB for MySQL内表语法没有区别，您可以方便地直接进行查询，如本步骤的示例代码所示。

```
select uid, other from oss_parquet_external_table where uid < 100 limit 10
```

- 对于数据量较大的CSV/TEXT数据文件，强烈建议您按照后续步骤导入AnalyticDB for MySQL后再做查询，否则查询性能可能会较差。
- 对于Parquet格式数据文件，直接查询的性能一般也比较高，您可以根据需要决定是否进一步导入到AnalyticDB for MySQL后再做查询。

- 通过 `CREATE TABLE`，在 `adb_demo` 数据库中创建目标表 `adb_oss_import_test` 存储从OSS中导入的数据。

```
CREATE TABLE IF NOT EXISTS adb_oss_import_test
(
  uid string,
  other string
)
DISTRIBUTED BY HASH(uid)
```

- 执行 `INSERT` 语句将OSS数据导入AnalyticDB for MySQL。

- 执行 `INSERT INTO` 导入数据。

```
insert into adb_oss_import_test
select * from oss_import_test_external_table
```

- 执行 `INSERT OVERWRITE INTO` 导入数据。

```
insert overwrite into adb_oss_import_test
select * from oss_import_test_external_table
```

- 异步执行 `INSERT OVERWRITE INTO` 导入数据。

```
submit job insert overwrite into adb_oss_import_test
select * from oss_import_test_external_table ;
+-----+
| job_id          |
+-----+
| 2020112122202917203100908203303000715 |
```

关于异步提交任务详情请参见 [异步提交导入导出任务](#)。

创建OSS外表语法

创建OSS CSV格式外表

示例的 `oss_import_test_data.txt` 文件为CSV格式，本节介绍CSV格式的OSS外表创建语法。

```
CREATE TABLE IF NOT EXISTS oss_import_test_external_table
(
  uid string,
  other string
)
ENGINE='OSS'
TABLE_PROPERTIES='{
  "endpoint":"oss-cn-hangzhou-internal.aliyuncs.com",
  "url":"oss://$bucketname/adb/oss_import_test_data.txt",
  "accessid":"LTAIF****5FsE",
  "accesskey":"Ccw****iWjv",
  "delimiter":";",
  "skip_header_line_count":1
}'
```

参数	说明
ENGINE='OSS'	表示该表是外部表，使用的存储引擎是OSS。
TABLE_PROPERTIES	用于告知AnalyticDB for MySQL如何访问OSS中的数据。
endpoint	<p>OSS的 EndPoint（域名节点）。</p> <p> 说明 目前仅支持AnalyticDB for MySQL通过OSS中ECS的VPC网络（内网）访问OSS。</p> <p>登录OSS控制台，单击目标Bucket，在Bucket概览页面查看 endpoint。</p>
url	<p>OSS中目标数据文件夹或文件的绝对地址。</p> <p>示例：oss://\$Bucket名称/adb/oss_import_test_data.txt</p>
accessid	<p>您在访问OSS中的文件时所持有的AccessKey ID。</p> <p>如何获取您的accessid和accesskey，请参见获取账号的AK信息。</p>
accesskey	您在访问OSS中的文件时所持有的Access Key Secret。
delimiter	定义CSV数据文件的列分隔符。

参数	说明
ossnull (可选)	<p>标识 NULL 值，包含以下四种取值。</p> <ul style="list-style-type: none"> • 默认值为1: EMPTY_SEPARATORS a,"",,c-->"a","",NULL,"c" • EMPTY_QUOTES a,"",,c-->"a",NULL,"","c" • BOTH a,"",,c-->"a",NULL,NULL,"c" • NEITHER a,"",,c-->"a","",,"","c"
skip_header_line_count (可选)	<p>CSV文件第一行为表头，导入数据时设置为1可自动跳过第一行。或设置为其它值表示开头跳过的行数。默认为0，也就是不跳过。</p>

创建OSS Parquet格式外表

```
CREATE TABLE IF NOT EXISTS oss_parquet_external_table
(
  uid string,
  other string
)
ENGINE='OSS'
TABLE_PROPERTIES='{
  "endpoint":"oss-cn-hangzhou-internal.aliyuncs.com",
  "url":"oss://****",
  "accessid":"LTAIF****5FsE",
  "accesskey":"Ccw****iWjv",
  "format":"parquet"
}'
```

参数	说明
ENGINE='OSS'	表示该表是外部表，使用的存储引擎是OSS。
TABLE_PROPERTIES	用于告知AnalyticDB for MySQL如何访问OSS中的数据。

参数	说明
endpoint	<p>OSS的 EndPoint（域名节点）。</p> <p> 说明 目前仅支持AnalyticDB for MySQL通过OSS中ECS的VPC网络（内网）访问OSS。</p> <p>登录OSS控制台，单击目标Bucket，在Bucket概览页面查看 endpoint。</p>
url	<p>OSS中目标数据文件或文件夹的绝对地址。</p> <p>示例：oss://\$Bucket名称/adb/oss_import_test_data.txt</p>
accessid	<p>您在访问OSS中的文件时所持有的AccessKey ID。</p> <p>如何获取您的accessid和accesskey，请参见获取账号的AK信息。</p>
accesskey	<p>您在访问OSS中的文件时所持有的Access Key Secret。</p>
format	<p>数据文件的格式，创建Parquet格式文件的外表时必须将其设置为parquet。</p>

创建Parquet格式文件的外表时，需要注意数据类型的对应关系，具体规则如下：

Parquet基本类型	Parquet的logicalType类型	可对应ADB类型
BOOLEAN	无	boolean
INT32	INT_8	tinyint
INT32	INT_16	smallint
INT32	无	int或integer
INT64	无	bigint
FLOAT	无	float
DOUBLE	无	double
<ul style="list-style-type: none"> FIXED_LEN_BYTE_ARRAY BINARY INT64 INT32 	DECIMAL	decimal
BINARY	UTF-8	<ul style="list-style-type: none"> varchar string json（如果已知Parquet该列内容为json格式）

Parquet基本类型	Parquet的logicalType类型	可对应ADB类型
-------------	-----------------------	----------

INT32	DATE	date
INT64	TIMESTAMP_MILLIS	timestamp或datetime
INT96	无	timestamp或datetime

 注意

- 外表定义中column的名称应与Parquet文件的中该column的名称必须完全对应（可忽略大小写），而顺序可以是随意的，但建议也保持同样顺序。
- 外表定义中的column可以只选择Parquet文件的部分列，未被外表定义的Parquet文件的列将被忽略；反之如果定义了Parquet文件中未包含的列，该列的查询将均为NULL。

创建OSS TEXT格式外表

```
CREATE TABLE IF NOT EXISTS oss_text_external_table
(
  uid string,
  other string
)
engine='oss' TABLE_PROPERTIES='{
  "endpoint":"oss-cn-hangzhou-internal.aliyuncs.com",
  "accessid":"LTAIF****5FsE",
  "accesskey":"Ccw****iWjv",
  "format":"text",
  "row_delimiter":"\n",
  "field_delimiter":"\n",
  "URL":"oss://****"
}';
```

参数	说明
ENGINE='OSS'	表示该表是外部表，使用的存储引擎是OSS。
TABLE_PROPERTIES	用于告知AnalyticDB for MySQL如何访问OSS中的数据。

参数	说明
endpoint	<p>OSS的 <code>EndPoint</code>（域名节点）。</p> <p> 说明 目前仅支持AnalyticDB for MySQL通过OSS中ECS的VPC网络（内网）访问OSS。</p> <p>登录OSS控制台，单击目标Bucket，在Bucket概览页面查看 <code>endpoint</code>。</p>
url	<p>OSS中目标数据文件夹或文件的绝对地址。</p> <p>示例：<code>oss://\$Bucket名称/adb/oss_import_test_data.txt</code></p>
accessid	<p>您在访问OSS中的文件时所持有的AccessKey ID。</p> <p>如何获取您的accessid和accesskey，请参见获取账号的AK信息。</p>
accesskey	<p>您在访问OSS中的文件时所持有的Access Key Secret。</p>
format	<p>数据文件的格式，创建TEXT格式文件的外表时必须将其设置为text。</p>
row_delimiter	<p>定义TEXT文件的行分割符，目前仅支持一种：<code>\n</code></p>
field_delimiter	<p>定义TEXT文件的列分隔符，只能为一个字符。设置为 <code>\n</code> 表示整行为一个字段。</p>

针对带有分区的Parquet/CSV数据文件创建OSS外表

有的OSS数据源是包含分区的，会在OSS上形成一个分层目录，类似如下内容：

```
parquet_partition_classic/
├── p1=2020-01-01
│   ├── p2=4
│   │   ├── p3=SHANGHAI
│   │   │   ├── 000000_0
│   │   │   └── 000000_1
│   │   └── p3=SHENZHEN
│   │       └── 000000_0
│   └── p2=6
│       └── p3=SHENZHEN
│           └── 000000_0
├── p1=2020-01-02
│   └── p2=8
│       ├── p3=SHANGHAI
│       │   └── 000000_0
│       └── p3=SHENZHEN
│           └── 000000_0
└── p1=2020-01-03
    └── p2=6
        ├── p2=HANGZHOU
        └── p3=SHENZHEN
            └── 000000_0
```

上述数据中p1为第1级分区，p2为第2级分区，p3为第3级分区。对应这种数据源，一般都希望以分区的方式进行查询，那么就需要在创建OSS外表时额外指明分区列。具体的建表语法示例如下（本例为Parquet格式，分区也支持CSV格式）：

```
CREATE TABLE IF NOT EXISTS oss_parquet_partition_table
(
  uid varchar,
  other varchar,
  p1 date,
  p2 int,
  p3 varchar
)
ENGINE='OSS'
TABLE_PROPERTIES='{
  "endpoint":"oss-xxxx.aliyuncs.com",
  "url":"oss://****/****/oss_parquet_data_dir",
  "accessid":"****",
  "accesskey":"****",
  "format":"parquet",
  "partition_column":"p1, p2, p3"
}'
```

② 说明

- 如上例所示，除了在table的列定义中声明p1、p2、p3及其类型，还需要在 `TABLE_PROPERTIES` 部分中的`partition_column`属性里声明它们为分区列。且`partition_column`属性里必须按“第1级,第2级,第3级.....”的严格顺序声明（例中p1为第1级分区，p2为第2级分区，p3为第3级分区），在列定义中也需保持相同顺序，并将分区列置于列定义列表的末尾。
- 可以作为分区列的数据类型有：boolean、tinyint、smallint、int/integer、bigint、float、double、decimal、varchar/string、date、timestamp。
- 查询时分区列和其它数据列的表现和用法没有区别。