

ALIBABA CLOUD

阿里云

数据湖分析

RDS

文档版本：20211014

 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. 一键同步RDS表结构	05
2. 创建MySQL数据库	08
2.1. 背景信息以及准备工作	08
2.2. 创建MySQL Schema	09
2.3. 创建MySQL表并读取数据	09

1. 一键同步RDS表结构

云原生数据湖分析DLA支持通过标准JDBC对RDS（MySQL、SQLServer、PostgreSQL）中的数据进行查询和分析，并可以将分析结果回写入RDS。RDS一键同步元数据功能，只需要在DLA中创建RDS Schema，然后通过 `MSCK REPAIR TABLE` 命令一键将数据库中所有表的信息同步至DLA中。本文以MySQL为例，详细介绍RDS一键同步元数据功能。

注意事项

在DLA中创建MySQL、SQLServer、PostgreSQL Schema之前，需要将IP地址段 `100.104.0.0/16` 加入到RDS的白名单列表中。具体操作请参见[通过客户端、命令行连接RDS MySQL实例](#)。

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

 **说明** 当您在RDS白名单中添加了 `100.104.0.0/16` IP地址段，即视为您同意阿里云利用VPC反向访问技术读写RDS数据。

准备工作

通过DLA读写RDS数据前，您需要通过以下操作在RDS中准备好测试数据：

1. [创建RDS MySQL实例](#)
2. [通过客户端、命令行连接RDS MySQL实例](#)
3. [创建数据库和账号](#)
4. [通过客户端、命令行连接RDS MySQL](#)
5. 创建表

操作步骤

您可以在DMS页面编写SQL创建MySQL Schema，也可以先通过MySQL客户端或者MySQL命令行工具连接DLA，然后再编写SQL创建MySQL Schema。

1. 创建RDS Schema
 - i. 登录[DLA控制台](#)。
 - ii. 在左侧导航栏单击**Serverless Presto > SQL访问点**。
 - iii. 在**SQL访问点**页面的**VPC网络**区域，定位到RDS所在的VPC网络，然后单击**登录到DMS**。
您也可以**在公网网络区域**，单击**登录到DMS**通过公网登录到DMS创建RDS Schema。

- iv. 在登录实例对话框中，选择RDS数据库类型、实例地区、实例ID，输入数据库账号和密码，单击登录。

登录实例

* 数据库类型 MySQL

* 实例地区 华东1 (杭州)

* 实例ID 请选择

* 数据库账号 请输入数据库账户

* 数据库密码

记住密码

测试连接 登录 取消

说明 您需要将DMS IP地址添加至RDS实例的白名单设置中，才能正常登录到DMS。具体操作请参见[设置IP白名单](#)。

- v. 在DMS的SQL information_schema页签中，创建MySQL Schema的语法如下所示：

```
CREATE SCHEMA hello_mysql_vpc_rds WITH DBPROPERTIES (  
  CATALOG = 'mysql',  
  LOCATION = 'jdbc:mysql://rm-2zer0vg58mf****.mysql.rds.aliyuncs.com:3306/rds_mysql_dbname',  
  USER = 'rds_mysqlldb_username',  
  PASSWORD = 'rds_mysqlldb_password',  
  INSTANCE_ID = 'rds_mysql_instance_id',  
  VPC_ID = 'rds_mysqlldb_vpcid'  
);
```

2. 执行MSCK REPAIR TABLE命令同步表信息。

- i. 先执行show tables查询Schema中是否存在表。

```
mysql> use hello_mysql_vpc_rds;  
Database changed  
mysql> show tables;  
Empty set (0.01 sec)
```

- ii. 再执行MSCK REPAIR TABLE, 将RDS数据库中的所有表一键同步至Schema中。

```
mysql> msck repair database hello_mysql_vpc_rds;
```

```
+-----+-----+
| Table Name      | Operation |
+-----+-----+
| admin_acc_log   | CREATED  |
| fractest        | CREATED  |
| person          | CREATED  |
| prep_stmt_test  | CREATED  |
| staples         | CREATED  |
| test_datetime   | CREATED  |
| top10_user      | CREATED  |
| type_test       | CREATED  |
| world1          | CREATED  |
| world2          | CREATED  |
+-----+-----+
```

```
17 rows in set (1.60 sec)
```

```
mysql> show tables;
```

```
+-----+
| Table Name      |
+-----+
| admin_acc_log   |
| fractest        |
| person          |
| prep_stmt_test  |
| staples         |
| test_datetime   |
| top10_user      |
| type_test       |
| world1          |
| world2          |
+-----+
```

```
17 rows in set (0.02 sec)
```

表信息同步成功之后, 接下来您就可以通过DLA管理RDS数据了。

2. 创建MySQL数据库

2.1. 背景信息以及准备工作

Data Lake Analytics（简称DLA）作为云上数据处理的枢纽，支持通过标准JDBC连接[阿里云关系型数据库](#)（Relational Database Service，简称RDS）（MySQL、SQLServer、PostgreSQL）数据库、[云服务器](#)（Elastic Compute Service，简称ECS）上的自建数据库（MySQL、SQLServer、PostgreSQL），并对其数据进行查询操作。

前提条件

通过DLA读写ECS自建MySQL数据前，必须先连接ECS，详情请参见[连接方式概述ECS远程连接操作指南](#)，然后通过以下操作创建MySQL数据库、写入测试数据。

 **注意** DLA将通过VPC连接ECS自建MySQL数据库，应确保DLA与ECS所属Region相同。

1. 创建MySQL数据库dla_db。

```
MySQL [(none)]> CREATE DATABASE dla_db;
```

2. 执行以下SQL，新增dla_user用户，并将适当的权限授权给dla_user。后续DLA将以该用户读取MySQL数据。

```
MySQL [(none)]> USE mysql;
MySQL [mysql]> CREATE USER 'dla_user'@'%' IDENTIFIED BY 'dla_userpasswd';
MySQL [mysql]> GRANT SELECT,INSERT,UPDATE,DELETE,CREATE,DROP,ALTER,INDEX on dla_db.* TO 'dla_user'@'%';
MySQL [mysql]> flush privileges;
```

 **注意** 若DLA与ECS所属账号相同（同一主账号或者子账号），则上述授权 `GRANT SELECT.INSERT.UPDATE.DELETE.CREATE.DROP.ALTER.INDEX on dla_db.* TO 'dla_user'@'%';` 可改为 `GRANT SELECT.INSERT.UPDATE,DELETE,CREATE,DROP,ALTER,INDEX on dla_db.* TO 'dla_user'@'100.104.0.0/255.255.0.0';` 授权方式，为了更加细粒度地控制权限，仅允许DLA可连接MySQL数据库。

3. 执行以下SQL，在MySQL数据库中创建person表，并写入测试数据。

```
CREATE TABLE person (
  id int,
  name varchar(1023),
  age int
);
```

```
INSERT INTO person VALUES
(1, 'james', 10),
(2, 'bond', 20),
```

说明

- 为了让DLA可以访问ECS自建MySQL数据库，需要在ECS的安全组中[添加安全组规则](#)：授权100.104.0.0/16 IP地址段。
- 添加安全组规则时，协议类型选择MySQL，授权对象设置为100.104.0.0/16。

2.2. 创建MySQL Schema

本文介绍如何在DLA中创建MySQL Schema。

1. 登录[DLA控制台](#)。
2. 在左侧导航栏，单击Serverless Presto > SQL访问点。
3. 在SQL访问点页面，单击登录到DMS，在DMS中创建Schema。创建MySQL Schema的语法如下所示：

```
CREATE SCHEMA hello_mysql_vpc_rds WITH DBPROPERTIES (
  CATALOG = 'mysql',
  LOCATION = 'jdbc:mysql://192.168.2.141:3306/dla_db',
  USER = 'dla_user',
  PASSWORD = 'dla_userpasswd',
  INSTANCE_ID = 'ecs_instance_id',
  VPC_ID = 'vpc_id'
);
```

参数	说明
CATALOG	指定创建的是MySQL Schema。
LOCATION	192.168.2.141是ECS的私网连接地址。
USER	MySQL数据库用户的用户名。
PASSWORD	MySQL数据库用户对应的用户名密码。
INSTANCE_ID	ECS实例ID
VPC_ID	ECS实例所属VPC ID或者DLA所属VPC ID。

2.3. 创建MySQL表并读取数据

本文介绍如何在DLA中创建MySQL表并读取数据。

1. 登录[DLA控制台](#)。
2. 单击左侧导航栏的SQL访问点，然后单击登录DMS，在DLA中创建如下表所示的MySQL表。

```
CREATE EXTERNAL TABLE hello_mysql_vpc_rds.tbl_person (  
  id int,  
  name varchar(1023),  
  age int  
) tblproperties (  
  table_mapping = "person"  
);
```

查询MySQL数据

在DLA中创建好了MySQL Schema和表后，您可以通过MySQL客户端或者MySQL命令行工具连接DLA，然后使用标准SQL语句读写MySQL数据库数据。

```
mysql> select * from person;  
+-----+-----+-----+  
| id | name | age |  
+-----+-----+-----+  
| 1 | james | 10 |  
| 2 | bond | 20 |
```