

ALIBABA CLOUD

阿里云

数据湖分析
连接数据湖分析

文档版本：20201207

 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.连接概览	05
2.MySQL命令行	06
3.Superset连接数据湖分析	07

1.连接概览

在应用开发中，您可以通过Data Lake Analytics管理控制台访问DLA，也可以通过代码或者客户端连接DLA。

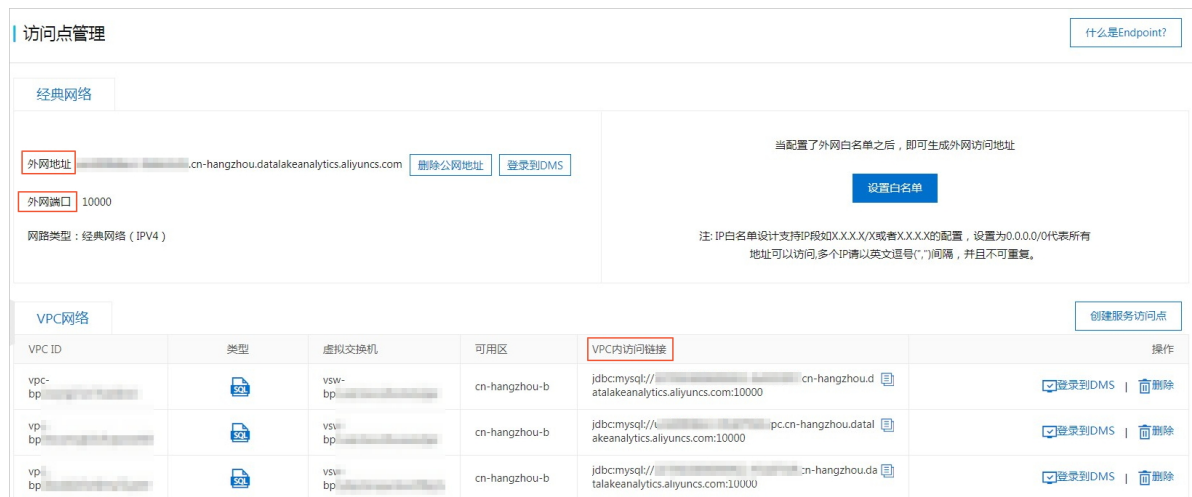
连接说明

连接DLA时，需要获取以下连接信息。

● 连接地址和端口

登录Data Lake Analytics管理控制台，单击左侧导航栏的系统管理>访问点管理，查看连接地址和端口号。

- 通过外网连接DLA时，使用外网连接地址。
- 通过内网连接DLA时，使用VPC连接地址。
- DLA中VPC专有网络是针对VPC的服务访问点（single tunnel），一个VPC服务访问点仅提供指定的VPC到DLA的服务访问通道。
- 如果要从多个不同的VPC访问DLA，需要针对每个VPC分别创建VPC服务访问点。每个VPC服务访问点的连接串信息不同。



- 用户名和密码，您可以通过数据库主账号（Root账号）或者数据库子账号连接DLA。

通过代码连接DLA

- PreparedStatement
- Druid
- Druid连接池
- JDBC

通过客户端连接DLA

- DBeaver
- Navicat for MySQL
- SQL WorkBench/J

通过MySQL命令行工具连接DLA

MySQL命令行

2.MySQL命令行

DLA支持MySQL兼容的JDBC，可以使用MySQL命令行工具通过JDBC连接DLA。

前提条件

- DLA对外提供的是MySQL访问协议，MySQL命令行工具协议支持MySQL 5.0及以上版本。


```
mysql --version
mysql Ver 14.14 Distrib 5.6.38, for Linux (x86_64) using EditLine wrapper
```

- 在创建JDBC连接前，需要先[设置服务访问点](#)来生成连接串，连接串由连接地址和端口号组成。

代码类型

不指定Schema

```
mysql -h<end_point_host> -P<end_point_port> -u<user_name> -p<password> -c -A
```

 **说明** MySQL client 默认发送到服务器的 SQL statements 中删除 comments（包括优化器提示），直到 MySQL 5.7.7。如果使用小于5.7.7版本，请使用 -c 选项支持添加 comments（包括优化器提示）。

参数

- end_point_host：连接地址，如何获取连接地址，请参见[设置服务访问点](#)中的[查看服务访问点信息](#)。
- end_point_port：端口号。
- user_name：用户名。
- password：密码。

连接DLA后，您可以通过use <schema>切换到对应的Schema下。

指定schema

```
mysql -h<end_point_host> -P<end_point_port> -u<user_name> -p<password> <schema> -c -A
```

连接DLA后，您已经在指定Schema下，可以通过select database()查看当前Schema。

3.Superset连接数据湖分析

本文主要介绍如何使用Superset连接DLA进行分析。

背景信息

DLA对外暴露的是MySQL的协议，大多数BI工具可以使用MySQL连接器直接连接DLA，由于Superset对于表的定义语句校验严格，同时DLA的建表语句与MySQL的建表语句存在差别，这导致Superset不能够直接连接DLA，但是我们支持一种兼容模式，可以通过修改用户名的方法来连接DLA。

操作方法

在用户名后加上字段`%lst=true`。

例如：原用户名是 `user`，密码是 `pass`，把用户名修改为 `user%lst=true` 后即可使用Superset连接DLA进行分析。