ALIBABA CLOUD

阿里云

数据湖分析

ElasticSearch

文档版本：20210708

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 `cd /d C:/window` 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	`bae log list --instanceid` *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.读写ElasticSearch数据 ---------- 05

1.读写ElasticSearch数据

云原生数据湖分析DLA（Data Lata Analytics）中支持接入Elasticsearch，通过标准SQL语句读取Elasticsearch中的数据，从而实现Elasticsearch中的数据与其他数据源的数据联合查询等业务需求。本文档主要介绍了DLA读写Elasticsearch数据的操作步骤。

前提条件

使用DLA读写Elasticsearch数据前，您需要通过以下操作在Elasticsearch中准备好测试数据。

1. 创建阿里云Elasticsearch实例

   > **说明**
   > - DLA和Elasticsearch所属地域必须相同，否则无法进行本文档操作。
   > - DLA只支持6.3及以上版本的Elasticsearch。

2. 配置Elasticsearch公网或私网访问白名单
3. 访问实例
4. 创建索引
5. 创建文档并插入数据

背景信息

Elasticsearch是一个基于Lucene的实时分布式的搜索与分析引擎，是遵从Apache开源条款的一款开源产品，是当前主流的企业级搜索引擎。它提供了一个分布式服务，可以使您快速的近乎于准实时的存储、查询和分析超大数据集，通常被用来作为构建复杂查询特性和需求强大应用的基础引擎或技术。

操作步骤

1. 创建Elasticsearch 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，单击**登录到DMS**，在DMS中执行以下SQL创建Schema。您也可以通过MySQL客户端或者程序代码等方式链接DLA，然后执行以下SQL创建Schema。

```
CREATE DATABASE `my_es_db`
WITH DBPROPERTIES (
catalog = 'elasticsearch',
location = 'http://****.elasticsearch.aliyuncs.com:9200',
user = '******',
password = '******',
VPC_ID = 'vpc-*****',
INSTANCE_ID = 'es-*****'
)
```

参数名称	参数说明
catalog	取值为elasticsearch，表示创建的是Elasticsearch Schema。
location	Elasticsearch的连接地址，由地址、协议、和端口三部分组成。您可以登录阿里云Elasticsearch控制台，在Elasticsearch实例的**基本信息**页面获取Elasticsearch实例的连接地址。 **注意** 您在输入LOCATION参数信息时，末尾不能添加符号/。
user	Elasticsearch的用户名。
password	Elasticsearch的密码。
VPC_ID	Elasticsearch实例所在的VPC ID。
INSTANCE_ID	Elasticsearch的实例ID。

2. 创建Elasticsearch表。例如通过以下SQL在 `my_es_db` Schema中创建 `product_info` 表。

```
CREATE EXTERNAL TABLE `product_info` (
`productname` string NULL COMMENT '',
`annual_rate` string NULL COMMENT '',
`describe` string NULL COMMENT ''
)
TBLPROPERTIES (
TABLE_MAPPING = 'product_info',
COLUMN_MAPPING = 'productname,productName;'
);
```

说明

- 默认情况下表名和ElasticSearch中对应的索引名保持一致。如果不一致可以通过 `TABLE_MAPPING` 属性指定要映射的索引名称。
- 由于 `productName` 中包含大写字母，而DLA中列名称不区分大小写，所以需要在 `COLUMN_MAPPING` 中增加映射。

3. 读取Elasticsearch中的表数据。例如通过 `SELECT` 在DLA中读取 `product_info` 表中的数据。

```
select * from product_info;
```