云数据仓库套件 Sparkling

快速入门

产品文档

腾讯云

文档目录

快速入门
创建集群

最近更新时间：2020-03-16 18:25:10

操作场景

Sparkling 集群是云数据仓库套件 Sparkling 为用户提供服务的载体。Sparkling 集群的大小决定了 Sparkling 数据仓库所能提供的存储能力和计算能力的上限，您可以根据业务需要定制 Sparkling 集群。

本文将为您介绍如何通过 Sparkling 控制台快速创建一个 Sparkling 集群。

前提条件

1. 已完成 腾讯云 账号注册，并完成 实名认证。
2. 在 云数据仓库套件 Sparkling 页面填写申请单，填写完成后单击【提交申请】完成线上申请。

> 说明：
> 腾讯云平台接到服务申请单后会进行服务需求审核，数据仓库团队工作人员将联系您进行初步需求确认、应用场景沟通与商务洽谈，审核通过后我们将为您开通内测资格。

创建集群

1. 进入 集群管理 页面，在左侧导航单击【集群】进入集群管理页面。
2. 在页面左上角选择集群所在区域，目前支持【广州】、【上海】、【北京】。
3. 在页面右上角单击【创建集群】进入集群创建页面。

4. 填写集群配置参数。

名称	含义	备注
存储集群名称	集群的名称	-
区域	集群实际工作区域	当前版本仅支持广州、上海和北京区域。 当前区域默认为集群管理页面所选择区域，如需更改请返回集群管理页面左上角重新选择区域。
可用区	选择区域下关联的可用分区	可以在【Master节点】查看该可用区是否有可用机型。 • 可用：显示请选择机型。 • 不可用：显示该可用区无可用机型。
选择网络	指定连接到 Sparkling 的私有网络	可以在控制台新建私有网络进行查询和规划。**选定 VPC 和子网后，不能更换。后续部署的计算集群只能部署在相同的 VPC 和子网下。**
选择子网组	指定连接到 Sparkling 的子网	可以在控制台新建子网进行查询和规划。
运行版本	Sparkling 内部组件版本	当前版本仅支持 0.1.0(Spark 2.3.2, Hadoop 2.7.3, Hive 2.1.0)。
Master	D1	当前版本可选择四种配置的大数据 D1 机型，根据需要选择主节点的内存及

节点		CPU 核心数。

5. 选择集群节点类型及数量。

名称	含义	备注
核心计算节点	负责集群的存储任务	当前版本支持四种配置的大数据型 D1 机型，可根据需要选择节点内存及 CPU 核心数。
弹性计算节点	负责集群的计算任务	当前版本支持多种配置的内存型机型，可根据需要选择节点内存及 CPU 核心数。
最少节点数量	所需的最少节点数	-
最多节点数量	所需的最多节点数	-

6. 单击【确定】创建集群。
   集群创建需要时间，初始状态为【创建中】，请等待一段时间。创建成功后状态更新为【正常】。
7. 在左侧导航单击【集群】进入集群管理页面。在左上角【广州】、【上海】、【北京】选择集群所在区域后即可查看当前区域下集群列表信息。

导入数据

最近更新时间：2020-03-16 18:25:47

操作场景

本文将为您介绍如何将云数据库 TencentDB for MySQL 中的数据全量导入到新建的 Sparkling 集群上。Sparkling 支持多样化的数据导入方式及定时导入功能，更多方式请参见 数据集成。

前提条件

已在腾讯云数据库 MySQL 控制台 中创建 MySQL 数据库实例，且您要导入的数据在该实例中已保存，建议数据库与 Sparkling 集群所在地域一致以保证访问的稳定性和速率。

关于 MySQL 数据库操作，请参见 云数据库 MySQL 入门。

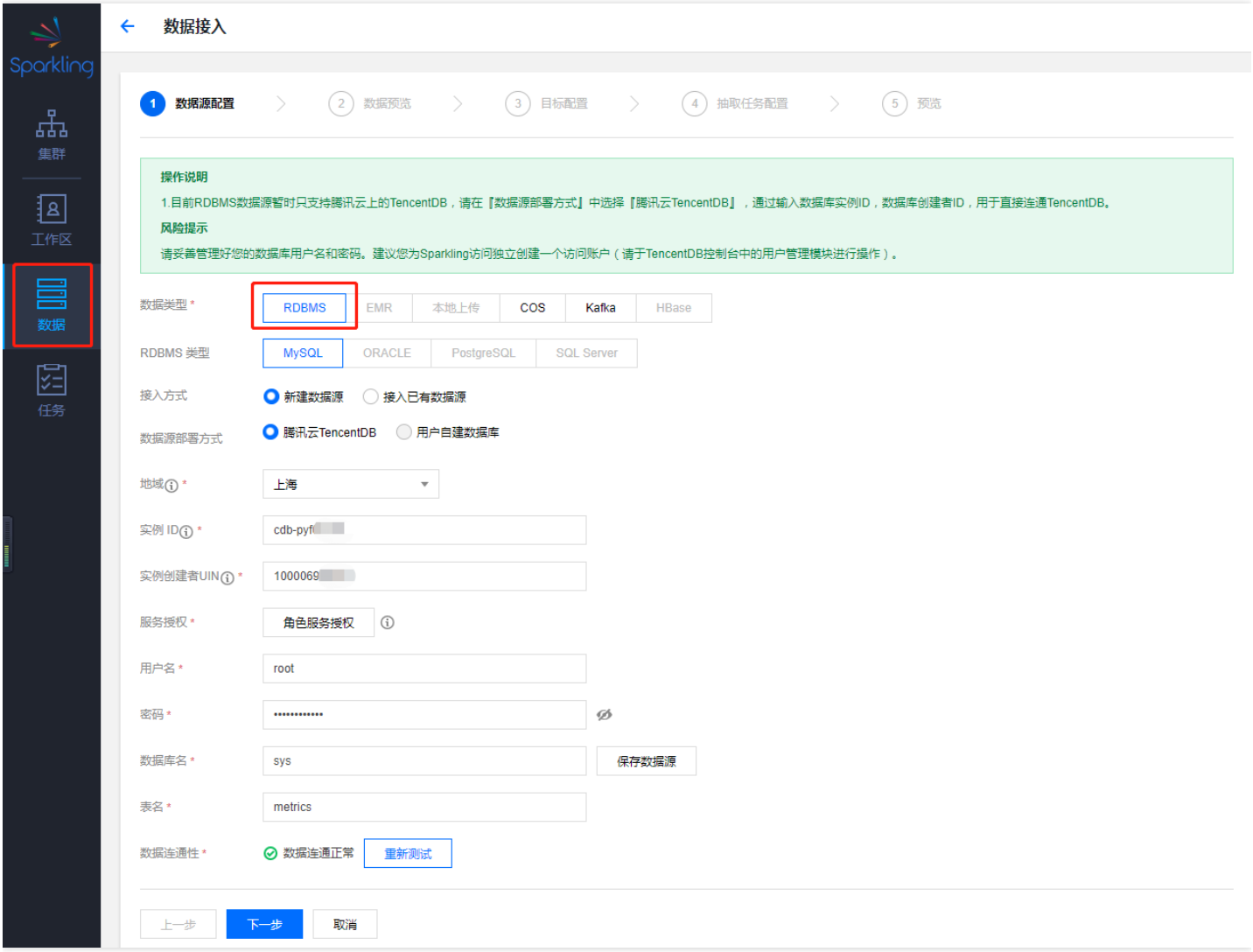
操作步骤

登录 Sparkling 控制台，在左侧导航单击【数据】进入数据接入页面，执行以下操作步骤完成 RDBMS 数据接入：

1. 数据源配置

以导入 MySQL 数据库中默认数据库 sys 中的默认数据表 metrics 为例，如图所示配置数据源，其中地域选择 MySQL 所在地域，MySQL 实例 ID 可在 MySQL 控制台 获取，实例创建者 UIN 在创建者 账号信息 中的账号 ID 栏

获取。

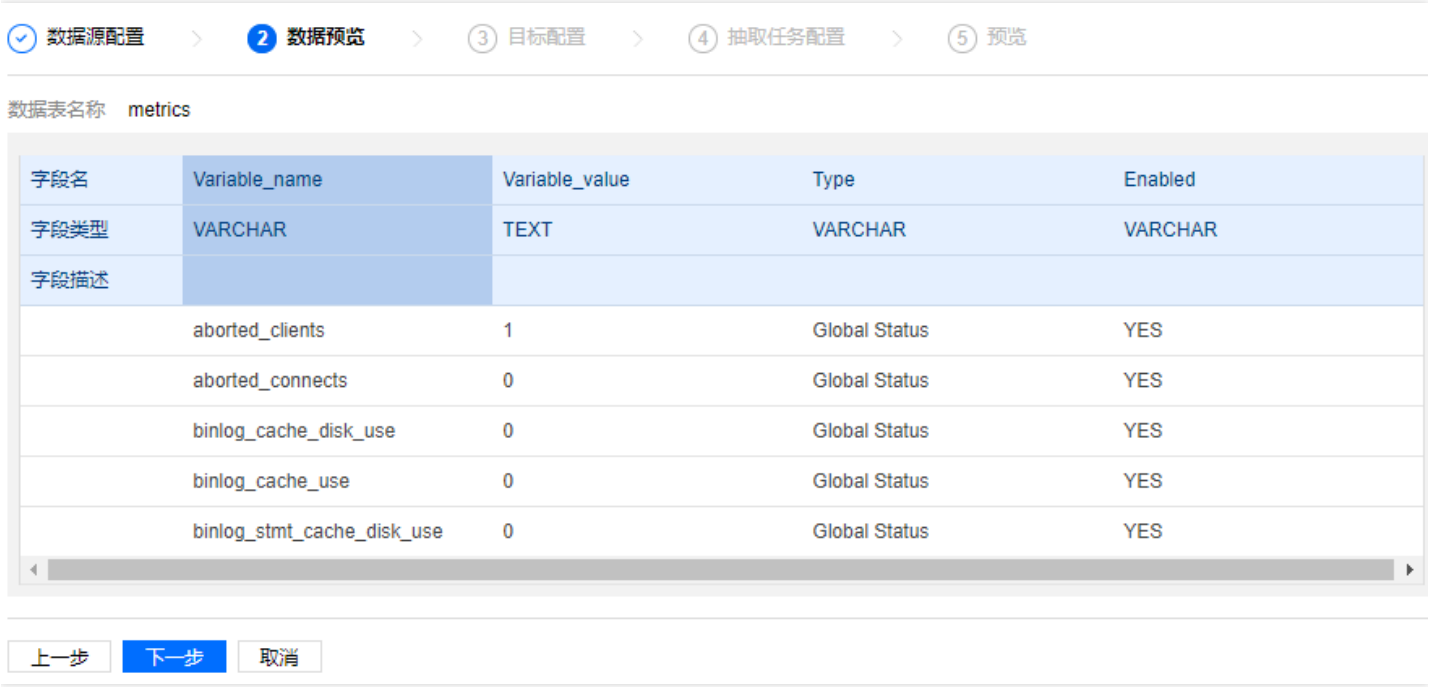


说明：

单击【测试连通性】确认是否可以连接到要接入数据表所在的数据库，待显示【数据连通正常】后，单击【下一步】完成数据源配置操作。

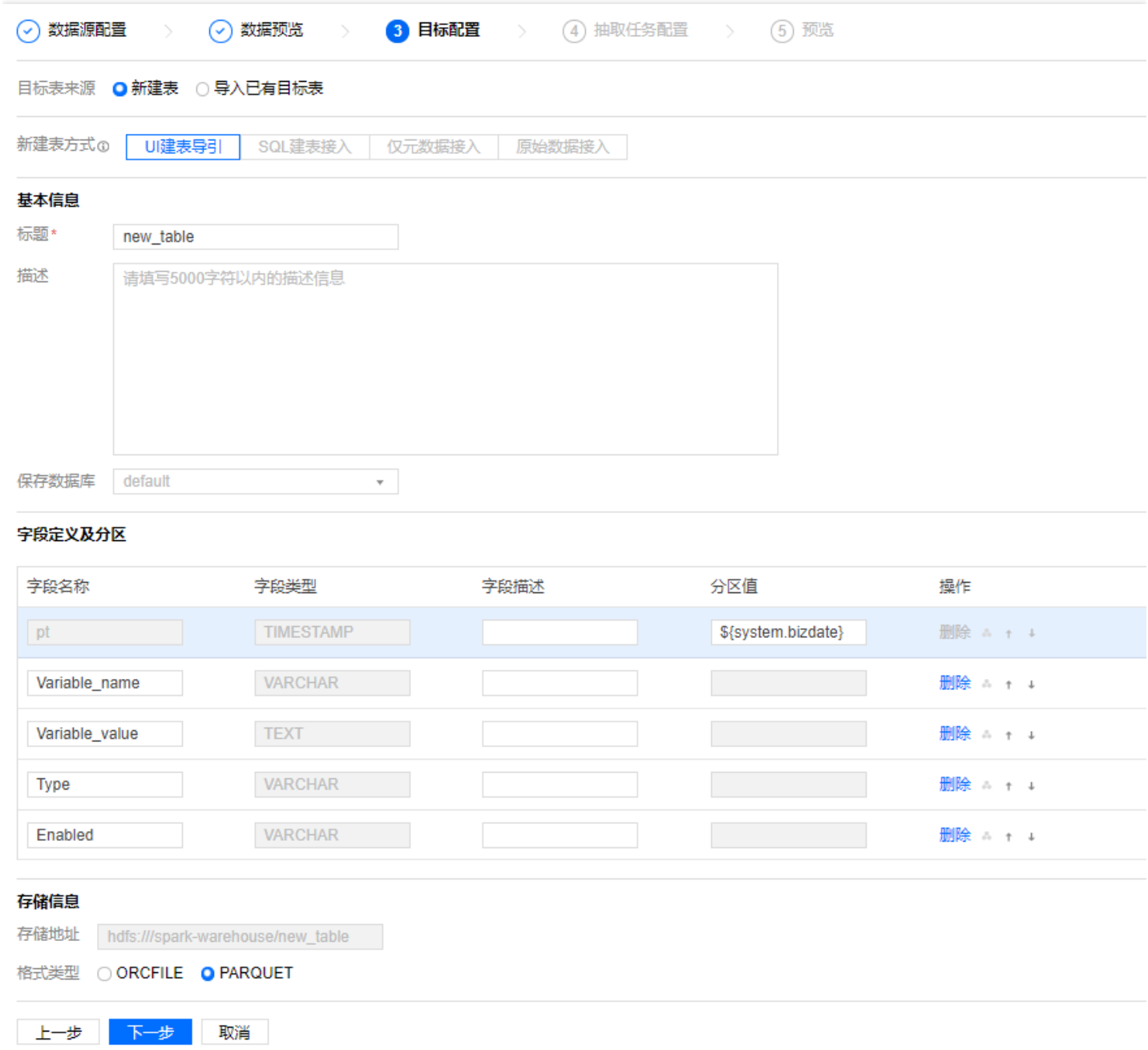
2. 数据预览

在【数据预览】页可以预览数据表中的字段信息，默认抽取五行数据进行预览，预览无误后单击【下一步】。



3. 目标配置

如图所示设置后，可以建立表名为“new_table”的新建表，字段设置为原 metrics 表中的所有字段，存储格式为 PARQUET。


数据源配置
数据预览
3 目标配置
4 抽取任务配置
5 预览
目标表来源 新建表 导入已有目标表
新建表方式 UI建表导引 SQL建表接入 仅元数据接入 原始数据接入
基本信息
标题* new_table
描述 请填写5000字符以内的描述信息
保存数据库 default
字段定义及分区
字段名称 字段类型 字段描述 分区值 操作
pt TIMESTAMP ${system.bizdate} 删除
Variable_name VARCHAR 删除
Variable_value TEXT 删除
Type VARCHAR 删除
Enabled VARCHAR 删除
存储信息
存储地址 hdfs:///spark-warehouse/new_table
格式类型 ORCFILE PARQUET
上一步 下一步 取消

4. 抽取任务配置

本文示例选择单次全量导入的方式导入数据，更多导入方式请参见 数据集成 。

5. 预览

在【预览】页查看当前设置的数据源信息、数据预览信息、目标表信息及任务调度信息，确认无误后单击【完成】即可完成数据导入任务设置。在弹出的对话框中单击【确定】，跳转到任务管理页面。

数据源配置 > 数据预览 > 目标配置 > 抽取任务配置 > 5 预览

数据类型 RDBMS
RDBMS类型 mysql
接入方式 新建数据源
部署方式 TencentDB
实例ID cdb-r
地区 广州
用户名 root
数据库名 sys
表名 metrics
数据连通性 数据连通正常

字段名称	pt	Variable_name	Variable_value	Type	Enabled
字段类型	TIMESTAMP	VARCHAR	TEXT	VARCHAR	VARCHAR
字段描述					
	${system.bizdate}	aborted_clients	1	Global Status	YES
	${system.bizdate}	aborted_connects	0	Global Status	YES
	${system.bizdate}	binlog_cache_disk_use	0	Global Status	YES
	${system.bizdate}	binlog_cache_use	0	Global Status	YES
	${system.bizdate}	binlog_stmt_cache_di...	0	Global Status	YES

目标表来源 新建表
新建表方式 UI建表导引
标题 new_table
保存数据库 default

数据加载方式 写入前清理已有数据 (Insert Overwrite)

上一步 完成 取消

简单的 SQL 数据分析

最近更新时间：2020-03-16 18:26:08

操作场景

本文为您介绍如何使用 Sparkling 笔记簿功能实现简单的 SQL 数据查询及数据可视化分析。更多数据开发细节，请参见 数据开发。

前提条件

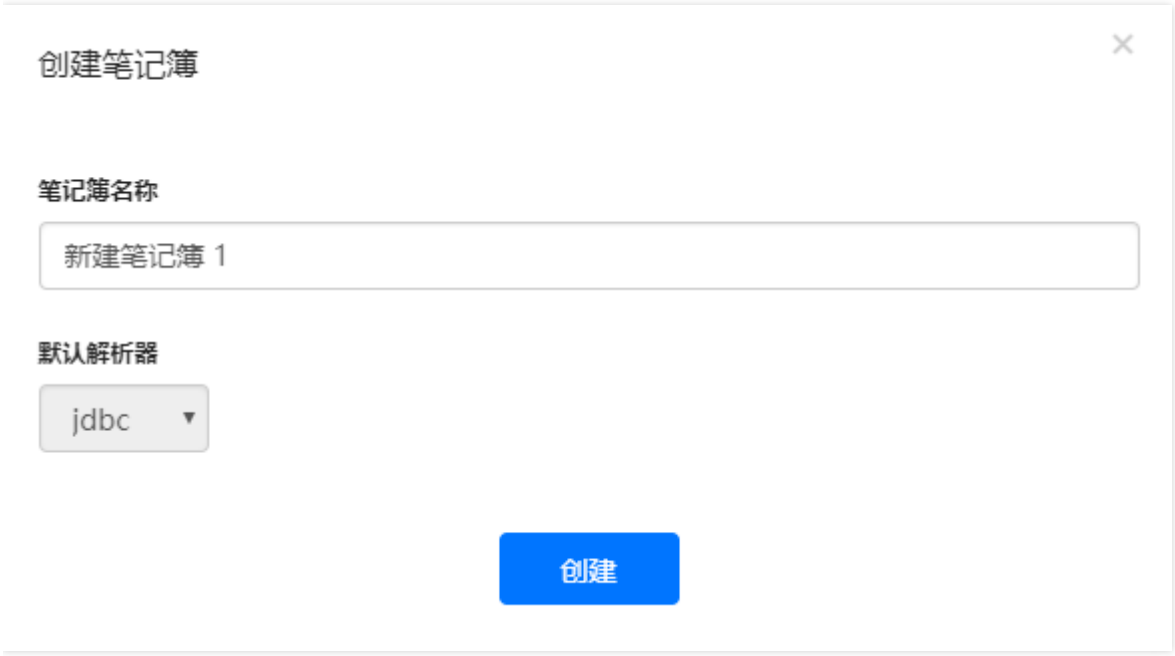
在进行数据分析之前，请确保您已根据 创建集群 指引建立 Sparkling 集群，并已根据 导入数据 指引将数据导入集群中。

操作步骤

进入 集群管理 页面，在左侧导航单击【工作区】进入数据开发页面。

1. 新建笔记簿

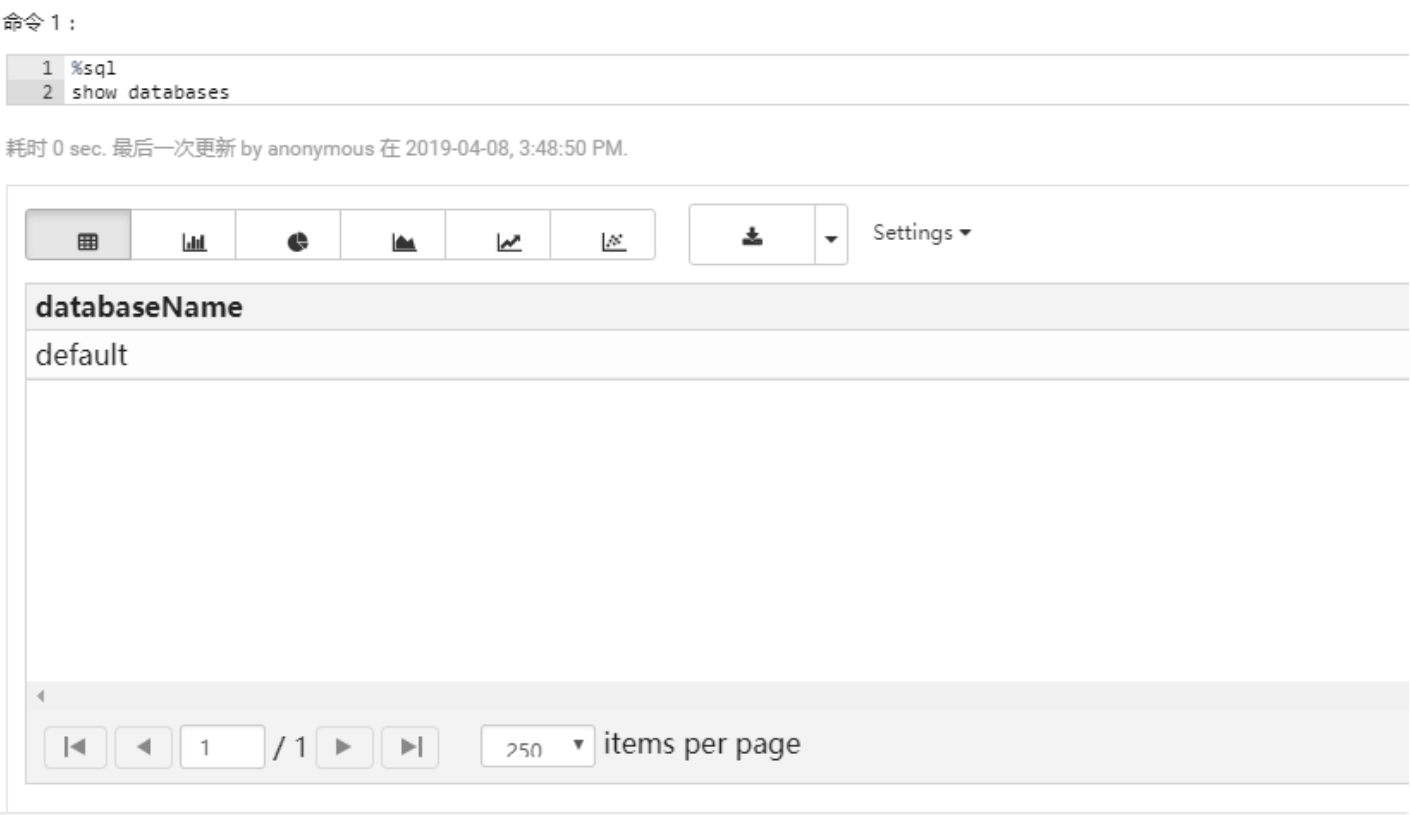
单击工作区左上角【+】，选择【新建笔记簿】，建立新笔记簿。



2. 查找数据库及数据表

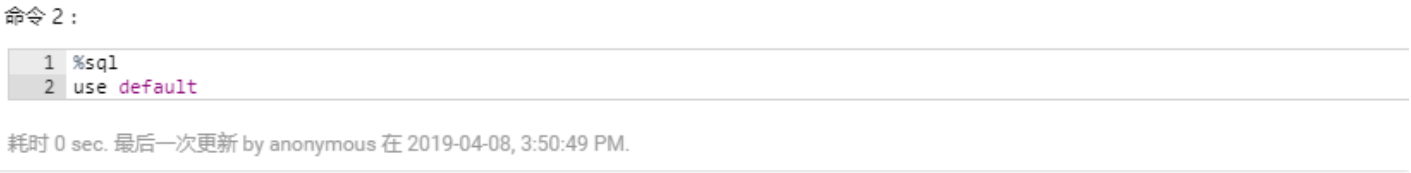
a. 在命令行输入以下命令后，使用快捷键 Shift + Enter 或单击右上角运行按钮运行该命令行，查找当前集群下包含的数据库名。

```
show databases
```



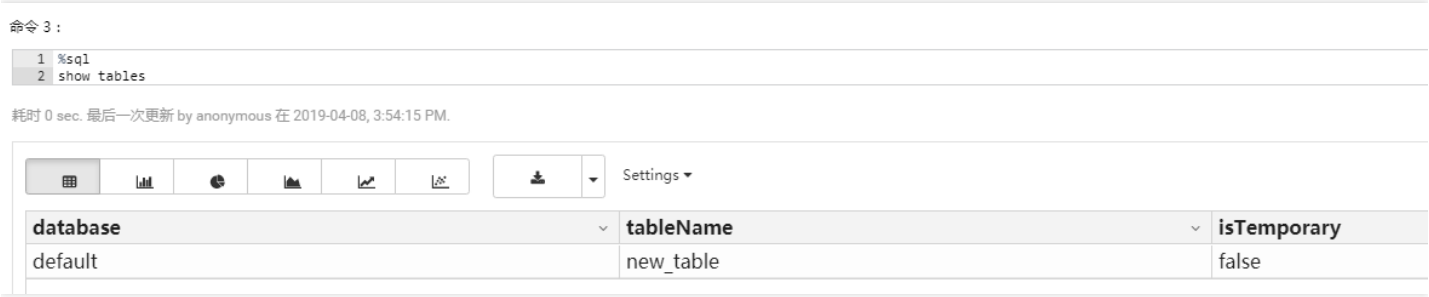
b. 输入以下命令进入 default 数据库。

```
use default
```



c. 输入以下命令查找 default 数据库中包含的数据表名，可以看到之前导入的数据表 new_table 已经存在于 Sparkling 集群中。

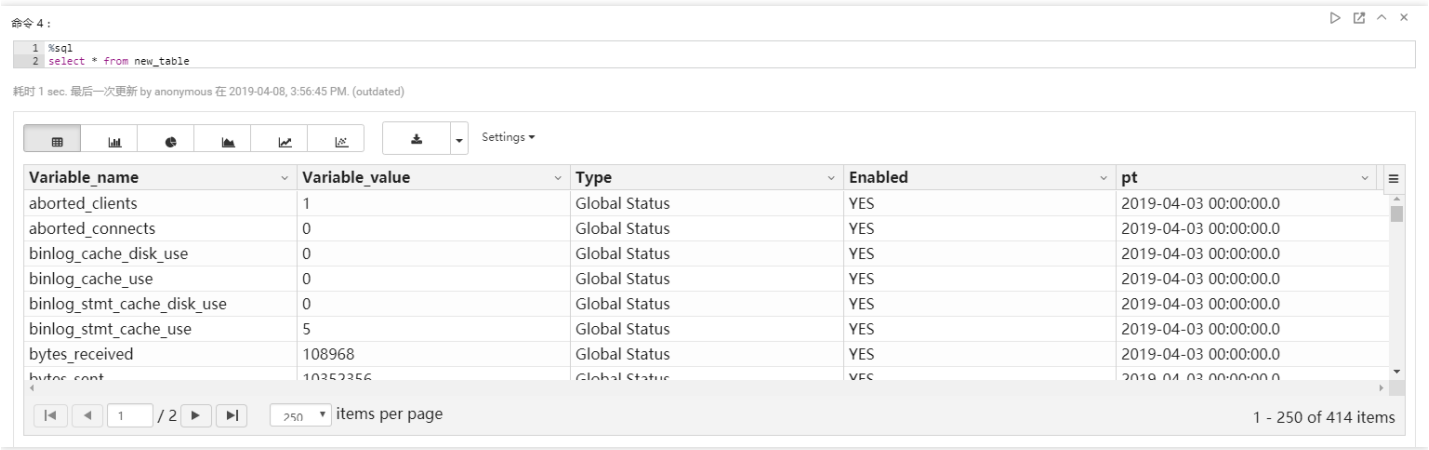
```
show tables
```



3. 执行简单的 SQL 语句

执行以下命令查看 new_table 中的数据信息，其中 pt 列是 Sparkling 集群导入时增加的一列时间戳，默认定义为数据导入日期的前一天00:00。

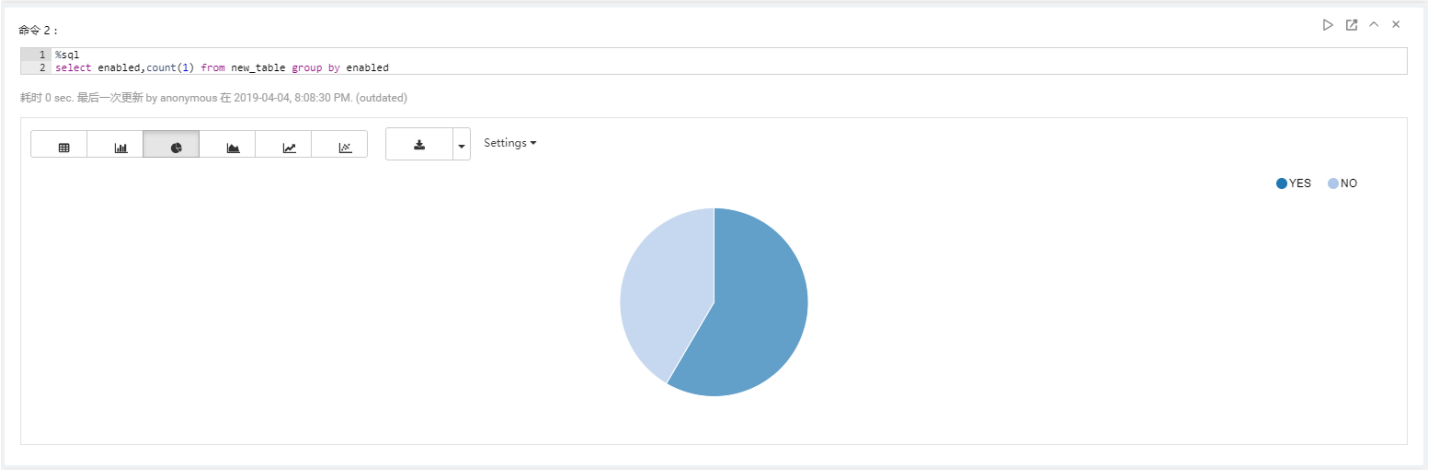
```
select * from new_table
```



4. 数据可视化分析

执行以下命令获取以 enabled 分组的检索行数，将结果绘制饼图如下：

```
select enabled,count(1) from new_table group by enabled
```



执行以下命令获取以 type 分组的检索行数，绘制柱状图如下，其中单击【Settings】可以设置 keys、groups、values 值。

```
select type,count(1) from new_table group by type
```