

尚硅谷大数据项目之尚品汇（可视化报表）

(作者：尚硅谷研究院)

版本：V4.0

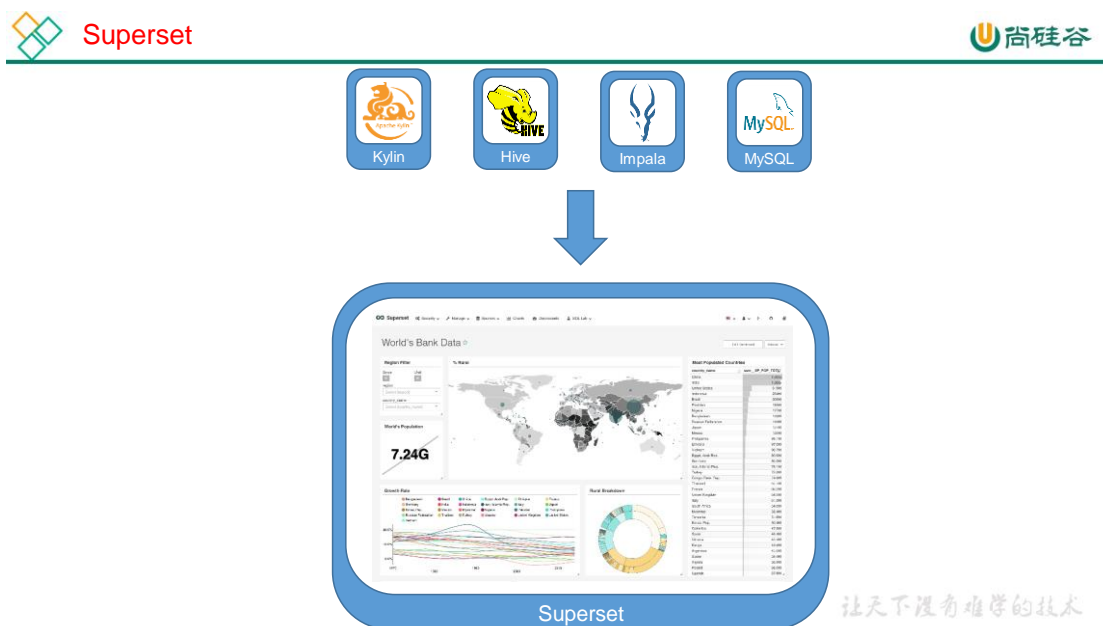
第 1 章 Superset 入门

1.1 Superset 概述

Apache Superset 是一个开源的、现代的、轻量级 BI 分析工具，能够对接多种数据源、拥有丰富的图表展示形式、支持自定义仪表盘，且拥有友好的用户界面，十分易用。

1.2 Superset 应用场景

由于 Superset 能够对接常用的大数据分析工具，如 Hive、Kylin、Druid 等，且支持自定义仪表盘，故可作为数仓的可视化工具。



第 2 章 Superset 安装及使用

Superset 官网地址：<http://superset.apache.org/>

2.1 安装 Python 环境

Superset 是由 Python 语言编写的 Web 应用，要求 Python3.7 的环境。

2.1.1 安装 Miniconda

conda 是一个开源的包、环境管理器，可以用于在同一个机器上安装不同 Python 版本的软件包及其依赖，并能够在不同的 Python 环境之间切换，Anaconda 包括 Conda、Python 以及一大堆安装好的工具包，比如：numpy、pandas 等，Miniconda 包括 Conda、Python。

此处，我们不需要如此多的工具包，故选择 MiniConda。

1) 下载 Miniconda (Python3 版本)

下载地址：https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh

2) 安装 Miniconda

(1) 执行以下命令进行安装，并按照提示操作，直到安装完成。

```
[atguigu@hadoop102 lib]$ bash Miniconda3-latest-Linux-x86_64.sh
```

(2) 在安装过程中，出现以下提示时，可以指定安装路径

```
Miniconda3 will now be installed into this location:
/home/atguigu/miniconda3
```

- Press ENTER to confirm the location
- Press CTRL-C to abort the installation
- Or specify a different location below

```
[/home/atguigu/miniconda3] >>> /opt/module/miniconda3
```

(3) 出现以下字样，即为安装完成

```
Thank you for installing Miniconda3!
```

3) 加载环境变量配置文件，使之生效

```
[atguigu@hadoop102 lib]$ source ~/.bashrc
```

4) 取消激活 base 环境

Miniconda 安装完成后，每次打开终端都会激活其默认的 base 环境，我们可通过以下命令，禁止激活默认 base 环境。

```
[atguigu@hadoop102 lib]$ conda config --set auto_activate_base
false
```

2.1.2 创建 Python3.7 环境

1) 配置 conda 国内镜像

```
(base) [atguigu@hadoop102 ~]$ conda config --add channels
https://mirrors.tuna.tsinghua.edu.cn/anaconda/pkg/free
(base) [atguigu@hadoop102 ~]$ conda config --add channels
https://mirrors.tuna.tsinghua.edu.cn/anaconda/pkg/main
(base) [atguigu@hadoop102 ~]$ conda config --set
show_channel_urls yes
```

2) 创建 Python3.7 环境

```
(base) [atguigu@hadoop102 ~]$ conda create --name superset
python=3.7
```

说明: conda 环境管理常用命令

创建环境: conda create -n env_name

查看所有环境: conda info --envs

删除一个环境: conda remove -n env_name --all

3) 激活 superset 环境

```
(base) [atguigu@hadoop102 ~]$ conda activate superset
```

激活后效果如下图所示

```
(superset) [atguigu@hadoop102 ~]$
```

说明: 退出当前环境

```
(superset) [atguigu@hadoop102 ~]$ conda deactivate
```

4) 执行 python 命令查看 python 版本

```
(superset) [atguigu@hadoop102 ~]$ python
Python 3.6.10 |Anaconda, Inc.| (default, Jan 7 2020, 21:14:29)
[GCC 7.3.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> quit();
```

2.2 Superset 部署

2.2.1 安装依赖

安装 Superset 之前, 需安装以下所需依赖

```
(superset) [atguigu@hadoop102 ~]$ sudo yum install -y gcc gcc-
c++ libffi-devel python-devel python-pip python-wheel python-
setuptools openssl-devel cyrus-sasl-devel openldap-devel
```

2.2.2 安装 Superset

1) 安装 (更新) setuptools 和 pip

```
(superset) [atguigu@hadoop102 ~]$ pip install --upgrade
setuptools pip -i https://pypi.douban.com/simple/
```

说明: pip 是 python 的包管理工具, 可以和 centos 中的 yum 类比

2) 安装 Superset

```
(superset) [atguigu@hadoop102 ~]$ pip install apache-superset -
i https://pypi.douban.com/simple/
```

说明: -i 的作用是指定镜像, 这里选择国内镜像

注: 如果遇到网络错误导致不能下载, 可尝试更换镜像

```
(superset) [atguigu@hadoop102 ~]$ pip install apache-superset
--trusted-host https://repo.huaweicloud.com -i
https://repo.huaweicloud.com/repository/pypi/simple
```

3) 初始化 Superset 数据库

```
(superset) [atguigu@hadoop102 ~]$ superset db upgrade
```

4) 创建管理员用户

```
(superset) [atguigu@hadoop102 ~]$ export FLASK_APP=superset
(superset) [atguigu@hadoop102 ~]$ superset fab create-admin
```

说明：flask 是一个 python web 框架，Superset 使用的就是 flask

5) Superset 初始化

```
(superset) [atguigu@hadoop102 ~]$ superset init
```

2.2.3 启动 Superset

1) 安装 gunicorn

```
(superset) [atguigu@hadoop102 ~]$ pip install gunicorn -i
https://pypi.douban.com/simple/
```

说明：gunicorn 是一个 Python Web Server，可以和 java 中的 TomCat 类比

2) 启动 Superset

(1) 确保当前 conda 环境为 superset，及下图所示

```
(superset) [atguigu@hadoop102 ~]$
```

(2) 启动

```
(superset) [atguigu@hadoop102 ~]$ gunicorn --workers 5 --timeout
120 --bind hadoop102:8787 "superset.app:create_app()" --daemon
```

说明：

--workers: 指定进程个数

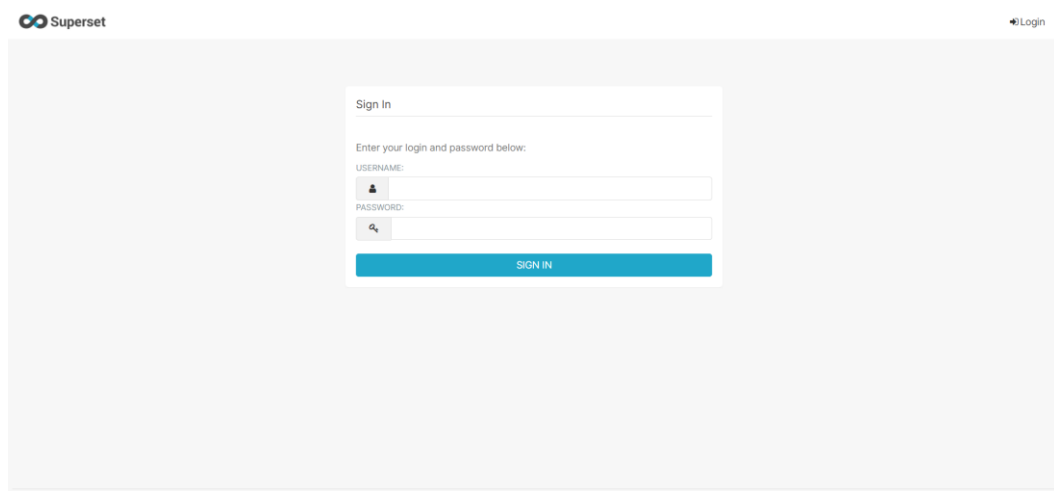
--timeout: worker 进程超时时间，超时会自动重启

--bind: 绑定本机地址，即为 Superset 访问地址

--daemon: 后台运行

(3) 登录 Superset

访问 <http://hadoop102:8787>，并使用 2.2.2 节中第 4 步创建的管理员账号进行登录。



3) 停止 superset

停掉 gunicorn 进程

```
(superset) [atguigu@hadoop102 ~]$ ps -ef | awk '/superset/ && !/awk/{print $2}' | xargs kill -9
```

退出 superset 环境

```
(superset) [atguigu@hadoop102 ~]$ conda deactivate
```

2.2.4 superset 启停脚本

1) 创建 superset.sh 文件

```
[atguigu@hadoop102 bin]$ vim superset.sh
```

内容如下

```
#!/bin/bash

superset_status(){
    result=`ps -ef | awk '/gunicorn/ && !/awk/{print $2}' | wc -l`
    if [[ $result -eq 0 ]]; then
        return 0
    else
        return 1
    fi
}

superset_start(){
    source ~/.bashrc
    superset_status >/dev/null 2>&1
    if [[ $? -eq 0 ]]; then
        conda activate superset ; gunicorn --workers 5 --timeout 120 --bind hadoop102:8787 --daemon 'superset.app:create_app()'
    else
        echo "superset 正在运行"
    fi
}

superset_stop(){
    superset_status >/dev/null 2>&1
    if [[ $? -eq 0 ]]; then
        echo "superset 未在运行"
    else
        ps -ef | awk '/gunicorn/ && !/awk/{print $2}' | xargs kill -9
    fi
}

case $1 in
    start )
        echo "启动 Superset"
        superset_start
        ;;
    stop )
```

```
    echo "停止 Superset"
    superset_stop
;;
restart )
    echo "重启 Superset"
    superset_stop
    superset_start
;;
status )
    superset_status >/dev/null 2>&1
    if [[ $? -eq 0 ]]; then
        echo "superset 未在运行"
    else
        echo "superset 正在运行"
    fi
fi
esac
```

2) 加执行权限

```
[atguigu@hadoop102 bin]$ chmod +x superset.sh
```

3) 测试

启动 superset

```
[atguigu@hadoop102 bin]$ superset.sh start
```

停止 superset

```
[atguigu@hadoop102 bin]$ superset.sh stop
```

第3章 Superset 使用

3.1 对接 MySQL 数据源

3.1.1 安装依赖

```
(superset) [atguigu@hadoop102 ~]$ conda install mysqlclient
```

说明：对接不同的数据源，需安装不同的依赖，以下地址为官网说明

<https://superset.apache.org/docs/databases/installing-database-drivers>

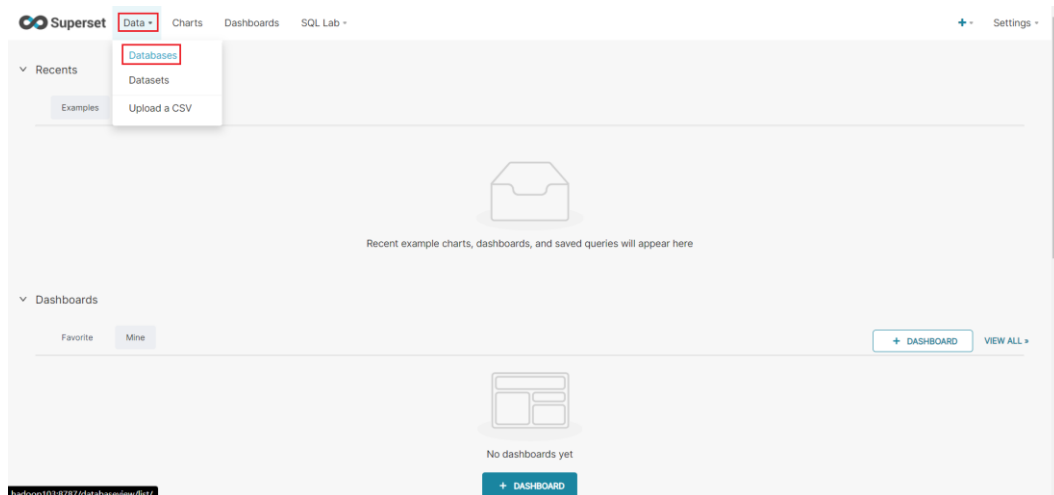
3.1.2 重启 Superset

```
(superset) [atguigu@hadoop102 ~]$ superset.sh restart
```

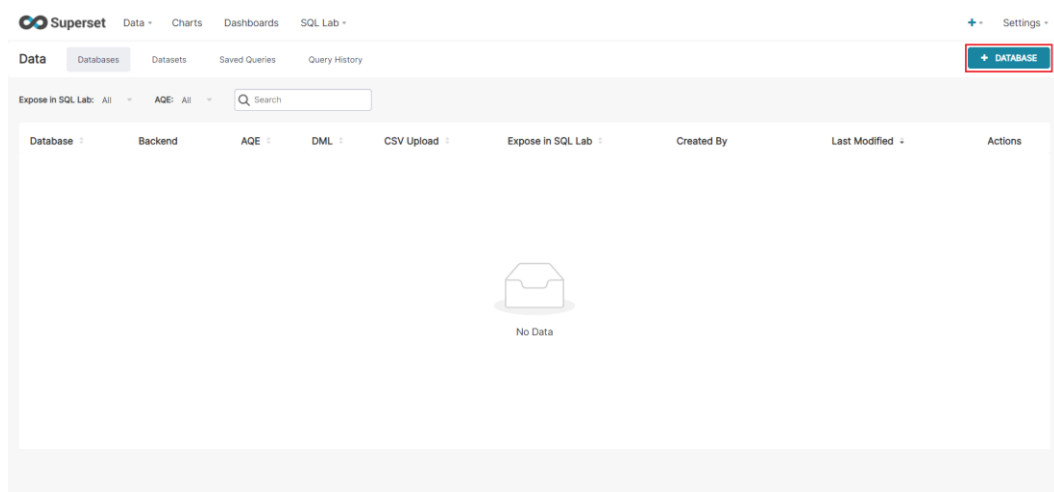
3.1.3 数据源配置

1) Database 配置

Step1: 点击 Data/Databases



Step2: 点击+DATABASE



Step3: 点击填写 Database 及 SQL Alchemy URI

注：SQLAlchemy URI 编写规范：`mysql://用户名:密码@主机名:端口号/数据库名称`
此处填写：

```
mysql://root:123456@hadoop102:3306/gmall_report?charset=utf8
```

Add Database

CONNECTION *

PERFORMANCE

SQL LAB SETTINGS

SECURITY

EXTRA

DATABASE NAME*

gmall_report

SQLALCHEMY URI*

mysql://root:123456@hadoop102:3306/gmall_report?charset=utf8

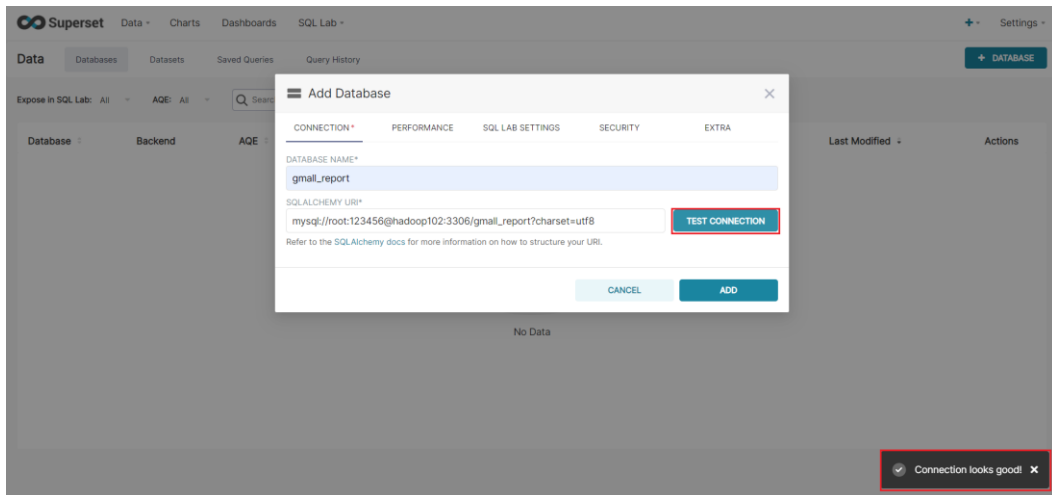
TEST CONNECTION

Refer to the [SQLAlchemy docs](#) for more information on how to structure your URI.

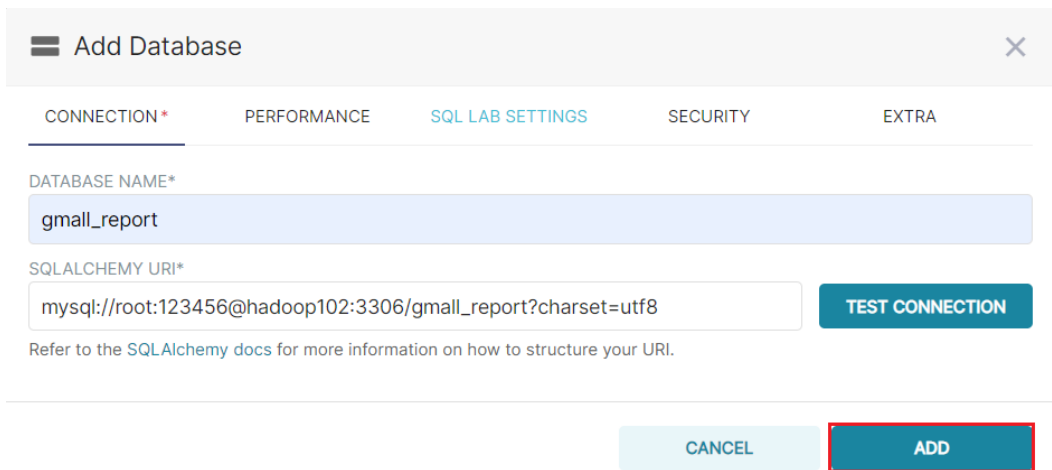
CANCEL

ADD

Step4: 点击 Test Connection, 出现 “Connection looks good!” 提示即表示连接成功

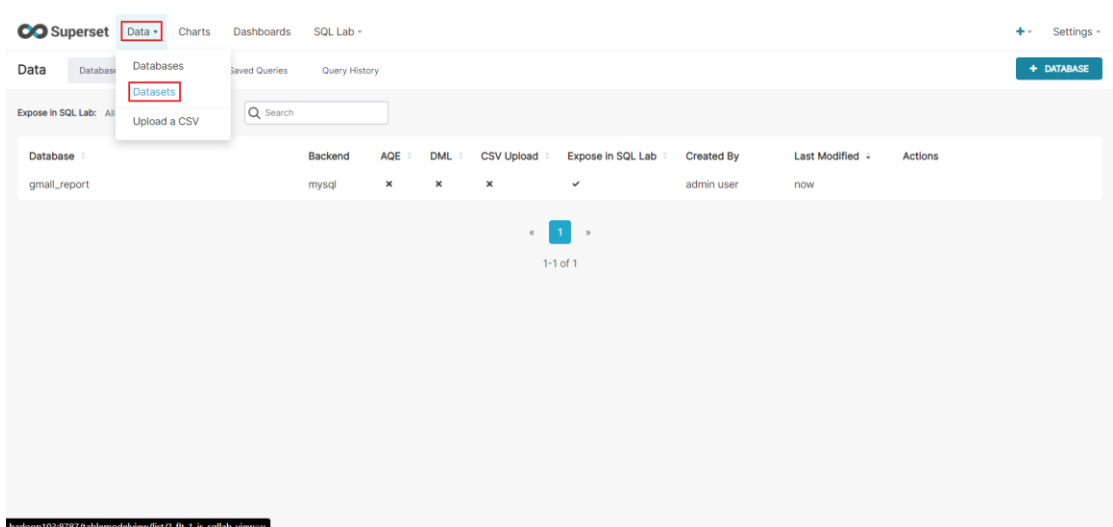


Step5: 点击 ADD

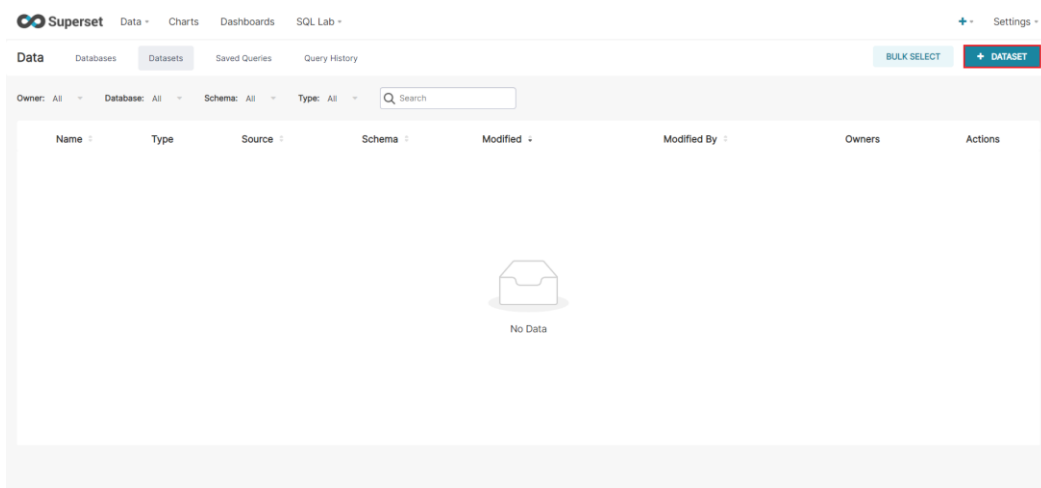


2) Table 配置

Step1: 点击 Data/Datasets



Step2: 点击 Data/ Datasets



Step3: 配置 Table

! Add Dataset

DATASOURCE

Database: mysql gmall_report

SCHEMA

Schema: gmall_report

SEE TABLE SCHEMA 31 INGMALL_REPORT

TABLE

ads_uv_count

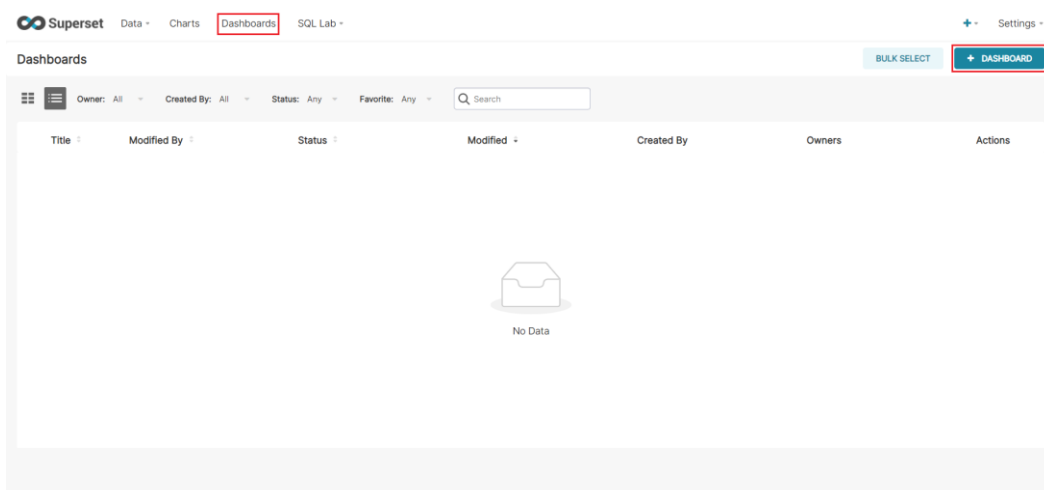
CANCEL

ADD

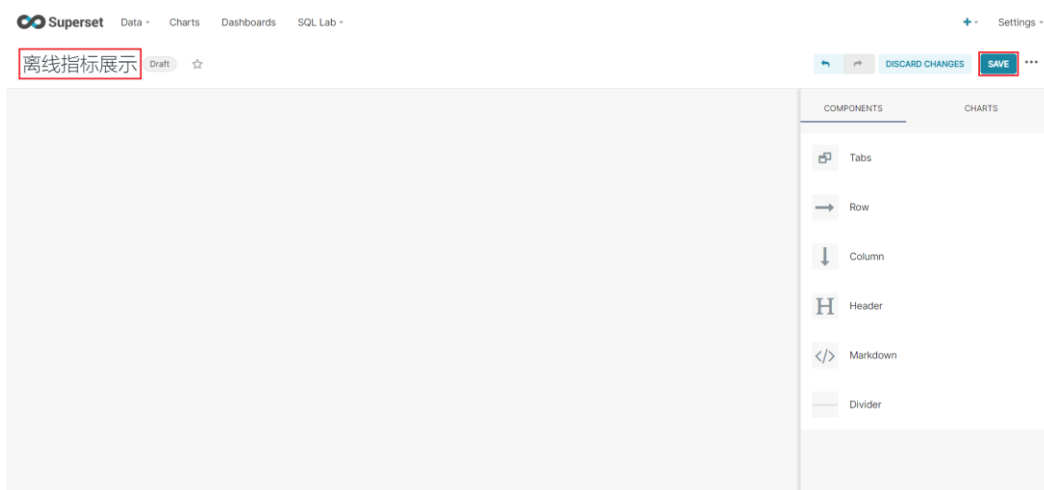
3.2 制作仪表盘

3.2.1 创建空白仪表盘

1) 点击 Dashboards/+DASHBOARDS

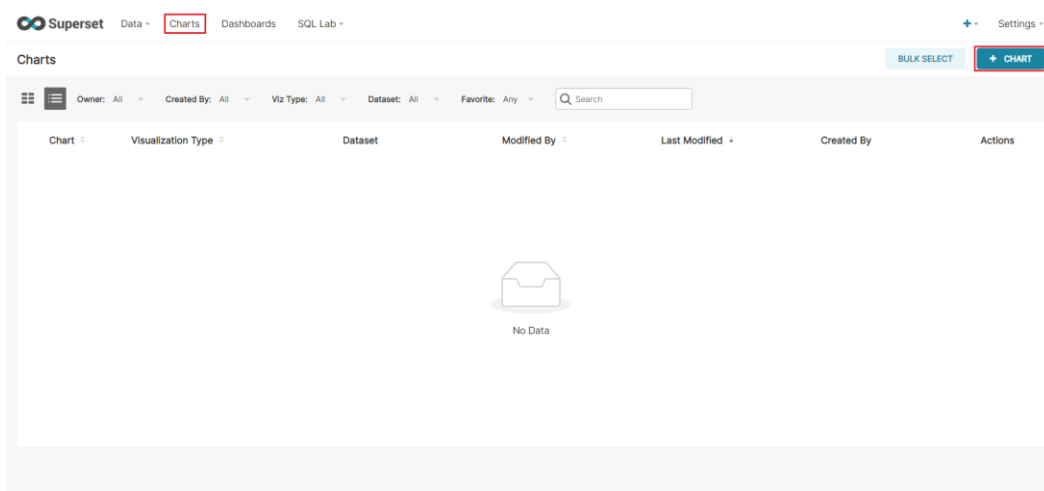


2) 命名并保存

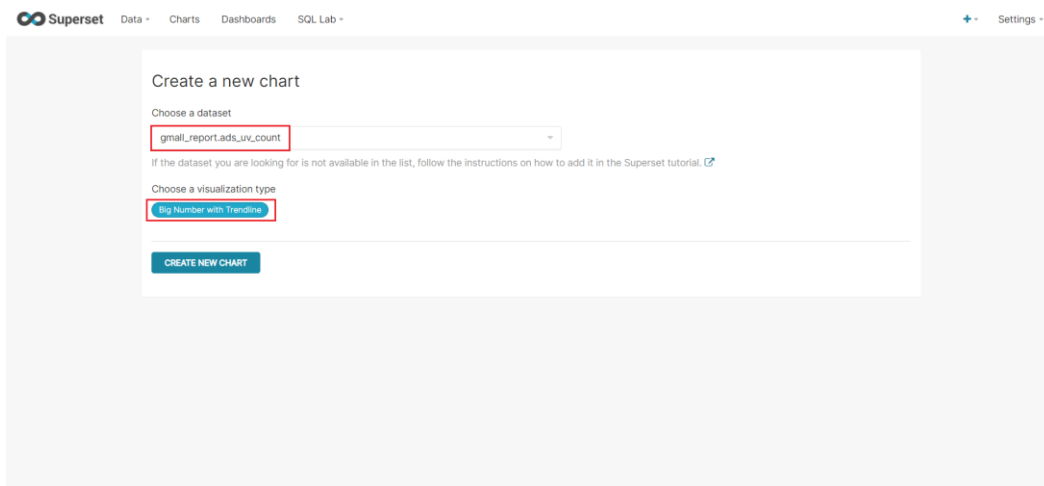


3.2.2 创建图表

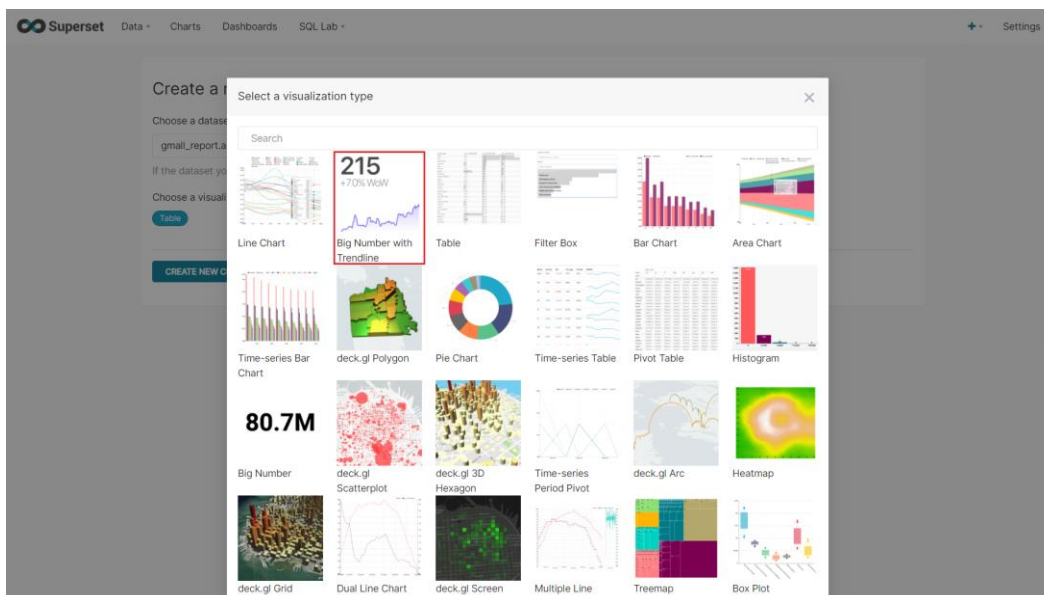
1) 点击 Charts/+CHART



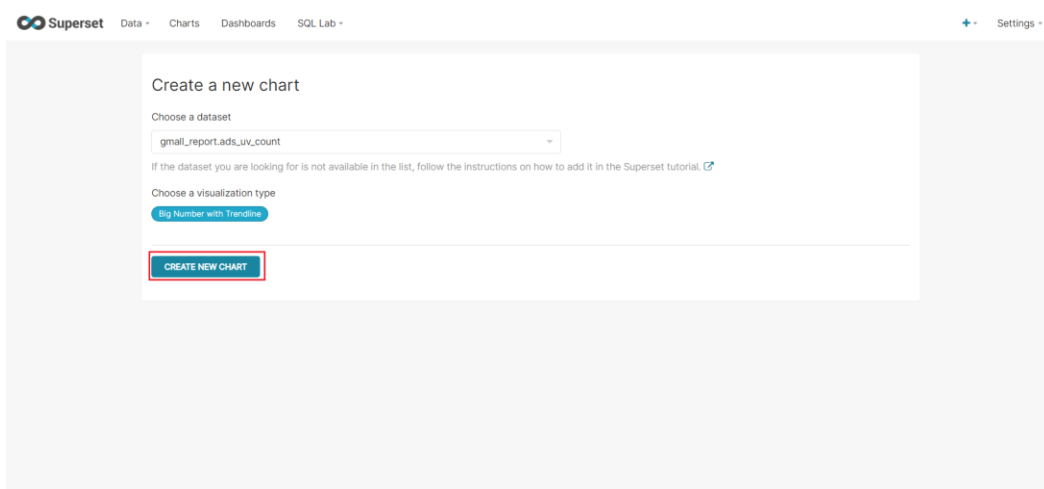
2) 选则数据源及图表类型



3) 选择何使的图表类型



4) 创建图表



5) 按照说明配置图表

Superset

Data

Charts

Dashboards

SQL Lab

+

Settings

→

RUN

SAVE

DATA

CUSTOMIZE

Chart Type

VISUALIZATION TYPE

Big Number with Trendline

Time

TIME COLUMN

dt

TIME GRAIN

day

TIME RANGE

$-\infty \leq col < \infty$

Query

METRIC

×

f(u) SUM(uv_count)

➤

FILTERS

+

Add filter

Options

COMPARISON PERIOD LAG

1

COMPARISON SUFFIX

与前一日相比

NUMBER FORMAT

Adaptative formatting

☒ SHOW TREND LINE

☒ START Y-AXIS AT 0

Advanced Analytics

- untitled

0 rows00:00:00.10

🔍

</>

.JSON

.CSV

☰

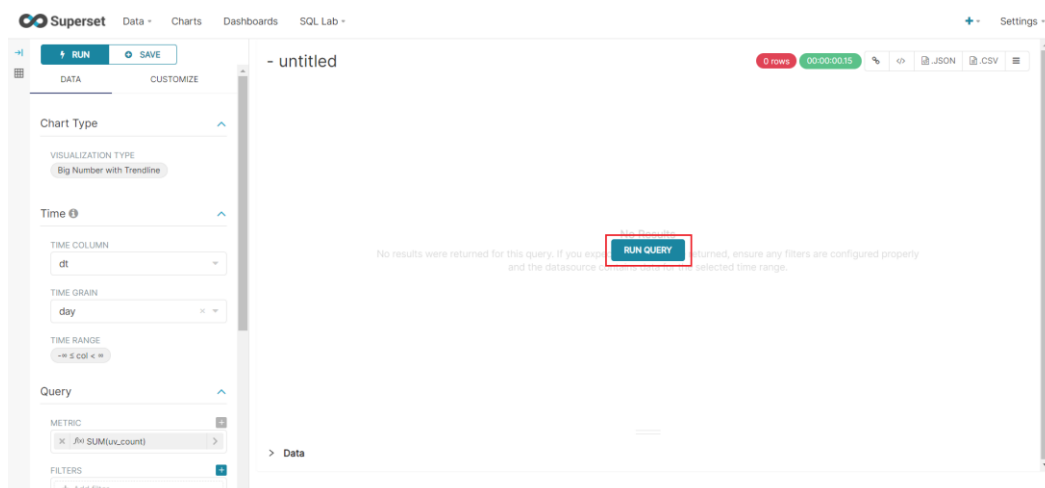
No Results

No results were returned for this query. If you expected results to be returned, please check that your filters are configured properly and the datasource contains data for the selected time range.

RUN QUERY

> Data

6) 点击 “Run Query”

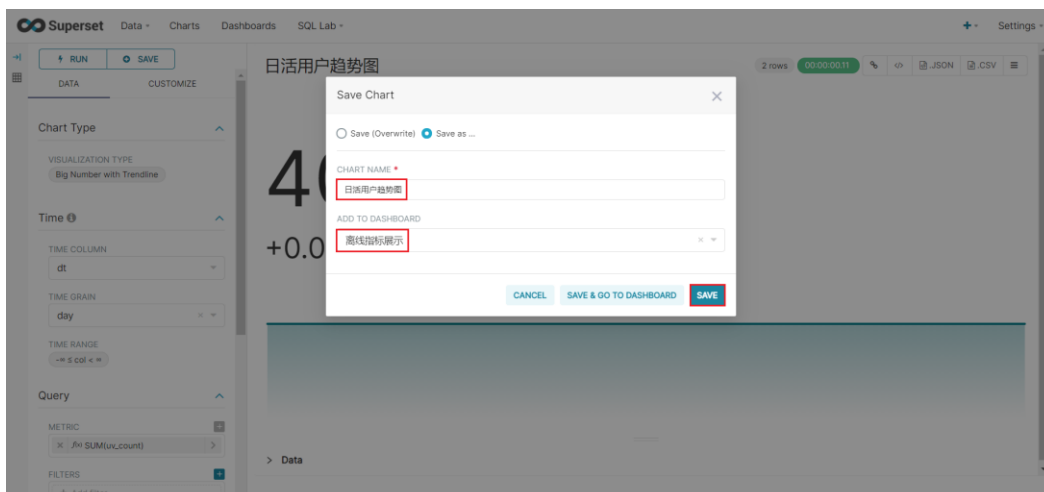


7) 如配置无误，可出现以下图标



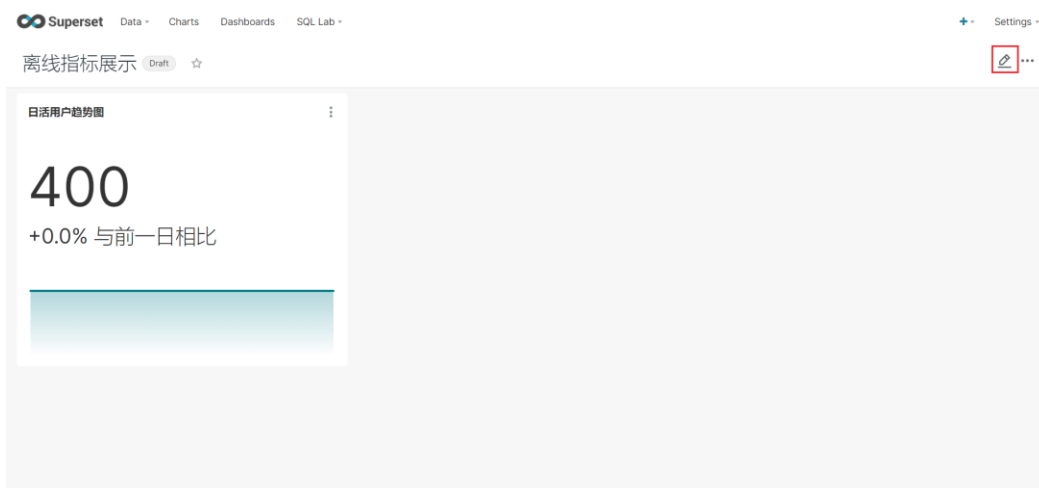
8) 命名该图表，并保存至仪表盘



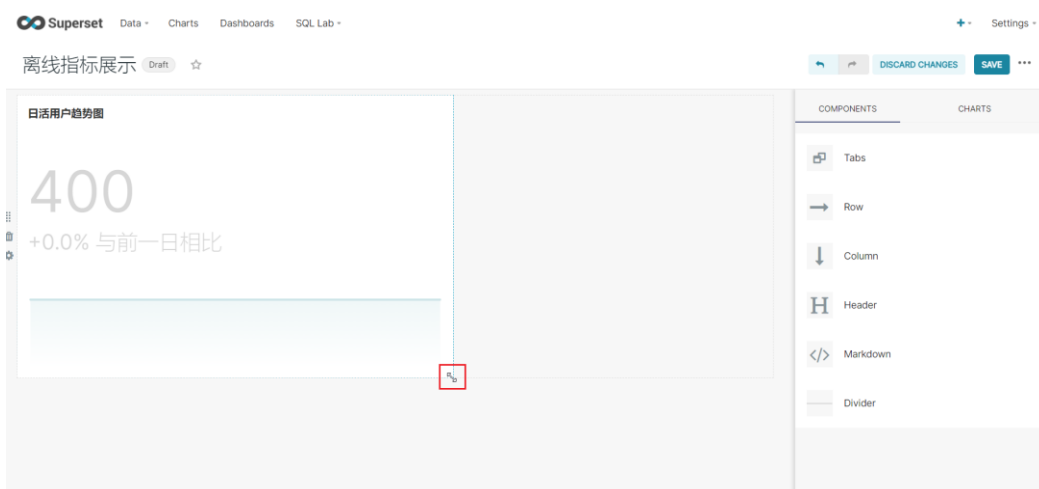


3.2.3 编辑仪表盘

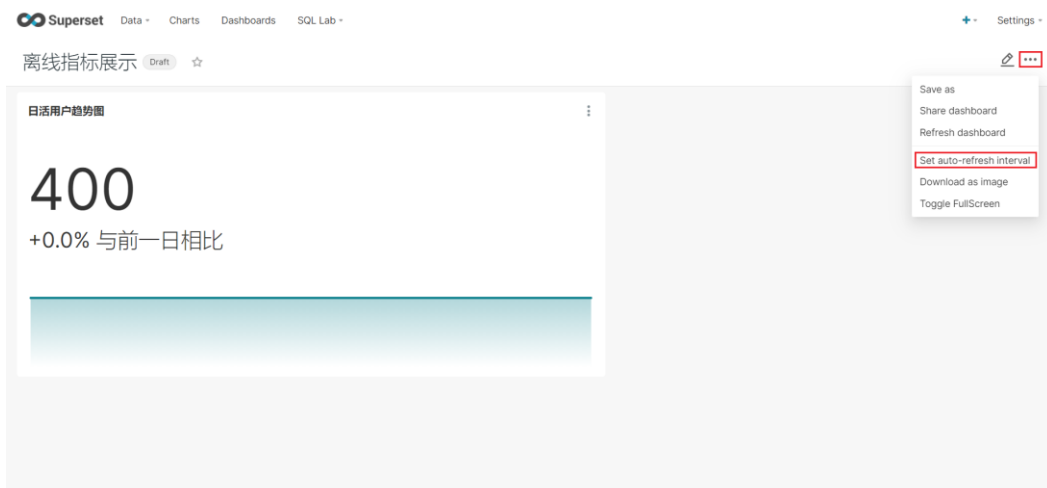
1) 打开仪表盘，点击编辑按钮



2) 调整图表大小以及图表布局



3) 点击下图中箭头，可调整仪表盘自动刷新时间



第 4 章 Superset 实战

4.1 制作地图

4.1.1 配置 Table

! Add Dataset

DATASOURCE

Database: mysql gmall_report

SCHEMA

Schema: gmall_report

SEE TABLE SCHEMA 31 IN GMALL_REPORT

TABLE

ads_order_count_by_area

CANCEL ADD

4.1.2 配置 Chart

Superset Data Charts Dashboards SQL Lab Settings

Create a new chart

Choose a dataset

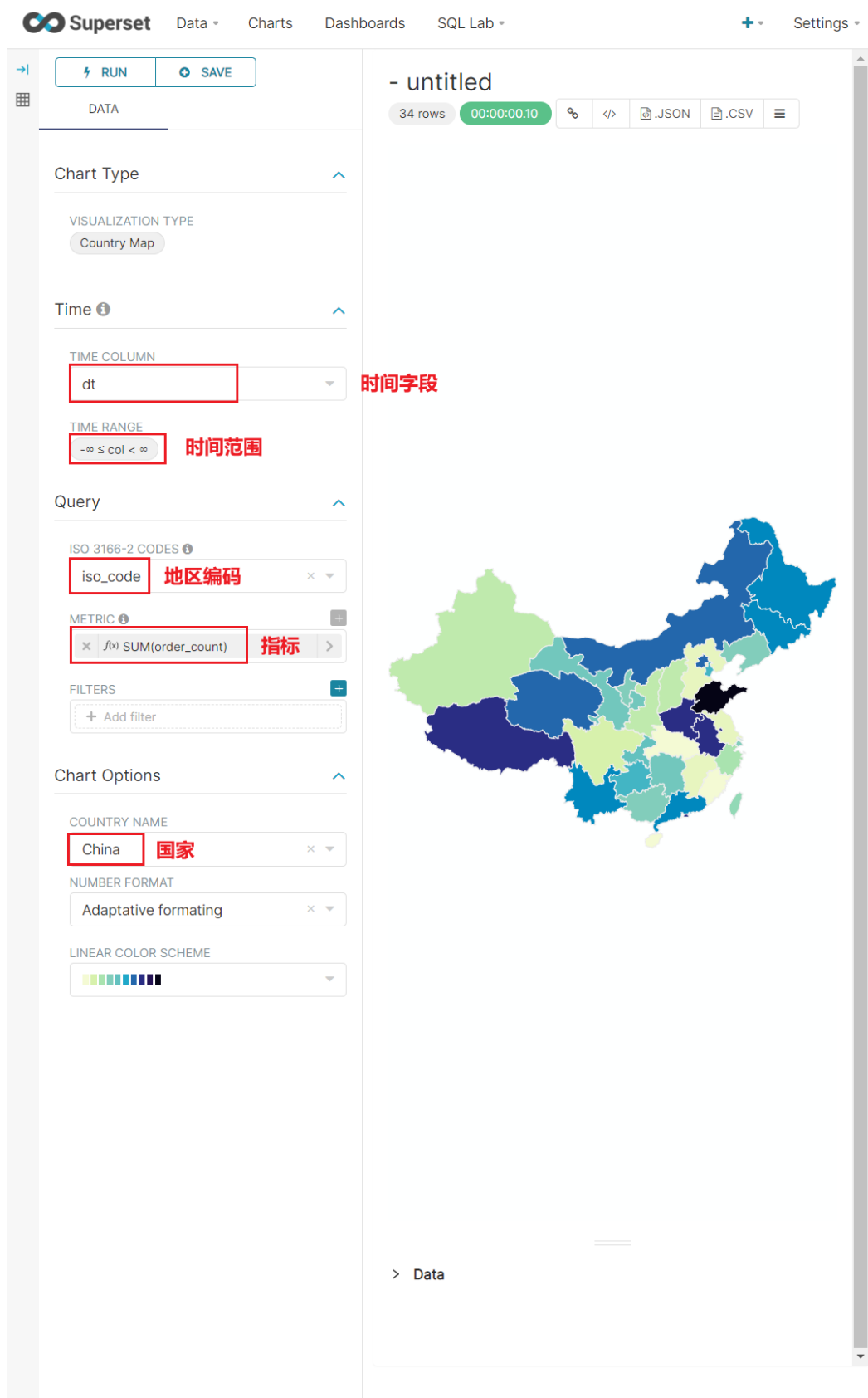
gmail_report_ads_order_count_by_area

If the dataset you are looking for is not available in the list, follow the instructions on how to add it in the Superset tutorial.

Choose a visualization type

Country Map

CREATE NEW CHART



4.2 制作饼状图

4.2.1 配置 Table

更多 [Java](#) - [大数据](#) - [前端](#) - [python](#) 人工智能资料下载，可百度访问：尚硅谷官网

! Add Dataset

DATASOURCE

Database: mysql gmail_report

SCHEMA

Schema: gmail_report

SEE TABLE SCHEMA 31 IN GMALL_REPORT

TABLE

ads_order_count_by_category

CANCEL

ADD

4.2.2 配置 Chart

Superset

Data Charts Dashboards SQL Lab

Settings

Create a new chart

Choose a dataset

gmail_report.ads_order_count_by_category

If the dataset you are looking for is not available in the list, follow the instructions on how to add it in the Superset tutorial.

Choose a visualization type

Sunburst Chart

CREATE NEW CHART

