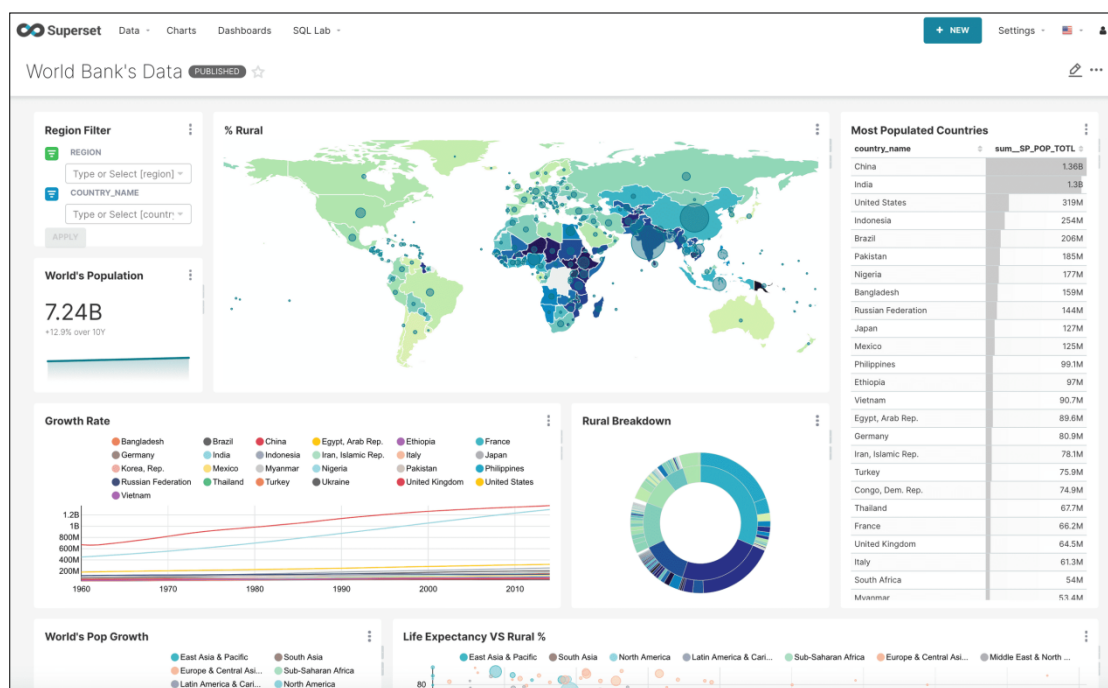


Superset

第 1 章 Superset 入门

1.1 Superset 概述

Apache Superset 是一个现代的数据探索和可视化平台。它功能强大且十分易用，可对接各种数据源，包括很多现代的大数据分析引擎，拥有丰富的图表展示形式，并且支持自定义仪表盘。



1.2 环境说明

本课程使用的服务器操作系统为 CentOS 7，Superset 对接的数据源为 MySQL 数据库。

第 2 章 Superset 安装

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 以上的环境。

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。

```
[iflytek@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 /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

```
[iflytek@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。

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

2.1.2 创建 Python3.8 环境

1) 配置 conda 国内镜像（此处可不配置，国外镜像比清华镜像更快）

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

(1) 查看 conda 的镜像 channel 配置

```
[iflytek@hadoop102 ~]$ conda config --show channels
- https://mirrors.tuna.tsinghua.edu.cn/anaconda/pkgs/main
- https://mirrors.tuna.tsinghua.edu.cn/anaconda/pkgs/free
- defaults
```

(2) 删除清华镜像,继续用默认的

```
[iflytek@hadoop102 ~]$ conda config --remove-key channels
[iflytek@hadoop102 ~]$ conda config --show channels
- defaults
```

2) 创建 Python3.8 环境

```
(base) [iflytek@hadoop102 ~]$ conda create --name superset python=3.8.16
```

➤ 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) [iflytek@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) [iflytek@hadoop102 ~]$ conda deactivate
```

4) 执行 `python -V` 命令查看 python 版本

```
(superset) [iflytek@hadoop102 ~]$ python -V
Python 3.8.16
```

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) [iflytek@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) 更新 pip

```
(superset) [iflytek@hadoop102 ~]$ pip install --upgrade 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

(1) 上传 base.txt 文件至任意路径

```
(superset) [iflytek@hadoop102 ~]$ ll ~
总用量 12
-rw-r--r--. 1 iflytek iflytek 5795 2月  3 15:27 base.txt
drwxrwxr-x. 2 iflytek iflytek 4096 2月  6 11:42 bin
```

该文件可用于指定 superset 依赖组件及版本, 下载地址及内容如下。

<https://raw.githubusercontent.com/apache/superset/2.0.0/requirements/base.txt>



base.txt

(2) 在 base.txt 所在目录下执行如下命令, 安装 SuperSet

```
(superset) [iflytek@hadoop102 ~]$ pip install apache-
superset==2.0.0 -i https://pypi.tuna.tsinghua.edu.cn/simple -r
```

base.txt

说明:

-i 的作用是指定镜像, 这里选择国内镜像

-r 的作用是指定 superset 依赖组件及相应版本, 指向 base.txt 文件即可

2.2.3 配置 Superset 元数据库

Superset 的元数据支持 MySQL、PostgreSQL, 此处采用 MySQL。

1) 在 MySQL 中创建 superset 元数据库

```
mysql> CREATE DATABASE superset DEFAULT CHARACTER SET utf8 DEFAULT  
COLLATE utf8_general_ci;
```

2) 创建 superset 用户

```
mysql> create user superset@'%' identified WITH  
mysql_native_password BY 'superset';  
mysql> grant all privileges on *.* to superset@'%' with grant option;  
mysql> flush privileges;
```

3) 修改 superset 配置文件

```
(superset) [iflytek@hadoop102 ~]$ vim  
/opt/module/miniconda3/envs/superset/lib/python3.8/site-  
packages/superset/config.py
```

修改内容如下: (184、185 行)

```
# SQLALCHEMY_DATABASE_URI = "sqlite:/// " + os.path.join(DATA_DIR,  
"superset.db")  
SQLALCHEMY_DATABASE_URI =  
'mysql://superset:superset@hadoop102:3306/superset?charset=utf8'
```

4) 安装 python mysql 驱动

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

5) 初始化 superset 元数据

```
(superset) [iflytek@hadoop102 ~]$ export FLASK_APP=superset  
(superset) [iflytek@hadoop102 ~]$ superset db upgrade
```

2.2.4 Superset 初始化

1) 创建管理员用户

```
(superset) [iflytek@hadoop102 ~]$ superset fab create-admin
```

2) 初始化 superset

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

2.2.5 启动 Superset

1) 安装 gunicorn

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

说明：

--workers: 指定进程个数

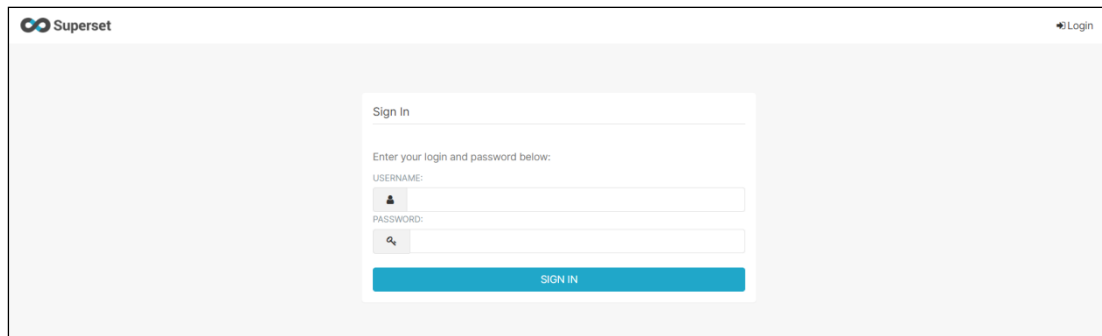
--timeout: worker 进程超时时间，超时会自动重启

--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

(1) 停掉 gunicorn 进程

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

(2) 退出 superset 环境

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

2.2.4 superset 启停脚本

1) 创建 superset.sh 文件

```
[iflytek@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
}
```

```

        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
    esac

```

2) 加执行权限

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

3) 测试

(1) 启动 superset

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

(2) 停止 superset

```
[iflytek@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（前边步骤已安装，可跳过）

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

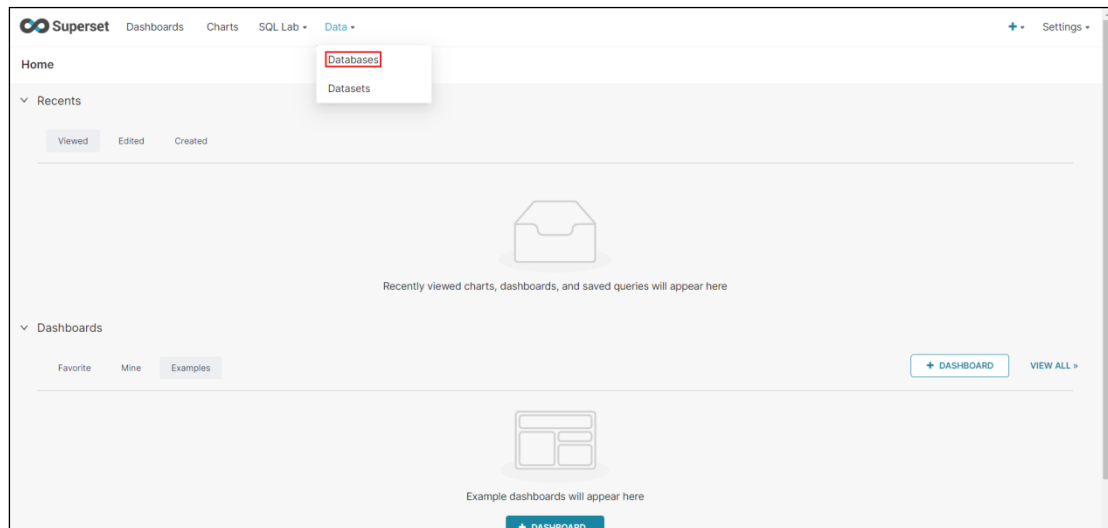
说明：对接不同的数据源，需安装不同的依赖，以下地址为官网说明。

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

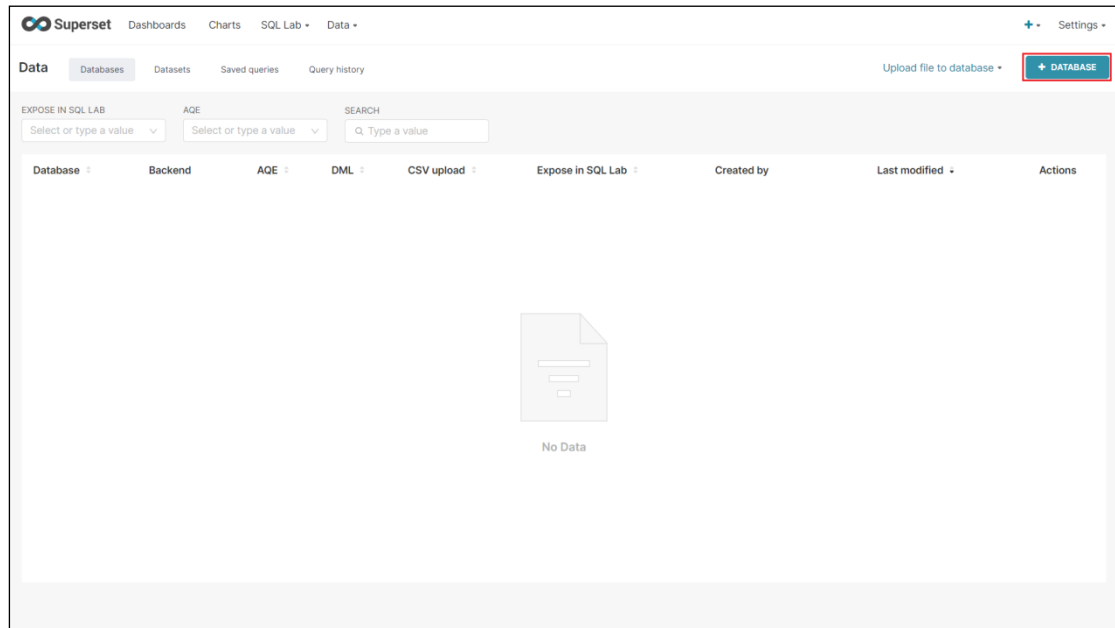
3.1.2 数据源配置

1) Database 配置

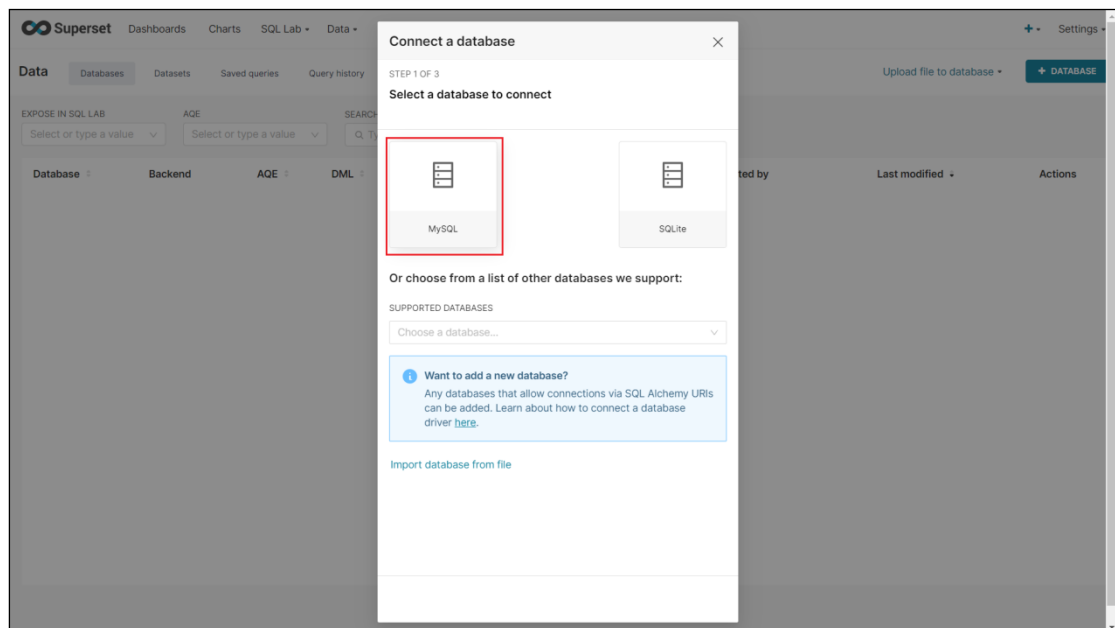
Step1: 点击 Data/Databases



Step2: 点击 +DATABASE



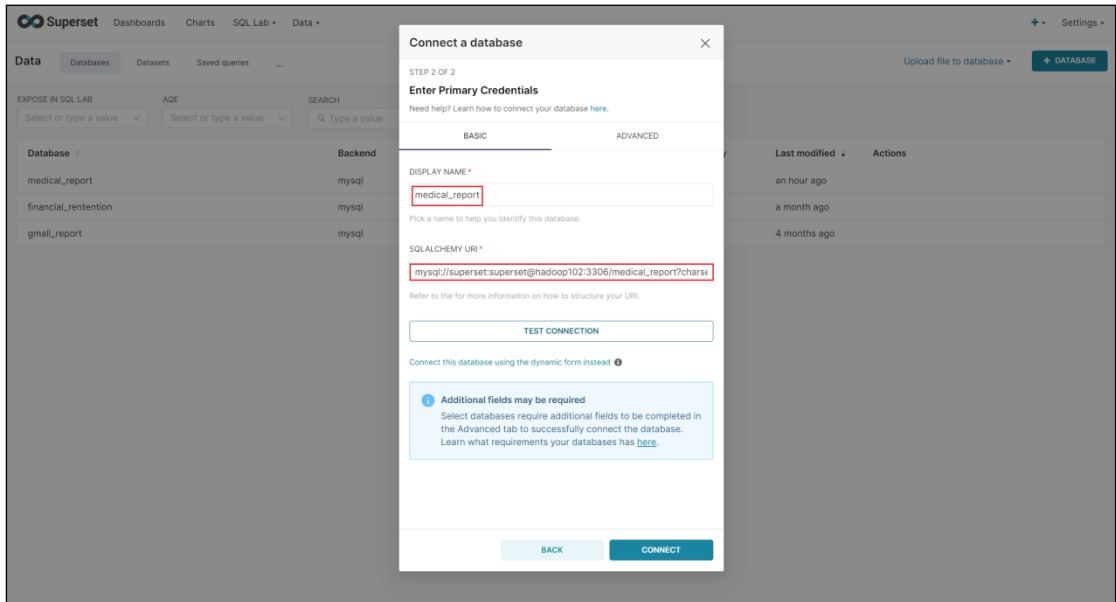
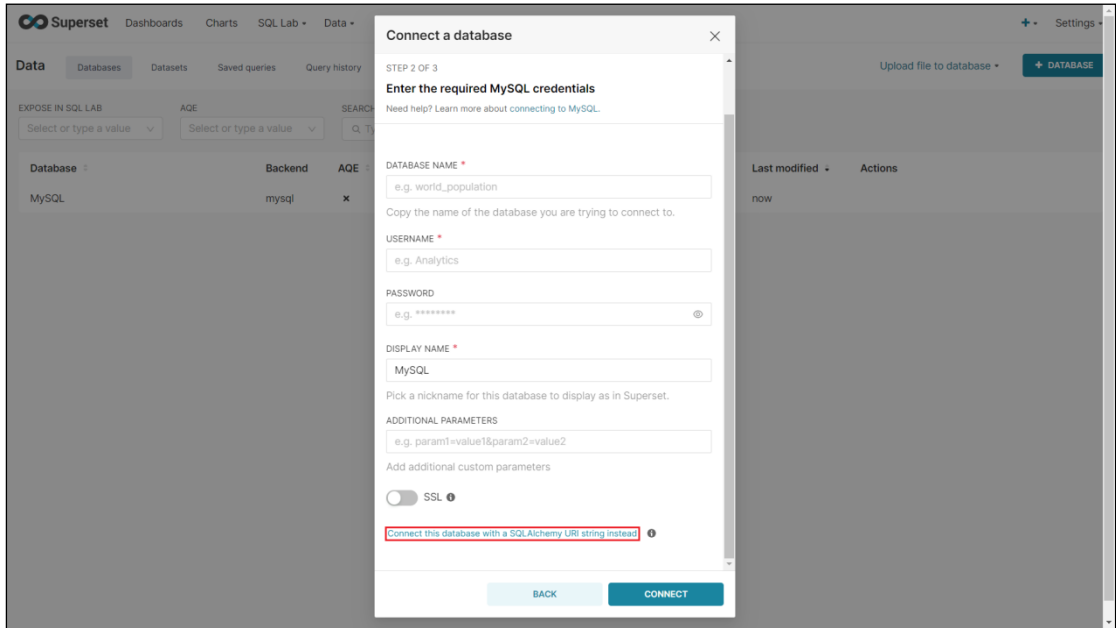
Step3: 点击填写 Database 及 SQL Alchemy URI



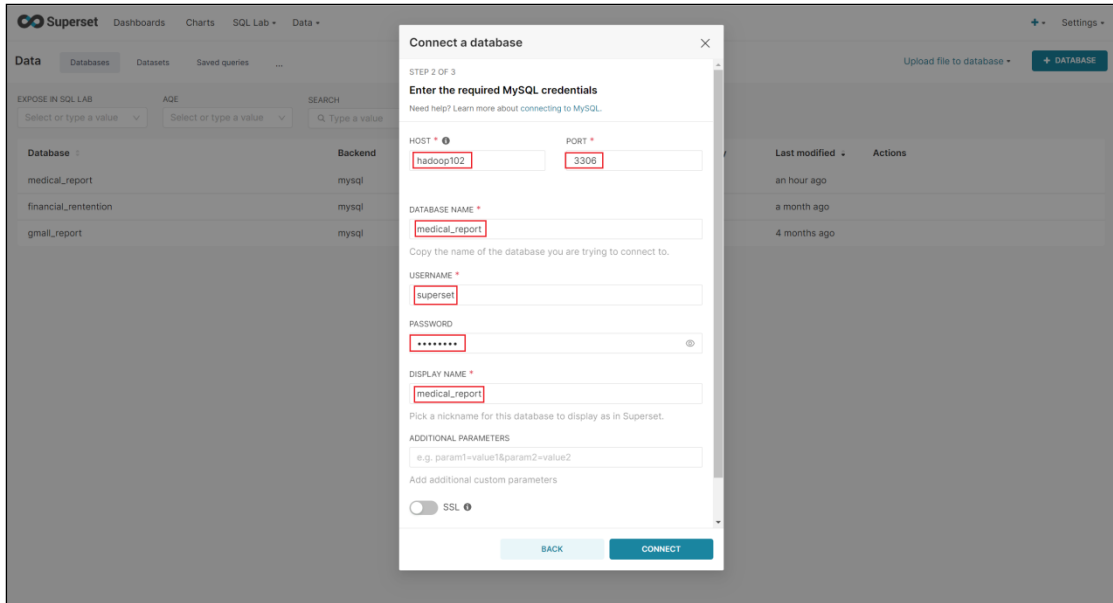
(1) 方式一：通过 URL 连接

注：SQL Alchemy 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://superset:superset@hadoop102:3306/medical_report?charset=utf8mb4

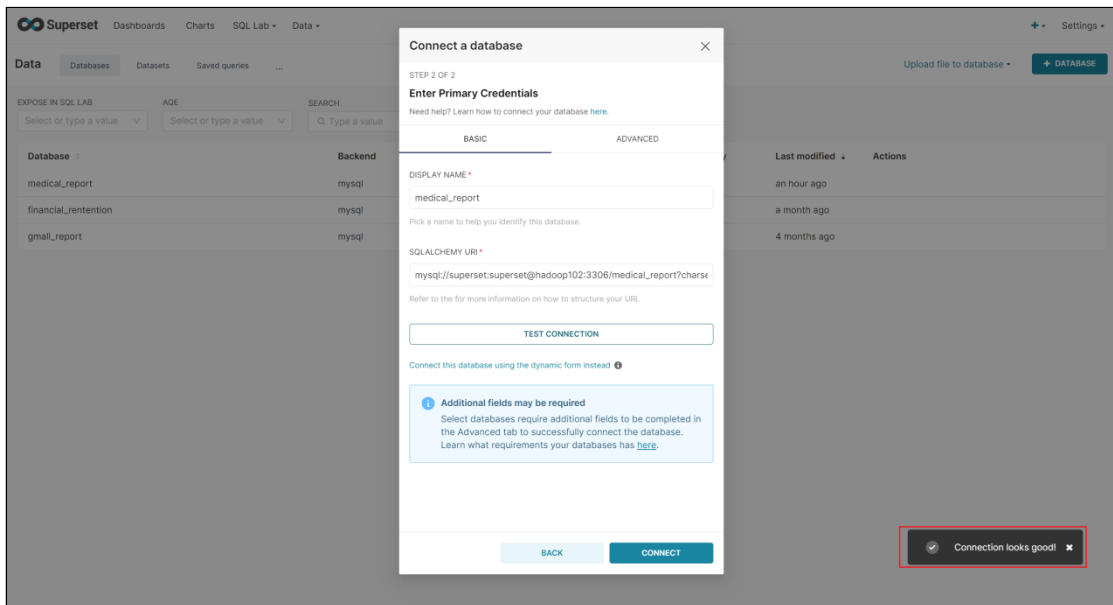


(2) 方式二：逐个输入认证信息

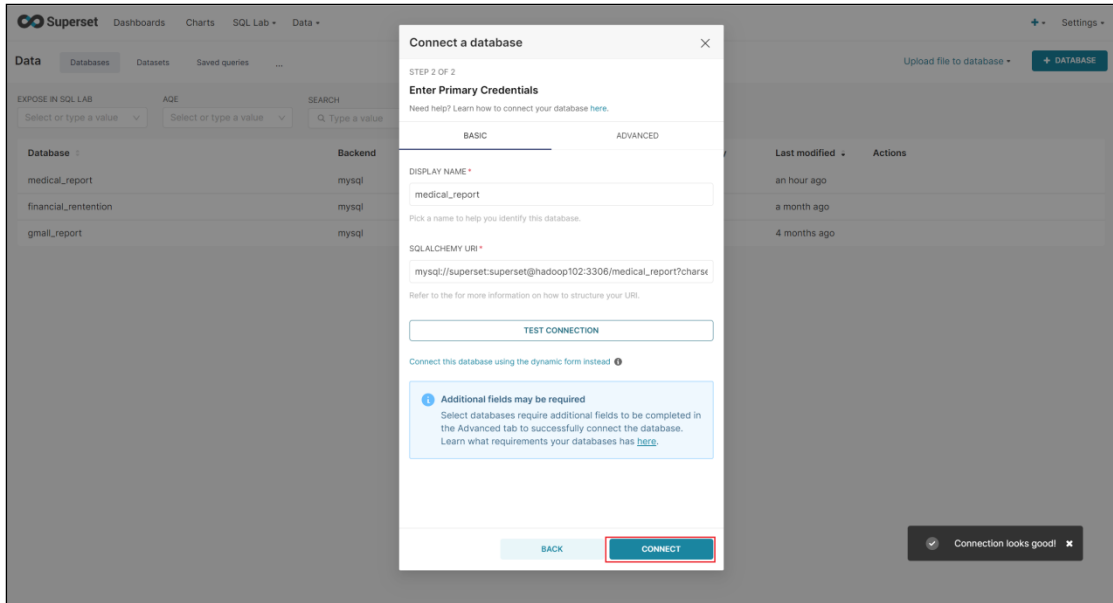


此处选择方式一。

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

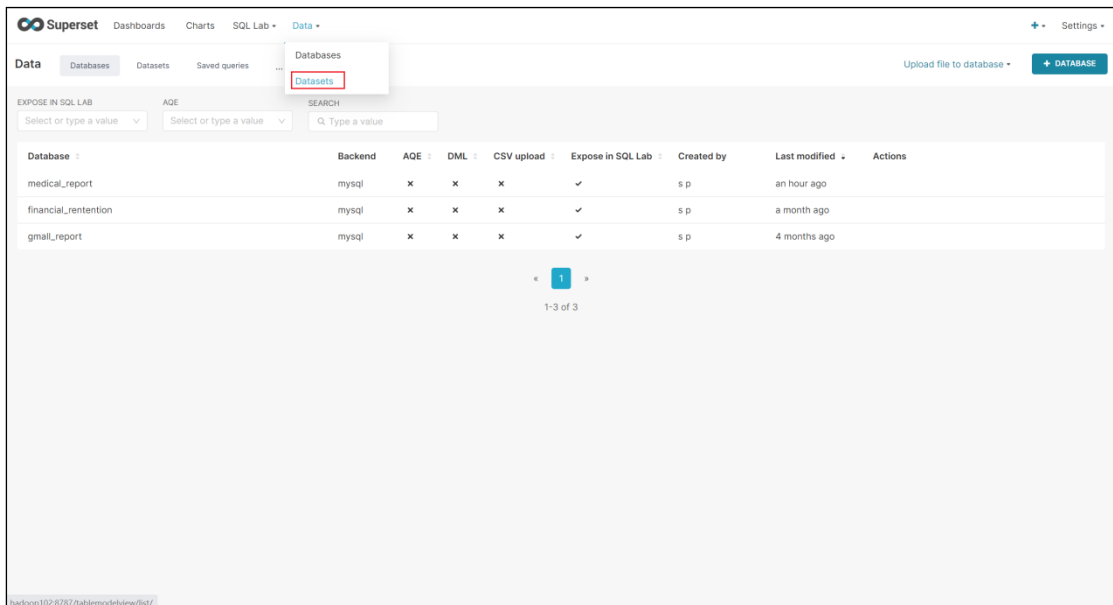


Step5: 点击 CONNECT

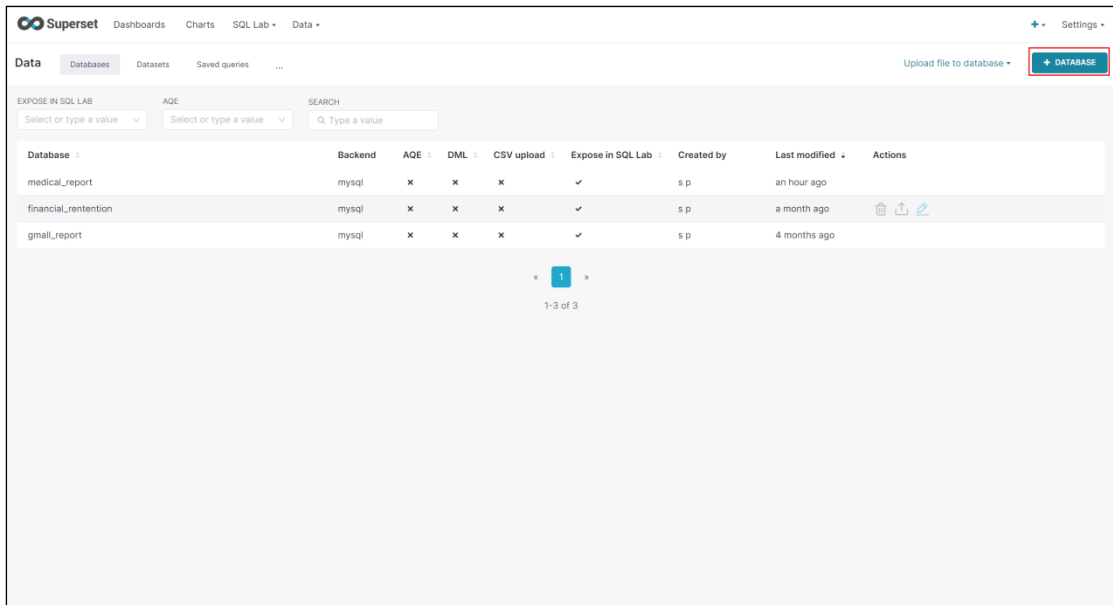


2) Table 配置

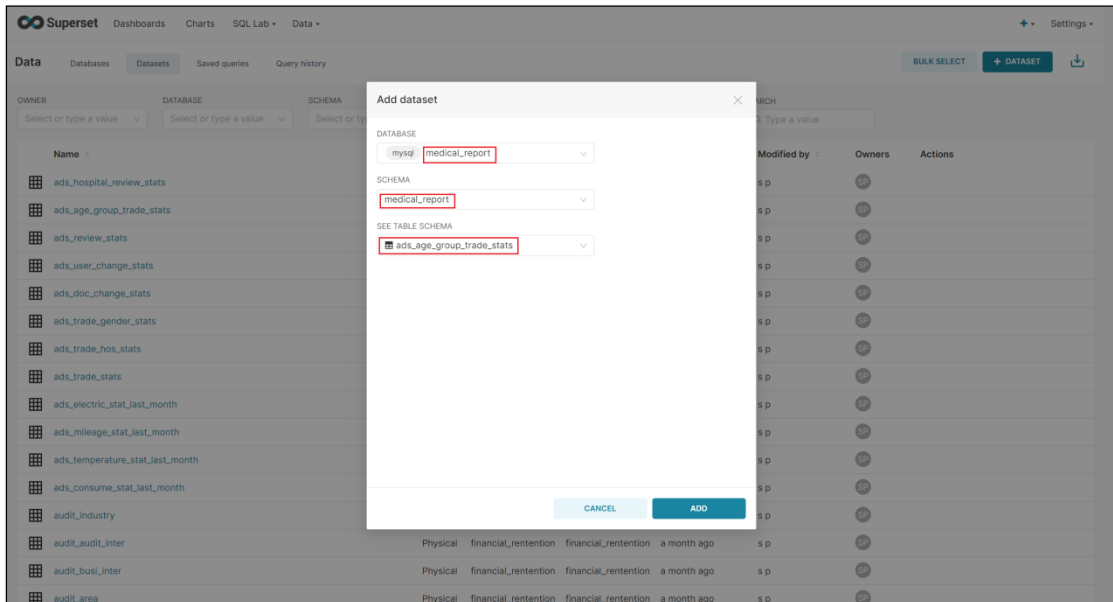
Step1: 点击 Data/Datasets



Step2: 点击+DATASET



Step3: 配置 Table



3.2 制作仪表盘

3.2.1 创建空白仪表盘

1) 点击 Dashboards/+DASHBOARDS

Superset

DashboardsChartsSQL LabData

Settings

Data

DatabasesDatasetsSaved queriesQuery history

BULK SELECT+ DATASET

OWNER

DATABASE

SCHEMA

TYPE

CERTIFIED

SEARCH

Select or type a value

Select or type a value

Select or type a value

Select or type a value

Select or type a value

Q. Type a value

Name	Type	Database	Schema	Modified	Modified by	Owners	Actions
ads_hospital_review_stats	Physical	medical_report	medical_report	35 minutes ago	s p		
ads_age_group_trade_stats	Physical	medical_report	medical_report	an hour ago	s p		
ads_review_stats	Physical	financial_rentention	medical_report	a month ago	s p		
ads_user_change_stats	Physical	financial_rentention	medical_report	a month ago	s p		
ads_doc_change_stats	Physical	gmail_report	medical_report	a month ago	s p		
ads_trade_gender_stats	Physical	financial_rentention	medical_report	a month ago	s p		
ads_trade_hos_stats	Physical	financial_rentention	medical_report	a month ago	s p		
ads_trade_stats	Physical	financial_rentention	medical_report	a month ago	s p		
ads_electric_stat_last_month	Physical	gmail_report	car_data_report	a month ago	s p		
ads_mileage_stat_last_month	Physical	financial_rentention	car_data_report	a month ago	s p		
ads_temperature_stat_last_month	Physical	financial_rentention	car_data_report	a month ago	s p		
ads_consume_stat_last_month	Physical	financial_rentention	car_data_report	a month ago	s p		
audit_industry	Physical	financial_rentention	financial_rentention	a month ago	s p		
audit_audit_inter	Physical	financial_rentention	financial_rentention	a month ago	s p		
audit_busi_inter	Physical	financial_rentention	financial_rentention	a month ago	s p		
hadoop102:8787/dashboard/list/	Physical	financial_rentention	financial_rentention	a month ago	s p		

Superset

DashboardsChartsSQL LabData

Settings

Dashboards

BULK SELECT+ DASHBOARD

OWNER

CREATED BY

STATUS

FAVORITE

CERTIFIED

SEARCH

Select or type a value

Select or type a value

Select or type a value

Select or type a value

Select or type a value

Q. Type a value

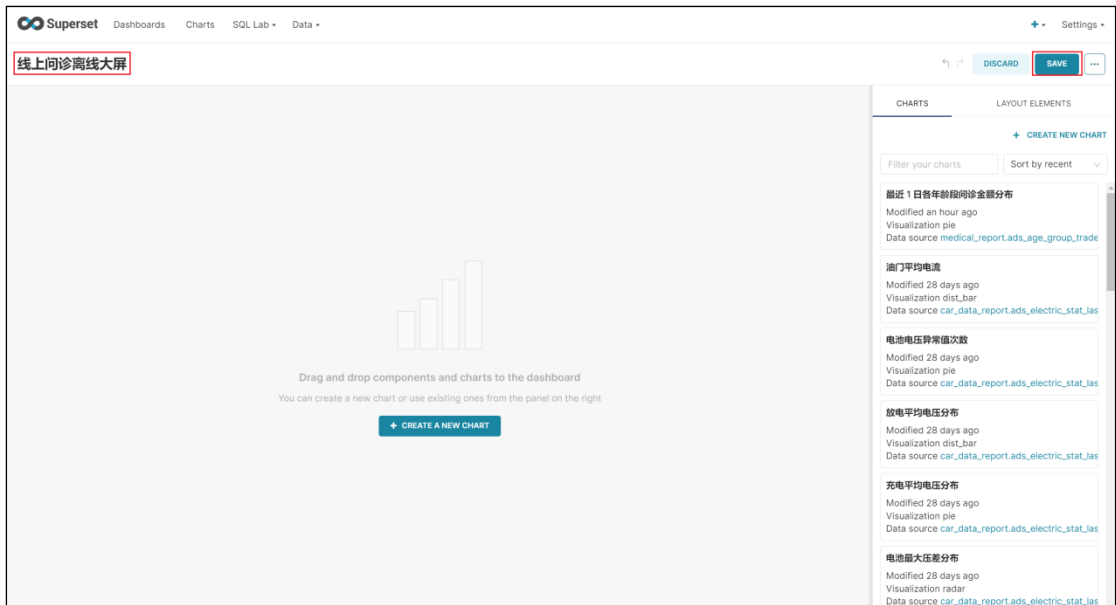
Title	Modified by	Status	Modified	Created by	Owners	Actions
☆ 线上问诊需求大屏	s p	Draft	41 minutes ago	s p		
☆ 新能源汽车可视化大屏二	s p	Draft	28 days ago	s p		
☆ 新能源汽车可视化大屏一	s p	Draft	28 days ago	s p		
☆ 医疗问诊可视化大屏二	s p	Draft	a month ago	s p		
☆ 医疗问诊可视化大屏一	s p	Draft	a month ago	s p		
☆ 新能源汽车可视化大屏三	s p	Draft	a month ago	s p		
☆ 信贷金融可视化	s p	Draft	a month ago	s p		
☆ gmail_report	s p	Draft	4 months ago	s p		

1

1-8 of 8

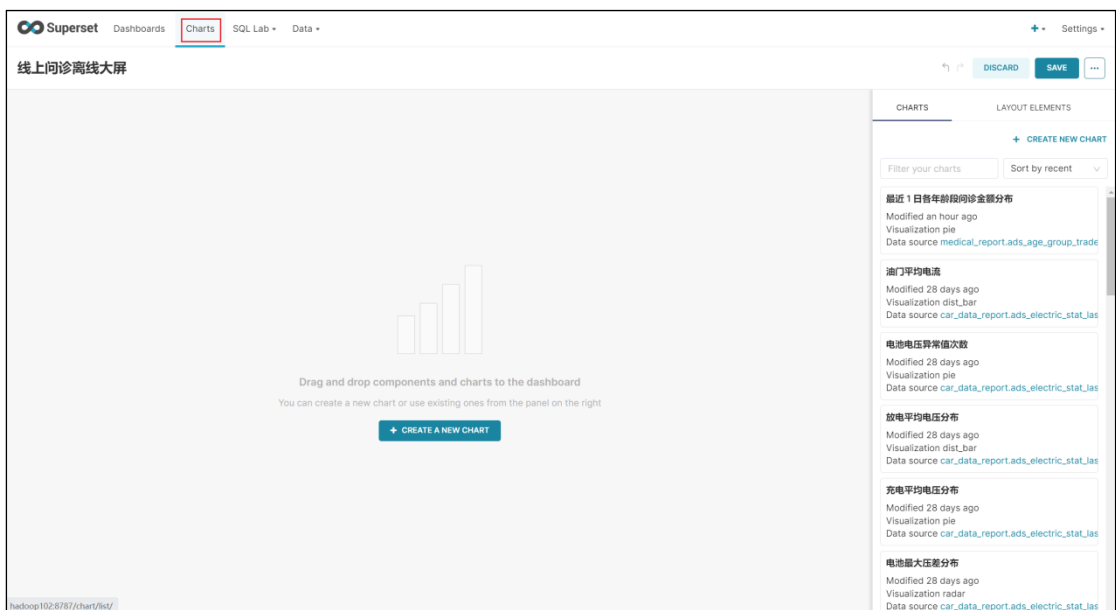
hadoop102:8787/dashboard/list/

2) 命名并保存



3.2.2 创建图表

1) 点击 Charts/+CHART



Superset

DashboardChartsSQL LabData

Settings

Charts

BULK SELECT

CHART

OWNER

CREATED BY

CHART TYPE

DATASET

FAVORITE

CERTIFIED

SEARCH

Select or type a value

Select or type a value

Select or type a value

Select or type a value

Select or type a value

Select or type a value

Q Type a value

Chart

Visualization type

Dataset

Modified by

Last modified

Created by

Actions

最近 1 日各年龄段问诊金额分布

Pie Chart

medical_report.ads_age_group_trade_stats

s p

in 7 hours

s p

油门平均电流

Bar Chart

car_data_report.ads_electric_stat_last_month

s p

a month ago

s p

电池电压异常值次数

Pie Chart

car_data_report.ads_electric_stat_last_month

s p

a month ago

s p

放电平均电压分布

Bar Chart

car_data_report.ads_electric_stat_last_month

s p

a month ago

s p

充电平均电压分布

Pie Chart

car_data_report.ads_electric_stat_last_month

s p

a month ago

s p

电池最大压差分布

Radar Chart

car_data_report.ads_electric_stat_last_month

s p

a month ago

s p

单体电池最高电压分布

Pie Chart

car_data_report.ads_electric_stat_last_month

s p

a month ago

s p

百公里急加减速次数分布

Pie Chart

car_data_report.ads_mileage_stat_last_month

s p

a month ago

s p

日均里程分布

Bar Chart

car_data_report.ads_mileage_stat_last_month

s p

a month ago

s p

电机控制器平均温度分布

Radar Chart

car_data_report.ads_temperature_stat_last...

s p

a month ago

s p

平均里程分布

Radar Chart

car_data_report.ads_mileage_stat_last_month

s p

a month ago

s p

电池温度异常值次数分布

Bar Chart

car_data_report.ads_temperature_stat_last...

s p

a month ago

s p

次均充电时长分布

Bar Chart

car_data_report.ads_consume_stat_last_mo...

s p

a month ago

s p

总次数及快充次数分布

Pie Chart

car_data_report.ads_consume_stat_last_mo...

s p

a month ago

s p

燃油消耗率及 SOC 效率

Radar Chart

car_data_report.ads_consume_stat_last_mo...

s p

a month ago

s p

电机平均温度分布

Pie Chart

car_data_report.ads_temperature_stat_last...

s p

a month ago

s p

每百公里SOC消耗分布

Pie Chart

car_data_report.ads_consume_stat_last_mo...

s p

a month ago

s p

2) 选则数据源及图表类型并创建图表

Superset

DashboardChartsSQL LabData

Settings

Create a new chart

CHOOSE A DATASET

ads_age_group_trade_stats

Instructions to add a dataset are available in the Superset tutorial

CHOOSE CHART TYPE

All charts

Recommended tags

Popular

ECharts

Advanced-Analytics

Category

Correlation

Distribution

Evolution

Flow

KPI

Search all charts

215

80.7M

Big Number with Trendline

Big Number

Table

Pivot Table v2

Time-series Line Chart

Time-series Area Chart

Time-series Bar Chart v2

Time-series Scatter Plot

Pie Chart

Bar Chart

World Map

Pie Chart

Aesthetic

Categorical

Circular

Comparison

Percentages

Popular

Proportional

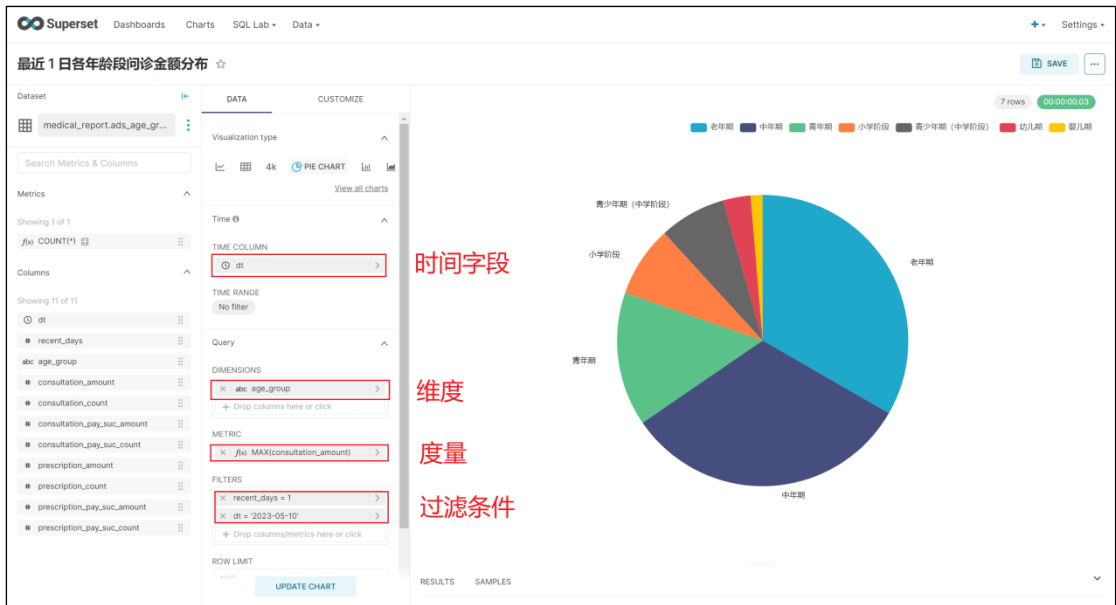
ECharts

The classic. Great for showing how much of a company each investor gets, what demographics follow your blog, or what portion of the budget goes to the military industrial complex. Pie charts are also used to show the distribution of data across categories. Pie charts are also used to show the distribution of data across categories.

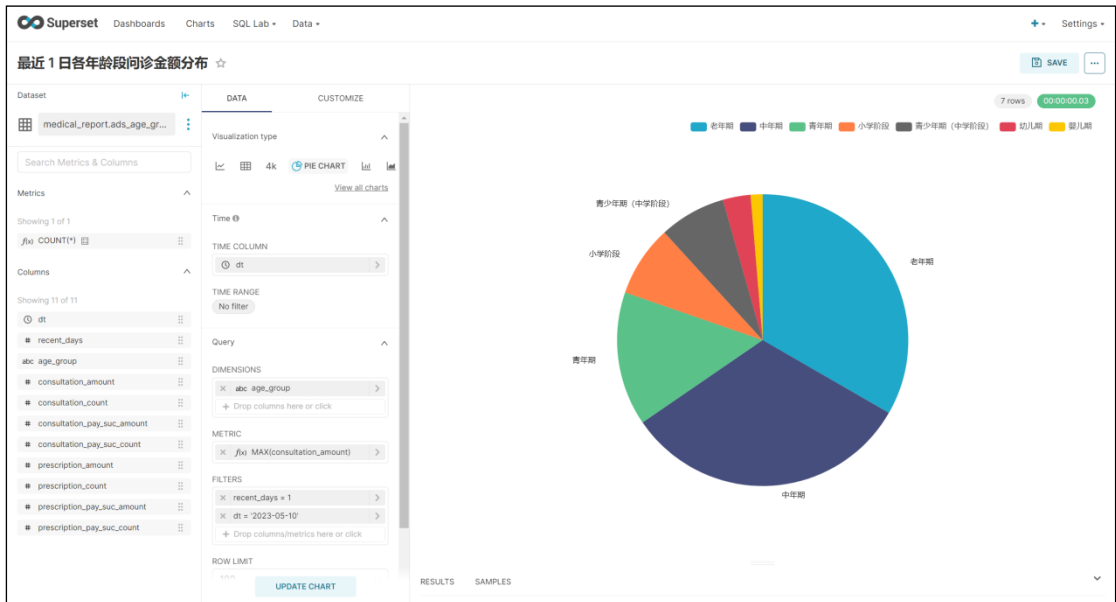
Examples

CREATE NEW CHART

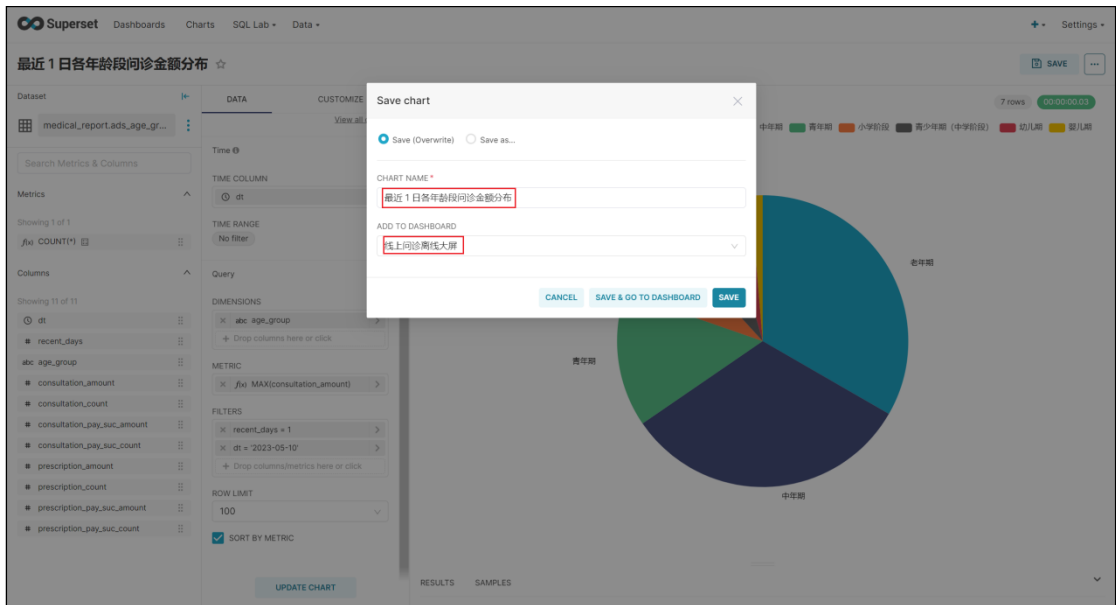
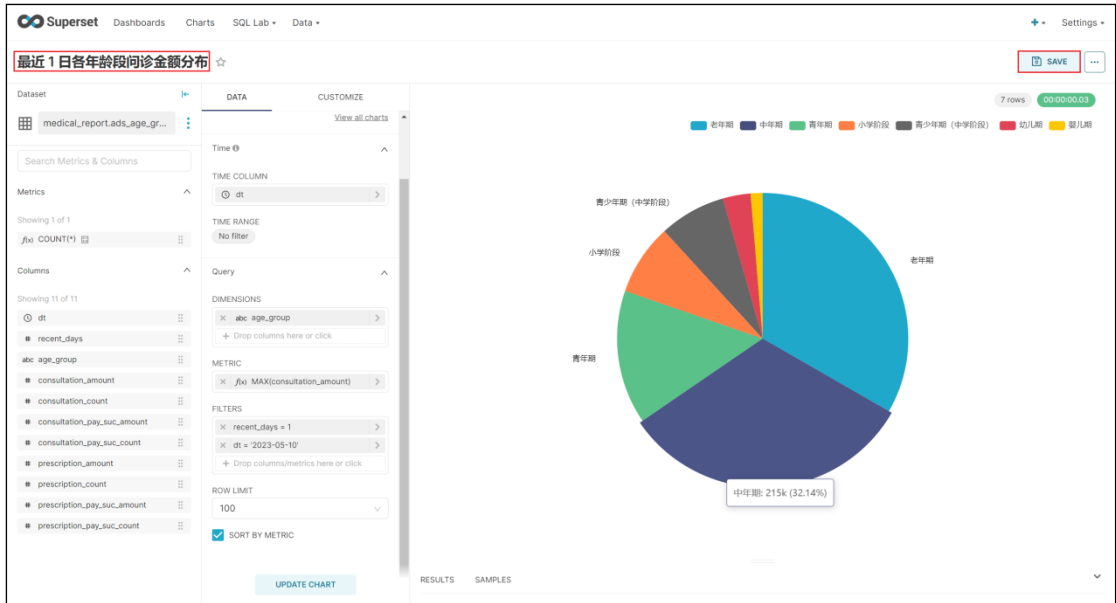
3) 按照说明配置图表并创建



4) 如配置无误，可出现以下图表

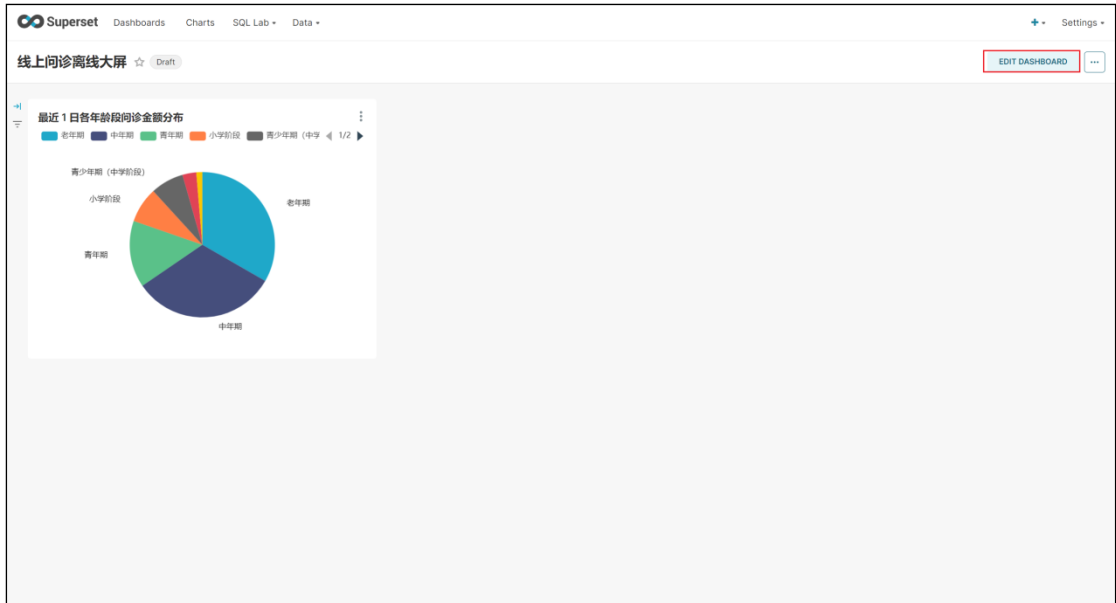


5) 命名该图表，并保存至仪表盘

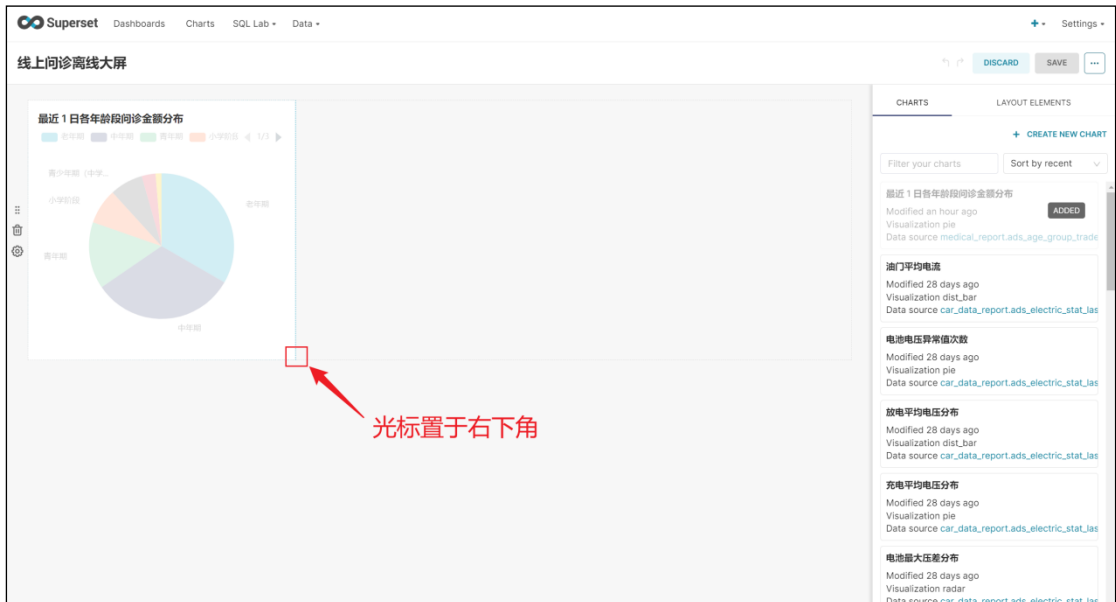


3.2.3 编辑仪表盘

1) 打开仪表盘，点击编辑按钮



2) 调整图表大小以及图表盘布局



3) 点击下图中箭头，可调整仪表盘自动刷新时间

Superset

DashboardsChartsSQL LabData

Settings

线上问诊离线大屏

最近 1 日各年龄段问诊金额分布

青少年期 (中学...)

小学阶段

青年期

中年期

老年期

DISCARD

SAVE

CHARTS

LAYOUT ELEMENTS

Filter your charts

Sort by recent

最近 1 日各年龄段问诊

Modified an hour ago

Visualization pie

Data source medical_report.ads_age_group_trade

油门平均电流

Modified 28 days ago

Visualization dist_bar

Data source car_data_report.ads_electric_stat_las

电池电压异常值次数

Modified 28 days ago

Visualization pie

Data source car_data_report.ads_electric_stat_las

放电平均电压分布

Modified 28 days ago

Visualization dist_bar

Data source car_data_report.ads_electric_stat_las

充电平均电压分布

Modified 28 days ago

Visualization pie

Data source car_data_report.ads_electric_stat_las

电池最大压差分布

Modified 28 days ago

Visualization radar

Data source car_data_report.ads_electric_stat_las

Edit properties

Edit CSS

Save as

Share

Set filter mapping

Set auto-refresh interval

Superset

DashboardsChartsSQL LabData

Settings

线上问诊离线大屏

最近 1 日各年龄段问诊金额分布

青少年期 (中学...)

小学阶段

青年期

中年期

老年期

DISCARD

SAVE

CHARTS

LAYOUT ELEMENTS

Filter your charts

Sort by recent

最近 1 日各年龄段问诊

Modified an hour ago

Visualization pie

Data source medical_report.ads_age_group_trade

油门平均电流

Modified 28 days ago

Visualization dist_bar

Data source car_data_report.ads_electric_stat_las

电池电压异常值次数

Modified 28 days ago

Visualization pie

Data source car_data_report.ads_electric_stat_las

放电平均电压分布

Modified 28 days ago

Visualization dist_bar

Data source car_data_report.ads_electric_stat_las

充电平均电压分布

Modified 28 days ago

Visualization pie

Data source car_data_report.ads_electric_stat_las

电池最大压差分布

Modified 28 days ago

Visualization radar

Data source car_data_report.ads_electric_stat_las

Refresh interval

REFRESH FREQUENCY

10 seconds

10 seconds

10 seconds

Don't refresh

30 seconds

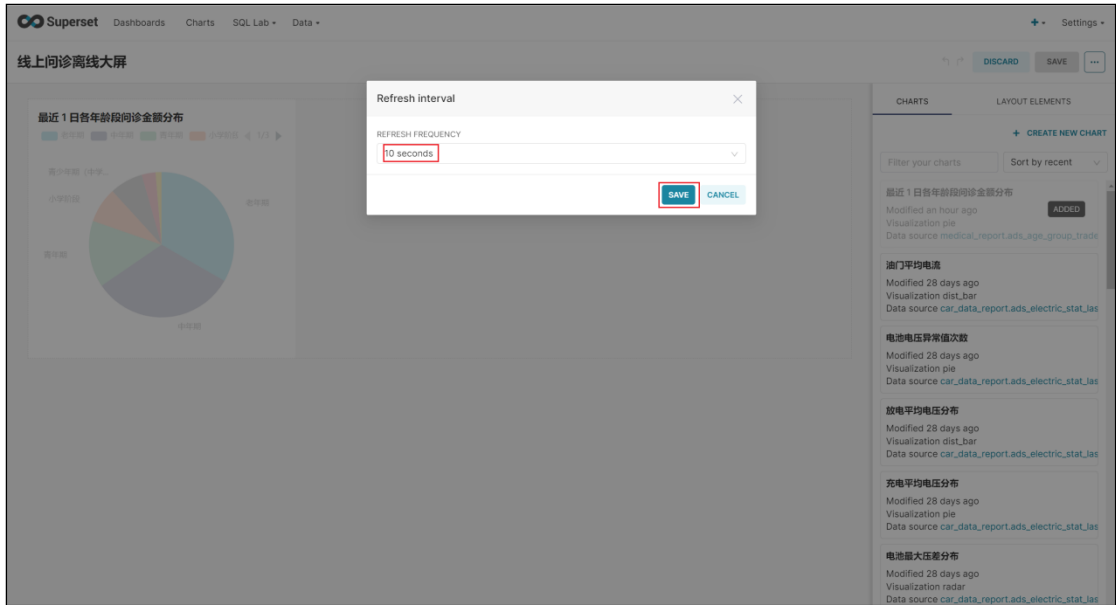
1 minute

5 minutes

30 minutes

1 hour

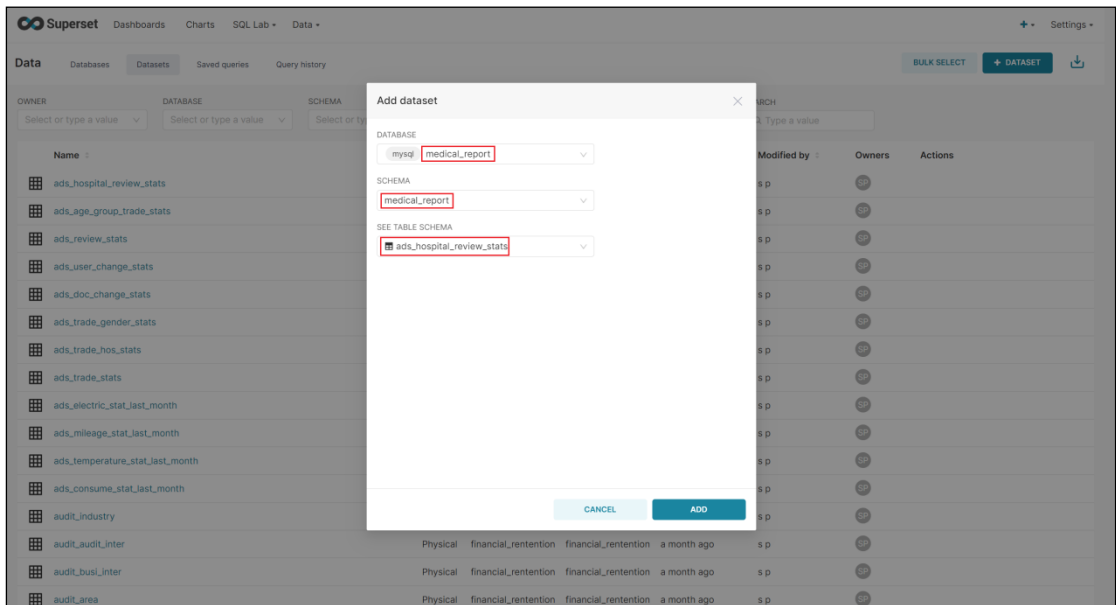
6 hours



第 4 章 Superset 实战

4.1 制作柱状图

4.1.1 配置 Table



4.1.2 配置 Chart

Superset

DashboardsChartsSQL LabData

Settings

Create a new chart

CHOOSE A DATASET

ads_hospital_review_stats

Instructions to add a dataset are available in the Superset tutorial.

CHOOSE CHART TYPE

All charts

Recommended tags

Popular

ECharts

Advanced-Analytics

Category

Correlation

Distribution

Evolution

Flow

KPI

Search all charts

215

80.7M

Big Number with Trendline

Big Number

Table

Pivot Table v2

Time-series Line Chart

Time-series Area Chart

Time-series Bar Chart v2

Time-series Scatter Plot

Pie Chart

Bar Chart

World Map

Bar Chart

AdditiveBarCategoricalComparisonDiscrete

LegacyPercentagesPopularStackedVertical

nv3

Compares metrics from different categories using bars. Bar lengths are used to indicate the magnitude of each value and

Examples

CREATE NEW CHART

Superset

DashboardsChartsSQL LabData

Settings

Add the name of the chart

SAVE

922 rows00:00:00.06

Dataset

medical_report_ads_hospita...

Search Metrics & Columns

Metrics

Showing 1 of 1

f(x) COUNT(*)

Columns

Showing 6 of 6

dt

abc_hospital_id

abc_hospital_name

review_user_count

review_count

good_review_rate

DATA

CUSTOMIZE

Time @

TIME COLUMN

dt

TIME RANGE

No filter

Query

METRICS

f(x) MN(good_review_rate)

FILTERS

dt = 2023-05-09

DIMENSIONS

abc_hospital_id

abc_hospital_name

BREAKDOWNS

ROW LIMIT

10000

SORT BY

UPDATE CHART

时间字段

度量

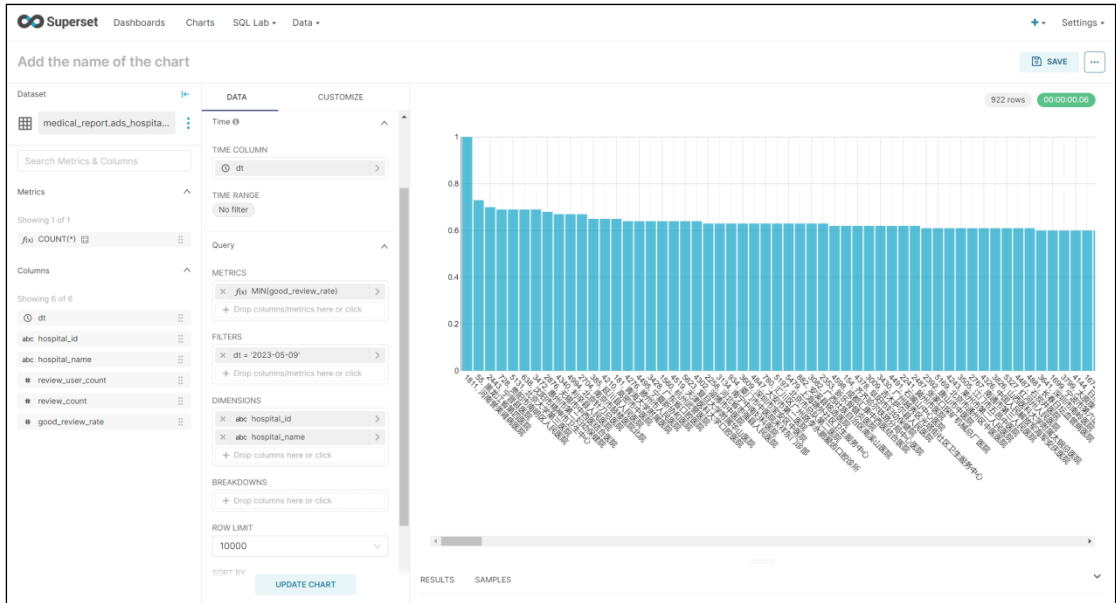
过滤条件

维度

Bar Chart

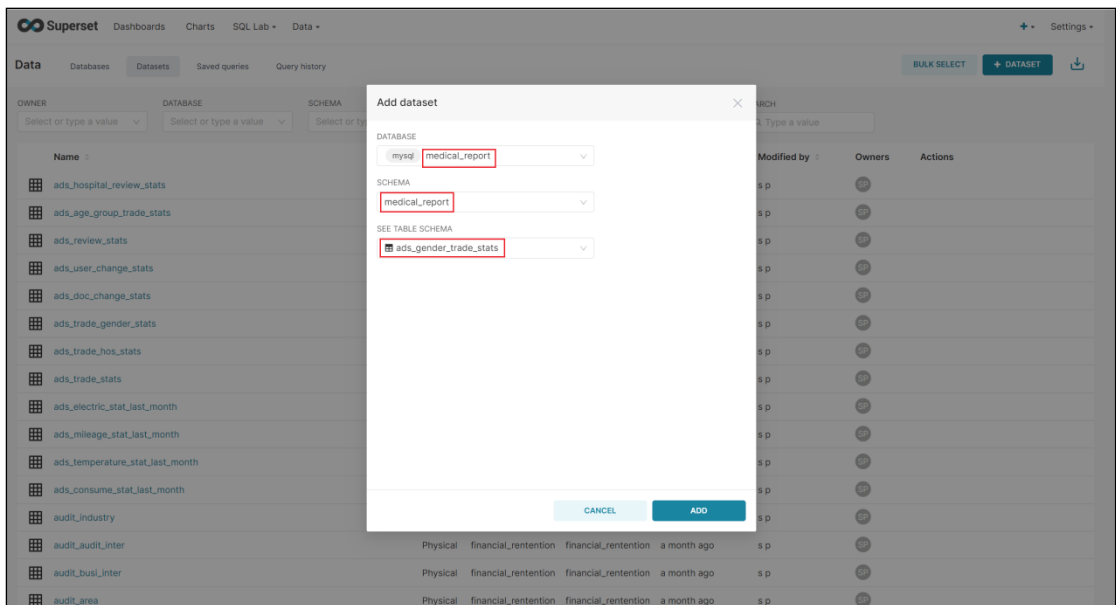
RESULTS

SAMPLES



4.2 制作饼状图

4.2.1 配置 Table



4.2.2 配置 Chart

