

Superset 使用简介

AUTHOR: 彭玲 TIME: 2022/10/31

Superset 使用简介

简介

数据源连接

PostgreSQL 连接

ClickHouse 连接

数据集

数据库表

SQL 查询

图表

数据库表

SQL 查询

看板

简介

Superset 是由 Airbnb 开源的数据分析与可视化平台。该工具主要特点是可自助分析、自定义仪表盘、分析结果可视化（导出）、用户/角色权限控制，还集成了一个 SQL 编辑器，可以进行 SQL 编辑查询对结果集进行保存可视化等。

Superset 特性如下：

- 支持几乎所有主流的数据库，包括 MySQL、PostgreSQL、Oracle、SQL Server、SQLite、SparkSQL 等，并深度支持 [Druid](#)。
- 丰富的可视化展示，支持自定义创建 Dashboard。
- 数据的展示完全可控，可自定义展示字段、聚合数据、数据源等。

数据源连接

Superset 支持丰富的数据源连接。

凡是 SQLAlchemy 支持的数据源，Superset 都支持，详见：<https://superset.apache.org/docs/databases/installing-database-drivers/>。

下面给出 PostgreSQL 和 ClickHouse 数据库连接示例。

PostgreSQL 连接

创建 `FS-PostgreSQL-pepca` 数据库连接。

Edit database ✕

POSTGRESQL

FS-PostgreSQL-pepca

BASIC ADVANCED

HOST * ⓘ PORT *

DATABASE NAME *

Copy the name of the database you are trying to connect to.

USERNAME *

PASSWORD

DISPLAY NAME *

Pick a nickname for this database to display as in Superset.

ADDITIONAL PARAMETERS

Add additional custom parameters

SSL ⓘ

CLOSE FINISH

ClickHouse 连接

创建 FS-ClickHouse-HR 数据库连接（注意本地连接端口）。

Edit database ✕

CLICKHOUSE

FS-ClickHouse-HR

BASIC ADVANCED

DISPLAY NAME *

Pick a name to help you identify this database.

SQLALCHEMY URI *

Refer to the for more information on how to structure your URI.

TEST CONNECTION

CLOSE **FINISH**

数据集

Superset 中数据集 来源为数据库表或者 SQL 查询语句。建立数据库连接后，即可创建数据集。

数据库表

通过 数据集 页面 +DATASET 按钮新增 cc_process_achievements 数据库表:

Edit Dataset cc_process_achievements

⚠ Be careful. Changing these settings will affect all charts using this dataset, including charts owned by other people.

SOURCE METRICS **1** COLUMNS **17** CALCULATED COLUMNS **0** SETTINGS

🔒 Click the lock to prevent further changes.

Physical (table or view) Virtual (SQL)

PHYSICAL ⓘ

DATABASE

postgresql FS-PostgreSQL-pepca

SCHEMA

public

TABLE

cc_process_achievements

CANCEL SAVE

SQL 查询

Superset 提供了 SQL 查询 web 界面, 通过 SQL -> SQL Lab 进入 SQL 查询界面。

在右上方 SQL 编辑器中, 输入查询语句, 点击 save dataset 将查询保存为数据集。

The screenshot shows the Apache Superset web interface. On the left, there are dropdown menus for 'DATABASE' (clickhouse FS-ClickHouse-HR), 'SCHEMA', and 'SEE TABLE SCHEMA'. The main area contains a SQL editor with the following query:

```
1 select _vacate.*, _user.name from (  
2   select  
3     hrm.vacate.pep_user_id as pepUserId,  
4     count(hrm.vacate.pep_process_story_id) as count  
5   from hrm.vacate  
6   group by hrm.vacate.pep_user_id  
7 ) as _vacate  
8 left join pepca8.user as _user on _vacate.pepUserId=_user.id
```

Below the query editor, there are buttons for 'RUN', 'LIMIT: 1 000', a timer '00:00:00.38', 'SAVE', 'COPY LINK', and a highlighted 'Save dataset' button. The 'RESULTS' section shows '255 rows returned' and a table with the following data:

| pepUserId | count | name |
|-----------|-------|------|
| 610 | 2 | 熊颖元 |
| 116 | 2 | 欧阳华强 |
| 742 | 7 | 匡晨 |
| 774 | 1 | 李伟02 |
| 874 | 2 | 金文强 |
| 650 | 4 | 张斌 |
| 526 | 1 | 曹秋红 |
| 504 | 1 | 陈淮 |

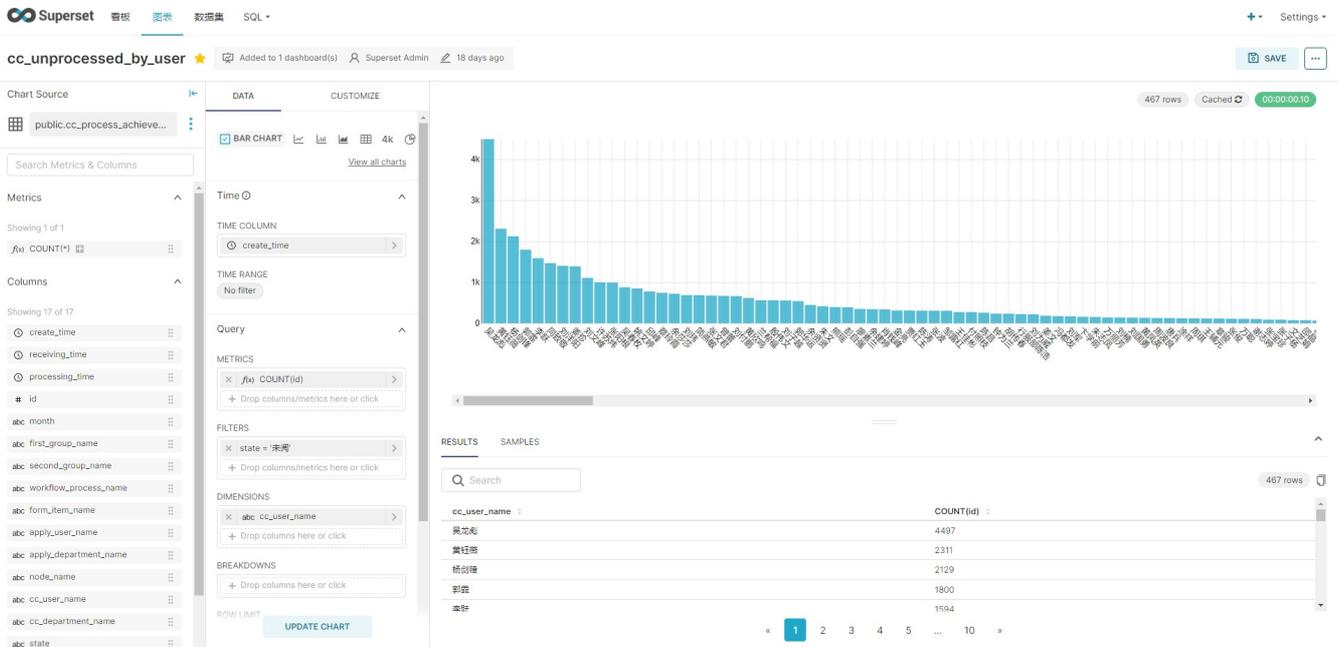
图表

数据库表

Superset 提供了丰富的图表类型来帮助我们进行数据可视化，除了常见的表格、时序图、柱状图外，还有很多其他类型，详见创建图表界面中的 [图表类型选择](#)。

Superset 中图表 是基于 [数据集](#) 进行创建的。以上面的 `cc_process_achievements` 数据集为例，对用户的抄送未处理记录数进行数据可视化：

- 过滤条件为 `state='未阅'` (数据集集中的未阅记录)
- 分组维度为数据库表字段 `cc_user_name` (抄送的用户名)
- 统计指标为 `count(id)` (按用户名分组的未读记录数)



SQL 查询

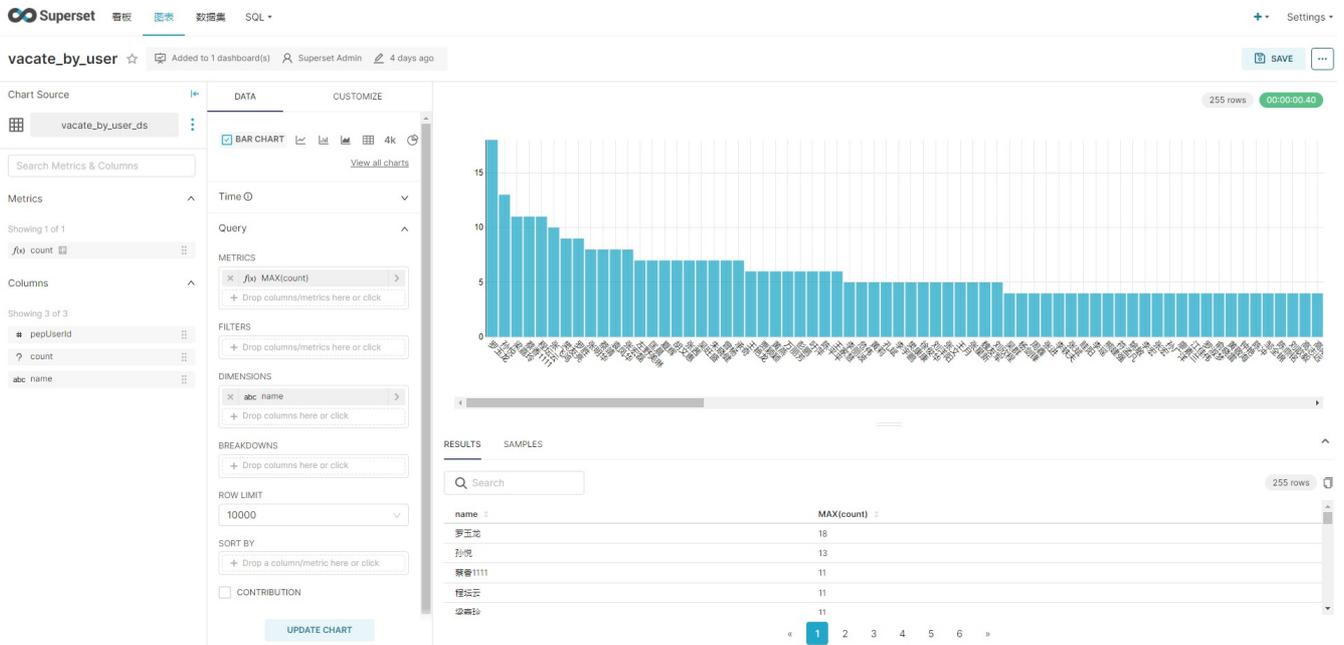
SQL 查询方式保存的数据集，默认不展示任何数据列，因此，在创建图表前需要先修改数据集，选择 COLUMNS 同步数据列，如下：

The screenshot shows the 'Edit Dataset vacate_by_user_ds' dialog. A warning message states: "Be careful. Changing these settings will affect all charts using this dataset, including charts owned by other people." The 'COLUMNS' tab is active, showing the following table:

| Column | Data type | Is temporal | Default datetime | Is filterable | Is dimension |
|-----------|-----------|--------------------------|-----------------------|-------------------------------------|-------------------------------------|
| pepUserId | INT32 | <input type="checkbox"/> | <input type="radio"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| count | UINT64 | <input type="checkbox"/> | <input type="radio"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| name | STRING | <input type="checkbox"/> | <input type="radio"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |

A red box highlights the 'SYNC COLUMNS FROM SOURCE' button. At the bottom right, there are 'CANCEL' and 'SAVE' buttons.

同步数据列后，创建图表：统计员工请假次数。



看板

Superset 提供了 Dashboard (看板) 功能，通过创建 看板可以自由组织图表，以快速、方便地查看图表，以及与图表进行交互。

