



从探索式数据分析到现代 BI 仪表盘

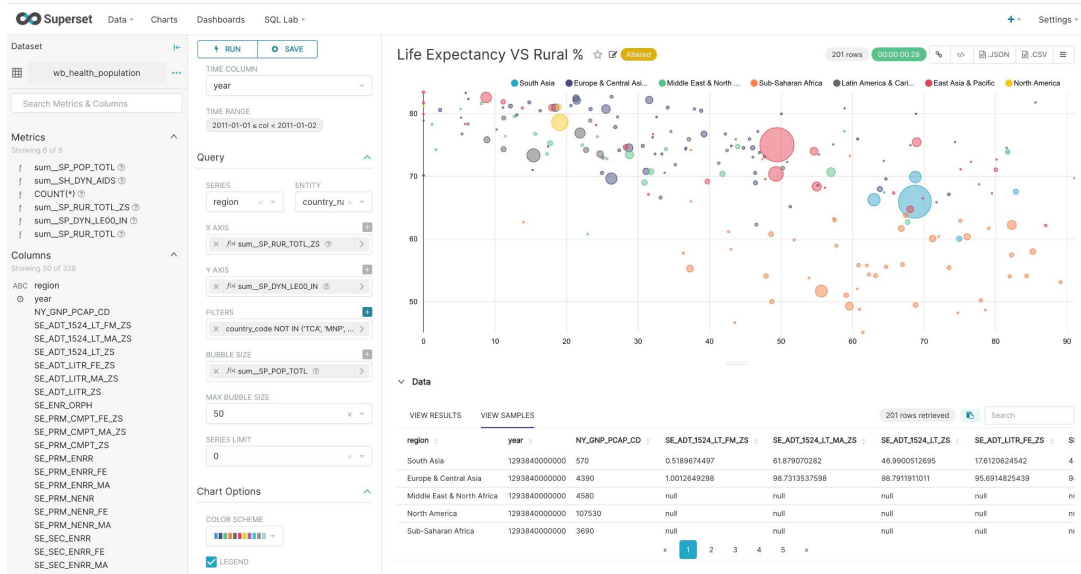
Apache Superset 2.0 介绍

2022-09-17

yongjiezhao@apache.org



Apache Superset 历史



Apache Superset 最初源于Airbnb 内部项目, 多年孵化, 现在是GitHub上 Star 最多的数据可视化项目。

- First Commit: 2015
- Version 1.0: 2022-01-18
- Version 1.4: 2022-01-28
- Version 1.5: 2022-05-17
- Version 2.0: 2022-07-19



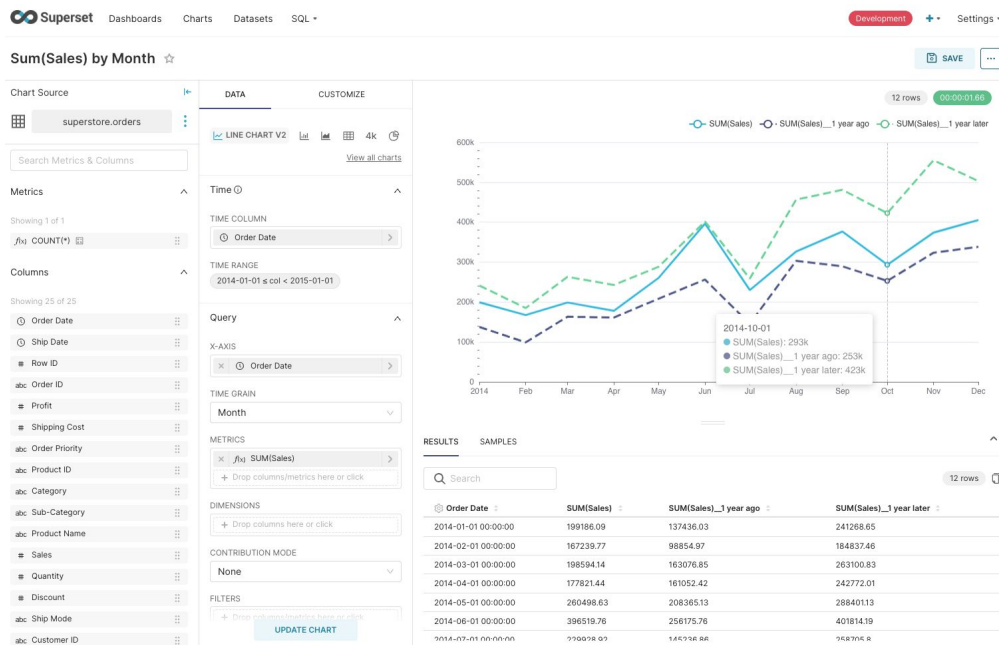
Apache Superset 功能

Apache Superset 是一个数据探索和数据可视化平台, 主要功能包括:

- 利用丰富的 Python DBAPI2 接口 (SQLAlchemy) 几乎可以支持市面上所有的关系型数据库(DB Connection).
- 全功能的, 强大的 SQL IDE(SQLLab).
- 数据探索工具可以生成 50 多种可视化图表, 依赖于前端插件系统, 可视化类型可以非常容易扩展(Explores).
- 数据探索产生的 Charts 可以生成 Dashboards. Dashboards 上可以做多种的数据筛选(Dashboards).
- 插件系统(Plugins).
- 定时报告(Reports).
- CSS 定制模版(CSS Template)
- 数据行级别权限, 根据登陆用户映射查询条件



数据探索



数据探索是一种数据分析的方法, 分析师可以用可视化手段分析数据.

Apache Superset 支持:

- 拖拽式UI设计
- 平滑切换不同可视化图形(new feature)
- 维度和指标均可以 customized SQL snippet
- SQL 预览/结果集/数据采样
- 高级时间分析: 时间平移/采样/滚动计算
- 高级时间分析: 预测分析
- 直接分析SQL 逻辑视图(来源于SQLLab)
- 注解层
- 定时发送报告
- 各种分享URL/email/embedded codes





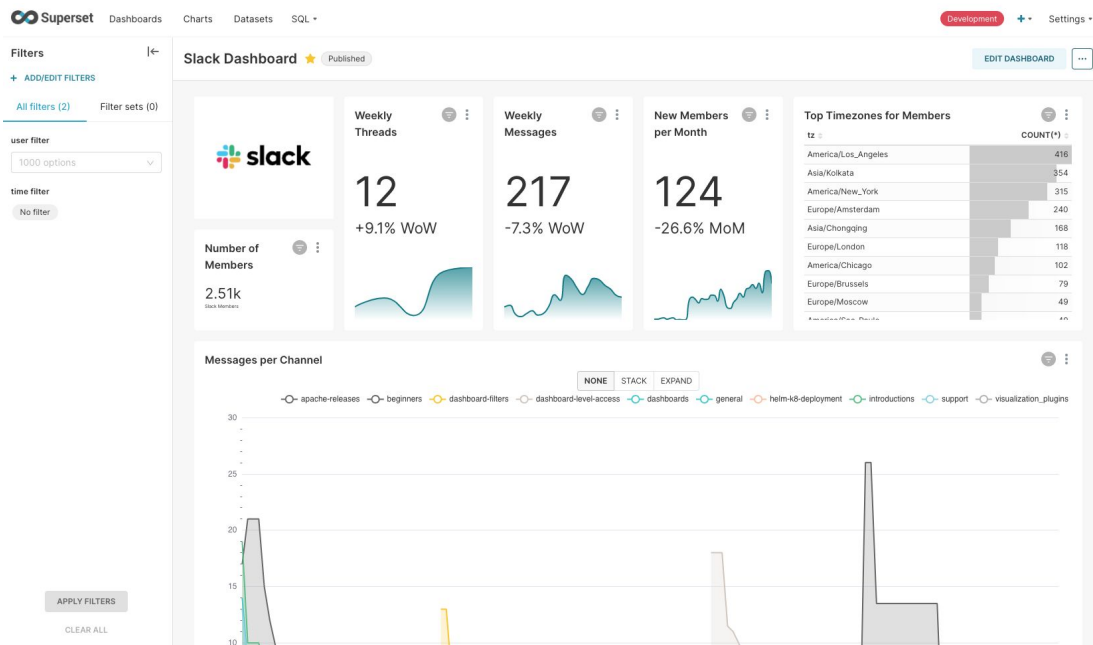
数据探索 Demo

我会使用 Superstore 数据集给大家演示:

1. 指标和维度拖拽
2. 智能X轴(Superset 2.0 new feature)
3. 时间过滤器
4. 时间移动分析
5. 数据预览和采样

请看我录制的视频

Dashboards



Charts 放入 Dashboards 后

Dashboard 可以对 Charts 做更丰富的操作.

- Native Filter - 独立于Charts 的筛选器
- 多标签
- 行和列的布局
- Markdown
- Cross Filter
- Drill Through (下钻到明细)
- 对 Charts 最大化, Charts 数据获取
- 自动刷新



Dashboards Demo

我会使用 Apache Superset 内建的 Slack Dashboard 给大家演示:

1. Native filter
2. Drill to Detail(Drill Through)
3. Charts 操作
4. 布局设计器

请看我录制的视频

Tips 1: 打开 Feature Flag 使用实验功能

```
# -----  
# Feature flags  
# -----  
# Feature flags that are set by default go here. Their values can be  
# overwritten by those specified under FEATURE_FLAGS in superset_config.py  
# For example, DEFAULT_FEATURE_FLAGS = { 'FOO': True, 'BAR': False } here  
# and FEATURE_FLAGS = { 'BAR': True, 'BAZ': True } in superset_config.py  
# will result in combined feature flags of { 'FOO': True, 'BAR': True, 'BAZ': True }  
DEFAULT_FEATURE_FLAGS: Dict[str, bool] = {  
    # allow dashboard to use sub-domains to send chart request  
    # you also need ENABLE_CORS and  
    # SUPERSET_WEBSERVER_DOMAINS for list of domains  
    "ALLOW_DASHBOARD_DOMAIN_SHARDING": True,  
    # Experimental feature introducing a client (browser) cache  
    "CLIENT_CACHE": False,  
    "DISABLE_DATASET_SOURCE_EDIT": False,  
    # When using a recent version of Druid that supports JOINS turn this on  
    "DRUID_JOINS": False,  
    "DYNAMIC_PLUGINS": False,  
    # With Superset 2.0, we are updating the default so that the legacy datasource  
    # editor no longer shows. Currently this is set to false so that the editor  
    # option does show, but we will be depreciating it.
```

1. superset/config.py
2. 查看想要设置的FF
3. `FEATURE_FLAGS: Dict[str, bool] = {}`

E.g.: `FEATURE_FLAGS: {
 "GENERIC_CHART_AXES": True,
 "DRILL_TO_DETAIL": True,
}`

Tips 2: 制作一个自定义 Viz 插件

The screenshot displays the Apache Superset web interface. On the left, the 'Datasets' panel shows the 'birth_names' dataset. Below it, the 'Metrics' and 'Columns' panels are visible. The main area is divided into two sections: 'DATA' and 'CUSTOMIZE'. In the 'CUSTOMIZE' section, the 'Chart type' is set to 'Hello World'. The 'Time' and 'Time Column' sections are also visible. The 'Query' section shows the 'Columns' panel with 'ABC name' selected. The right panel shows the 'Hello, World!' visualization, which displays a JSON array of data points. The top navigation bar includes 'Superset', 'Dashboards', 'Charts', 'SQL Lab', 'Data', and 'Settings'.

Superset Dashboards Charts SQL Lab Data Settings

Dataset: birth_names

Search Metrics & Columns

Metrics

Showing 2 of 2

- COUNT(*)
- sum_num

Columns

Showing 9 of 9

- ds
- New Col
- gender
- name
- num
- state
- num_boys
- num_girls

Chart type: Hello World

Visualization Type: Hello World

Time

Time Column: ds

Time Range: No filter

Query

Columns: ABC name

Drop columns here or click

- untitled 250 rows cached 00:00:00.13

Hello, World!

```
{
  "name": "Aaron",
  "sum_num": 453835
},
{
  "name": "Abigail",
  "sum_num": 210324
},
{
  "name": "Adam",
  "sum_num": 459794
},
{
  "name": "Adrian",
  "sum_num": 101297
},

```

Data

Ref: <https://preset.io/blog/building-custom-viz-plugins-in-superset-v2>



Thank You for Attending My Talk

1. Issues/PRs/Discussions in the GitHub Repo: <https://github.com/apache/superset>
2. Slack Community
3. Apache Mailing List
4. Preset Blog: <https://preset.io/blog/>