

1. Superset安装配置

1.0 新建X86主机

由于Superset依赖的一个组件pyarrow不支持ARM 64架构，故需要首先在华为云控制台新建一台X86架构的虚拟机，并配置 `/etc/hosts`，使其能通过主机名访问Hadoop系统各主机。

1.1 环境准备

```
# Python环境
$ apt install python3-pip

# 安装依赖包
$ apt install build-essential libssl-dev libffi-dev python3-dev libsasl2-dev libldap2-dev

# Python虚拟环境
$ apt install -y python3-venv
$ mkdir -p /opt/pyenv
$ cd /opt/pyenv
# 创建虚拟环境bdcourse
$ python3 -m venv bdcourse
# 激活虚拟环境bdcourse
$ source /opt/pyenv/bdcourse/bin/activate
# 退出虚拟环境: deactivate

# 配置pypi源为清华大学源
$ pip install -i https://pypi.tuna.tsinghua.edu.cn/simple pip -U
$ pip config set global.index-url https://pypi.tuna.tsinghua.edu.cn/simple

# 安装依赖的Python包
$ pip install wheel PyMySQL pyhive thrift_sasl pythrifthiveapi
```

1.2 安装、配置Superset

```
# 安装superset
$ pip install apache-superset
# 初始化数据库
$ superset db upgrade

# 创建管理员用户
$ export FLASK_APP=superset
$ flask fab create-admin

Username [admin]: admin
User first name [admin]: Dengcheng
User last name [user]: Yan
Email [admin@fab.org]: yanzhou@ahu.edu.cn
Password: Bdcourse2020
Repeat for confirmation: Bdcourse2020
logging was configured successfully
```

```
# 加载样例数据
$ superset load_examples

# 创建默认角色和权限
$ superset init

$ vi /opt/pyenv/bdcourse/lib/python3.6/site-packages/superset/config.py
# 修改如下内容
SUPERSET_WEBSERVER_TIMEOUT = 3600
SQLLAB_TIMEOUT = 3600
```

1.3 启动Superset

Linux系统启动的时候会启动一批守护进程（服务），早期的Linux版本一般采用 `init` 进程来启动服务，所有服务都是 `init` 进程的子进程。但是 `init` 只能串行启动，速度慢。现在一般采用Systemd，它是一组命令，对系统所有资源进行管理，系统服务（Service）只是其中一种。

注意：需要在华为云平台安全组中打开8888端口。

```
$ vi /etc/systemd/system/superset.service

[Unit]
Description=Superset

[Service]
User=root
Group=root
Type=simple
WorkingDirectory=/root/
ExecStart=/opt/pyenv/bdcourse/bin/superset run --host 0.0.0.0 --port 8888 --with-threads
StandardOutput=null
Restart=always
RestartSec=10

[Install]
WantedBy=multi-user.target

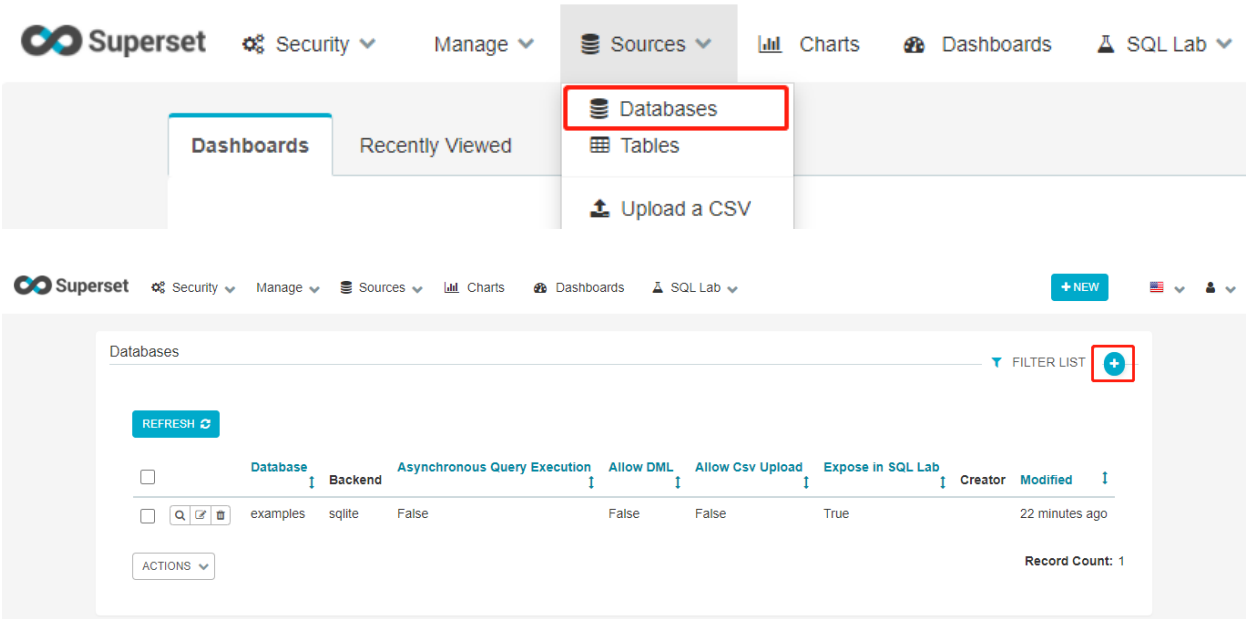
# 重新加载服务的配置文件
$ systemctl daemon-reload
# 设置开机启动 (通过在/etc/systemd/system/multi-user.target.wants/创建软连接实现)
$ systemctl enable superset
# 启动服务
$ systemctl start superset
# 查看服务状态
$ systemctl status superset
```

浏览器打开124.70.135.59:8888访问Superset。

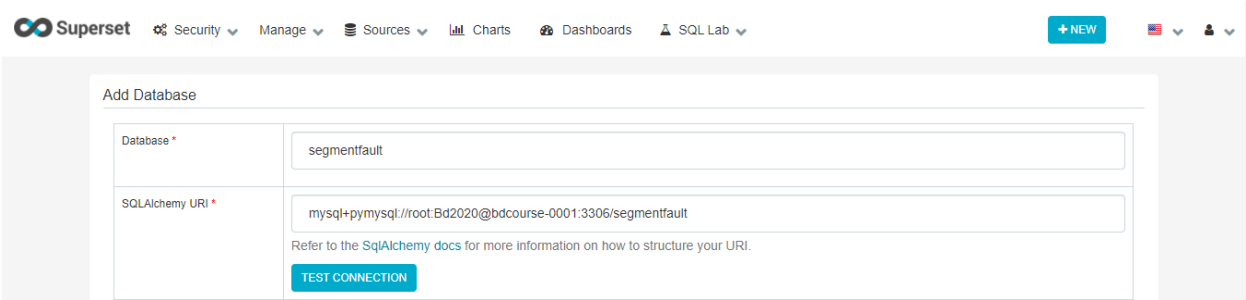
2. Superset基本操作

2.1 MySQL

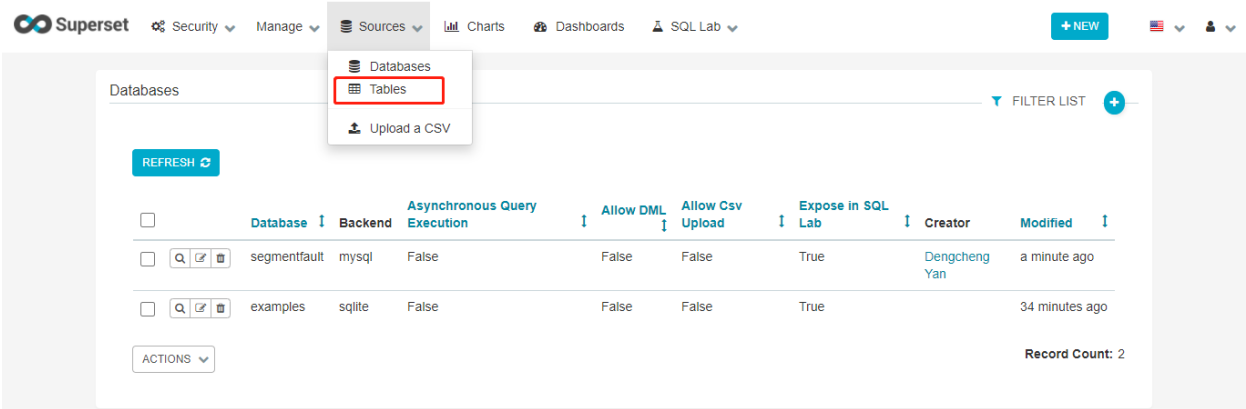
添加新数据源



填写新数据库连接信息



添加新表



Superset Security Manage Sources Charts Dashboards SQL Lab + NEW

Tables FILTER LIST

x SQL LAB VIEW Not Equal to

SEARCH

	Table	Database	Changed By	Modified
<input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> sf_population_polygons	examples		26 minutes ago
<input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> paris_iris_mapping	examples		26 minutes ago

Superset Security Manage Sources Charts Dashboards SQL Lab + NEW

Import a table definition

Database * segmentfault

Schema Schema
Schema, as used only in some databases like Postgres, Redshift and DB2

Table Name * sampleusers
Name of the table that exists in the source database

SAVE

2.2 Hive

添加Hive源

Superset Security Manage Sources Charts Dashboards SQL Lab + NEW

Add Database

Database * ghtorrent-hive

SQLAlchemy URI * hive://bdcourse-0001:10000/default?auth=NONE
Refer to the [SQLAlchemy docs](#) for more information on how to structure your URI.

TEST CONNECTION

添加Hive表

Superset Security Manage Sources Charts Dashboards SQL Lab + NEW

Import a table definition

Database * ghtorrent-hive


Schema Schema
Schema, as used only in some databases like Postgres, Redshift and DB2


Table Name * users
Name of the table that exists in the source database


SAVE


2.3 Dashboard


创建Dashboard


 Superset


 Security


 Manage


 Sources


 Charts


 Dashboards


 SQL Lab







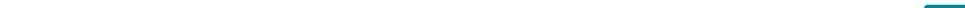





Add Dashboard

Title	<input type="text" value="DMiner"/>
Slug	<input type="text" value="Slug"/> To get a readable URL for your dashboard
Owners	<input type="text" value="x Dengcheng Yan"/> Owners is a list of users who can alter the dashboard.

新建Charts

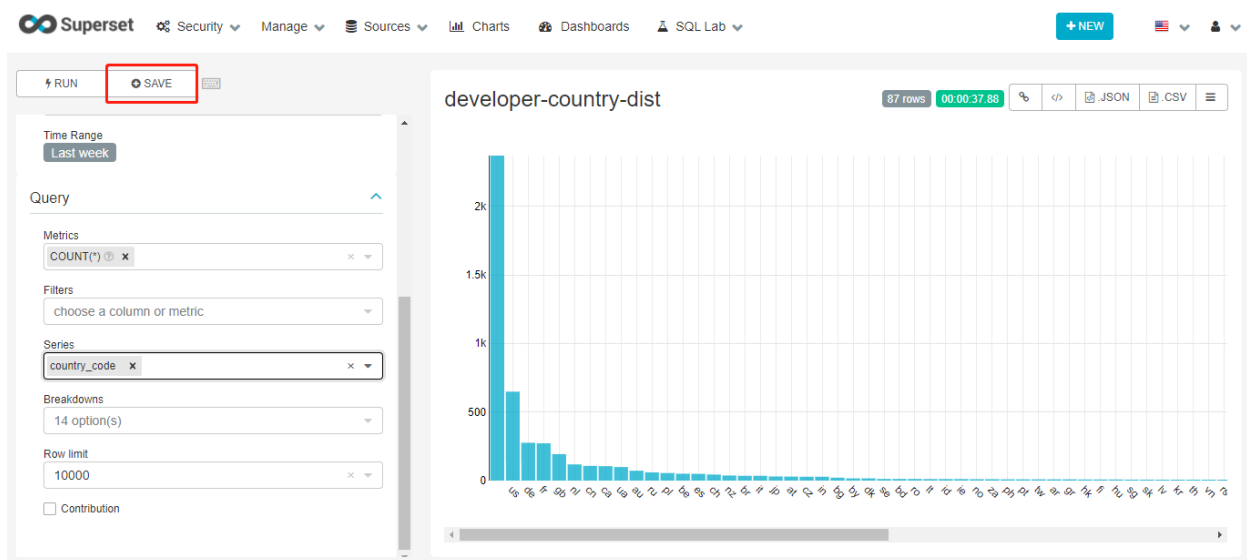


The screenshot shows the Superset web interface. The top navigation bar includes links for Security, Manage, Sources, Charts, Dashboards, and SQL Lab. A dropdown menu is open for the 'Charts' link, showing options for 'SQL Query', 'Chart' (highlighted with a red box), and 'Dashboard'. The main content area shows a 'DMiner' dashboard in 'Draft' status.



The screenshot shows the 'Create a new chart' page in Apache Superset. The top navigation bar includes the Superset logo and links for Security, Manage, Sources, Charts, Dashboards, and SQL Lab. A '+NEW' button is in the top right. The main content area has the heading 'Create a new chart'. Below it, the 'Choose a datasource' section shows a dropdown menu with 'users' selected. A note below the dropdown states: 'If the datasource you are looking for is not available in the list, follow the instructions on how to add it in the Superset tutorial.' The 'Choose a visualization type' section shows 'Bar Chart' selected. At the bottom, there is a 'CREATE NEW CHART' button.

The screenshot shows the Superset web interface. The top navigation bar includes links for Security, Manage, Sources, Charts, Dashboards, and SQL Lab, along with a '+NEW' button. The left sidebar contains several configuration sections: 'Time Range' set to 'Last week', 'Query' with a red box around the 'Metrics' field containing 'COUNT(*)', 'Filters' set to 'choose a column or metric', 'Series' with a red box around the 'country_code' field, 'Breakdowns' set to '14 option(s)', 'Row limit' set to '10000', and 'Contribution' which is unchecked. The main content area displays the query name 'developer-country-dist' and a red box around the 'RUN QUERY' button. The right sidebar shows export options: JSON, CSV, and a menu icon.



Save A Chart

Save as developer-country-dist

Do not add to a dashboard

Add chart to existing dashboard

DMiner

Add to new dashboard [dashboard name]

SAVE SAVE & GO TO DASHBOARD

