

Th.S LÊ NGỌC THẠCH

**ỨNG DỤNG
PHÂN TÍCH DỮ LIỆU
VÀ
TRÍ TUỆ NHÂN TẠO
VỚI PYTHON**

2021

Đây là eBook của riêng bạn – đề nghị không chia sẻ cho ai khác nhé!

Lời nhắn

eBook "ỨNG DỤNG PHÂN TÍCH DỮ LIỆU VÀ TRÍ TUỆ NHÂN TẠO VỚI PYTHON" này dự kiến phát hành vào tháng 12/2021. Bạn có thể đặt hàng ngay bây giờ với ưu đãi giảm 50% bằng 2 cách sau:

Cài **App MinePI** cho điện thoại tại theo link:

<https://minepi.com/thachln>

Sử dụng invitation code: **thachln**

Thử dùng điện thoại để đào Pi Coin. eBook được chấp nhận thanh toán Pi Coin với giá tương đương 399K đồng (Xem hình thức thanh toán tiền mặt).

Phiên bản bạn đang nhận là bản nháp trong quá trình hoàn thiện.

Bạn được gửi riêng để tham khảo hoặc để góp ý. Vì thế bạn được toàn quyền sử dụng và **KHÔNG** chia sẻ với bất kỳ ai khác nhé, **KHÔNG** lưu trữ trên internet nói chung để hạn chế đến tay người không thật sự cần nó!

Về nội dung bạn thu lượm được từ eBook dưới dạng các bài tóm tắt, đánh giá, hoặc đề nghị bổ sung thì rất được **KHUYẾN KHÍCH** chia sẻ công khai.



Đặc biệt khuyến khích bạn chia sẻ link:

<https://ThachLN.github.io>

Lê Ngọc Thạch

Hãy cài app [MinePI](#) ngay với Invitation Code là **thachln** để nhận ngay bản nháp (hơn 600 trang) nhé!

Hình thức thanh toán tiền mặt – Đặt hàng ngay bây với 199K, tiết kiệm 200K qua Mono hoặc Ngân hàng (dự kiến 31/12/2021 sẽ có bảng full của eBook):

① MoMo	② Chuyển khoản
<div><p>Thanh toán qua MoMo</p><p>0908550642 Lê Ngọc Thạch</p><p>Nội dung tin nhắn: email sdt Ví dụ: abc@gmail.com 0908550642 AIPYTHON Email và sdt của người nhận eBook.</p><p>Trường hợp tặng bạn bè thì ghi thông tin email và sdt của bạn.</p><p>Quét mã QR thanh toán 199K.</p><p>LE NGOC THACH 0908550642</p><p>199.000đ</p></div>	<div><p>Thanh toán qua NH Tiên Phong</p><p>Lê Ngọc Thạch, Ngân Hàng Tiên Phong, CN HCM Số tài khoản: 00002888001 Nội dung tin nhắn: email sdt AIPYTHON Vd tin nhắn: abc@gmail.com 0908456321 AIPYTHON Quét mã QR để thanh toán cho:</p><p>LE NGOC THACH 0000 2888 001</p><p>Quét mã vạch này để giao dịch</p></div>

Mục lục

Quy ước.....	7
Ngày 1 – Chủ đề: Giới thiệu về ngôn ngữ thống kê, ngôn ngữ lập trình	11
Bài 1: Tóm tắt về thống kê (Statistics).....	13
Bài 2: Ngôn ngữ lập trình Python.....	20
Bài 3: Ngôn ngữ Python và phần mềm Anaconda	28
Bài 4: Cài đặt thêm phần mềm.....	50
Bài 5: Nhập liệu, biên tập, lưu trữ dữ liệu với Python.....	55
Bài tập ngày 1	73
Thử thách cho bạn!.....	75
Ngày 2 – Chủ đề: Biểu đồ	76
Bài 6: Các loại biểu đồ.....	78
Bài 7: Vẽ biểu đồ trong Python	84
Bài 8: Nguyên tắc soạn biểu đồ	101
Bài 9: Giới thiệu Matplotlib	103
Bài 10: Giới thiệu Bokeh.....	129
Bài 11: Khai phá Bokeh.....	137
Ngày 3 – Phân tích mô tả.....	163
Bài 12: Phân tích mô tả dữ liệu Bank Marketing	165
Bài 13: Phân tích dữ liệu Marketing #2.....	178
Bài 14: So sánh 2 tỉ lệ	189
Bài 15: Mô hình kiểm định giả thuyết.....	200
Bài 16: Ứng dụng minh họa kiểm định giả thuyết	201
Bài 17: Phân tích mối tương quan	211
Ngày 4 – Chủ đề: Dữ liệu lớn	219
Bài 18: Cách xử lý tập hợp dữ liệu lớn.....	220
Bài 19: Sử dụng Ubuntu	257
Bài 20: Cài đặt Hadoop 3.2.....	266
Bài 21: Trải nghiệm Hadoop với Python.....	274
Ngày 5 – Chủ đề: Dự báo bằng mô hình hồi qui tuyến tính	280

Đây là eBook của riêng bạn – đề nghị không chia sẻ cho ai khác nhé!

Bài 22: Giới thiệu mô hình hồi qui tuyến tính.....	281
Bài 23: Diễn giải mô hình hồi qui tuyến tính.....	286
Bài 24: Mô hình hồi qui tuyến tính đa biến.....	299
Bài 25: Dự báo bằng mô hình hồi qui tuyến tính	305
Ngày 6 – Chủ đề: Dự báo bằng mô hình hồi qui logistic.....	309
Bài 26: Giới thiệu mô hình hồi qui logistic.....	310
Bài 27: Mô hình hồi qui logistic đa biến (Multiple logistic regression model).....	314
Bài 28: So sánh mô hình.....	324
Bài 29: Dự báo bằng mô hình hồi qui logistic	331
Ngày 7 – Chủ đề: Phân tích đa biến.....	340
Bài 30: Xử lý giá trị trống	341
Bài 31: Mô hình phân tích phân định (Linear discriminant analysis)..	345
Bài 32: Mô hình thành phần (Principal Component Analysis).....	353
Bài 33: Mô hình phân tích cụm/nhóm (cluster analysis).....	361
Ngày 8 – Chủ đề: Machine Learning.....	369
Bài 34: Giới thiệu Machine learning.....	370
Bài 35: Mô hình SVM.....	372
Bài 36: Mô hình Random Forest	380
Bài 37: Mô hình Artificial Neural Network.....	384
Bài 38: Machine Learning với Python Tensorflow.....	390
Ngày 9 – Chủ đề: Recommendation	419
Bài 39: Giới thiệu phương pháp gợi ý Collaborative filtering.....	420
Bài 40: Triển khai phương pháp gợi ý Collaborative filtering bằng R	430
Ngày 10 – Chủ đề: Natural Language Processing.....	436
Bài 41: Các kỹ thuật cơ bản	437
Bài 42: Trích đặc trưng (Feature extraction).....	442
Bài 43: Giới thiệu ứng dụng phân tích cảm xúc (Sentiment Analysis).....	452
Bài 44: Giới thiệu ứng dụng phân tích từ vựng (Word Embedding) ...	463
Bài 45: Giới thiệu ứng dụng xác định chủ đề (Topic Modeling)	475
Ngày 11 – Chủ đề: Computer Vision.....	486
Bài 46: Giới thiệu Face recognition	487

Đây là eBook của riêng bạn – đề nghị không chia sẻ cho ai khác nhé!

Bài 47: Giới thiệu mô hình CNN.....	502
Ngày 12 – Chủ đề: Nhận diện tiếng nói (Speech Recognition).....	521
Bài 48: Giới thiệu đặc trưng của âm thanh.....	522
Bài 49: Các thao tác cơ bản với file âm thanh.....	528
Bài 50: Mô hình Chuyển giọng nói thành văn bản.....	532
Ngày 13 – Chủ đề: Phân tích dữ liệu theo trường phái Bayes.....	535
Bài 51: Gia nhập trường phái Bayes.....	536
Học liệu tham khảo:	541
Tạm kết thúc.....	542
Phụ lục	543
Đóng gói chương trình Python.....	544
Sử dụng Flask.....	546
Cài đặt nginx proxy cho Flask API	550
Quan sát giao dịch cổ phiếu VNM (Vinamilk).....	552
Đọc và vẽ tín hiệu âm thanh.....	562
Tải sách nói “Từ tốt đến vĩ đại”.....	565
Đọc ảnh y khoa DiCOM.....	568
Áp dụng biến đổi Fourier cho ảnh.....	571
Sử dụng Git	575
Khảo sát ảnh và ma trận.....	604
Phát triển ứng dụng với Python	606
Xử lý file pdf.....	613
Khảo sát file âm thanh	621
Phân tích âm thanh với thư viện mutagen.....	624
Khám phá Python trong WSL2	625
Crawl dữ liệu bằng Selenium.....	627
Sử dụng OpenCV để phân tích dữ liệu ảnh.....	628
Cài đặt OpenCV.....	629
Sử dụng OpenCV để phát hiện đối tượng trong video	630
Đóng gói chương trình Python.....	631
Tải file video từ Youtube.....	634
Sinh code Restful API từ database.....	635

Đây là eBook của riêng bạn – đề nghị không chia sẻ cho ai khác nhé!

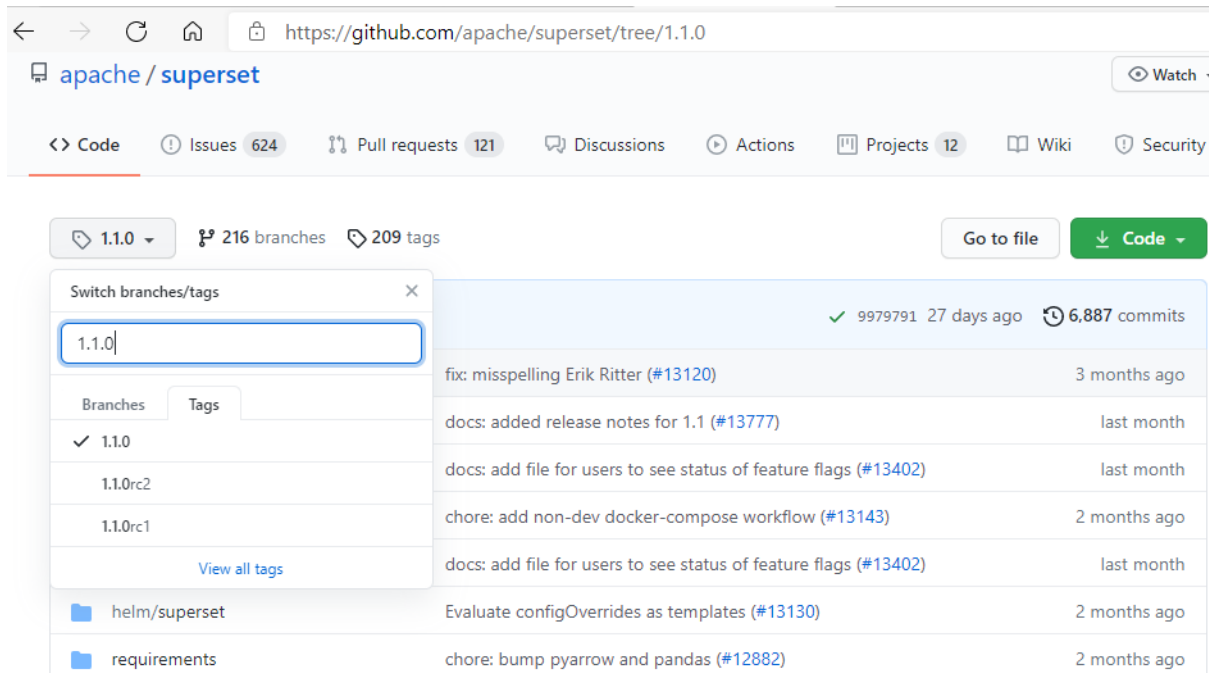
Trải nghiệm Restful API với Flask	636
Trải nghiệm Kafka.....	638
Trải nghiệm Apache NiFI.....	641
Trải nghiệm Prefect.....	646
Giới thiệu superset.....	647
Cài đặt	648
Khởi động lại superset	654
Cài đặt và cấu hình Nginx	656
Truy cập nginx trên Ubuntu từ máy Windows	657
Khai thác superset.....	659
Cài đặt service cho superset.....	663
Cài đặt Ubuntu Server 20.04.2.....	669
Superset introduction	670
Giới thiệu Airflow	671
Cài đặt	672
Giới thiệu PowerBI.....	674
Bài 1 – Phân tích mô tả với Power BI.....	675
Phân tích biểu đồ với dữ liệu Bank Marketing	687
Minh họa biểu đồ theo thời gian.....	697
Bài bổ sung.....	700
Bài 101: Phát triển ứng dụng với PyCharm	701
Bài 102: Xử lý dữ liệu.....	707
Bài 103: Sử dụng thư viện XGBoost	712
Bài 104: Sử dụng môi trường ảo cho Python.....	713
Bài 105: Sửa code.....	716
Bài 106: Sử dụng Swagger	717
Bài 107: Viết module cho Python	718

Giới thiệu superset

Bài này giúp bạn khai thác dự án <https://github.com/apache/superset> để làm Data Visualization.

Cài đặt

Dưới đây là tóm tắt quá trình tôi cài đặt superset từ source code phiên bản 1.1.0 trên <https://github.com/apache/superset>, tag 1.1.0.



Tài liệu tham khảo chính thức ở đây:

<https://github.com/apache/superset/blob/master/CONTRIBUTING.md#fask-server>

Tuy nhiên trong quá trình cài đặt có phát sinh vài lỗi và phải giải quyết bằng cách bổ sung thêm vài thư viện (được **bôi vàng**).

Yêu cầu

Ubuntu 20: Nếu bạn định dùng máy ảo trên Windows để cài Ubuntu thì tham khảo [Bài 14 trong Ngày 4](#).

Cập nhật hệ thống

```
sudo apt update
```

Kiểm tra và cài đặt Python

Mặc định khi cài Ubuntu thì Python phiên bản 3 được cài. Kiểm tra lại bằng lệnh sau:

```
python3 -V
```

```
Python 3.8.5
```


Nâng cấp python (không bắt buộc)

Trong trường hợp bạn dùng Ubuntu thì có khả năng Python được cài là 3.6.x. Bạn cần nâng cấp lên Python 3.8.x. Phiên bản Python tại thời điểm phần này được viết là 3.8.10. Hãy làm theo tài liệu sau:

<https://linuxize.com/post/how-to-install-python-3-8-on-ubuntu-18-04/>

Tiếp tục thực hiện các lệnh sau để backup file /usr/bin/python3 và ánh xạ python3.8 thành python3.

```
sudo cd /usr/bin
sudo mv ./python3 ./python3.bak
sudo ln -nsf ./python3.8 python3
```

Bổ sung thư viện

```
sudo apt install git
sudo apt install python3-virtualenv
```

Cài đặt back-end

Thực hiện các lệnh sau:

```
mkdir ~/projects && cd ~/projects

echo Checkout source code version 1.1.0
git clone https://github.com/apache/superset
cd superset
git checkout -b 1.2.0

export VENV_NAME=venv-superset
echo Create a new virtual environment with name: $VENV_NAME
virtualenv $VENV_NAME
source $VENV_NAME/bin/activate

# Upgrade pip
sudo apt install python3-pip
python3 -m pip install --upgrade pip
sudo apt-get install build-essential libmysqlclient-dev
python3-dev python3-dev libsasl2-dev
pip3 install -r requirements/local.txt
```

```
pip3 install -e .
superset fab create-admin
superset db upgrade
superset init
superset load-examples
FLASK_ENV=development superset run -p 8088 --with-threads --
reload --debugger
```

Nếu bạn dùng Ubuntu 18 thì có thể thay dòng màu đỏ ở trên bằng 2 lệnh sau:

```
pip3 install virtualenv
sudo virtualenv $VENV NAME
sudo apt-get install python-dev python3-dev libpython3.8-dev
```

Lệnh cuối cùng ở trên là để khởi động superset với port 8088. Bạn có thể truy cập <http://localhost:8088> ngay trong Ubuntu. Tuy nhiên lúc này superset chỉ có phần backend nên sẽ không có giao diện đầy đủ. Hãy nhấn Ctrl + C để tắt chương trình superset và chuyển sang phần [Cài đặt front-end](#) bên dưới.

Ghi chú

Nếu vì lý do nào đó mà quá trình làm việc của bạn bị gián đoạn, phải login lại thì thực hiện các lệnh sau sau khi login để kích hoạt môi trường ảo của Python:

```
cd ~/projects/superset
export VENV_NAME=venv-superset
source $VENV_NAME/bin/activate
```

Cài đặt front-end

Chuẩn bị

- NodeJS
- npm

Một trong các cách chuẩn bị NodeJS và npm là dùng nvm (<https://github.com/nvm-sh/nvm>) với lệnh sau:

```
sudo apt install curl
curl -o- https://raw.githubusercontent.com/nvm-sh/nvm/v0.37.0/install.sh | bash
```

Logout và login lại. Sau đó chạy lại các lệnh sau:

```
cd ~/projects/superset
export VENV_NAME=venv-superset
source $VENV_NAME/bin/activate
```

Chuẩn bị môi trường nodejs và npm:

```
cd superset-frontend
npm install
npm use
```

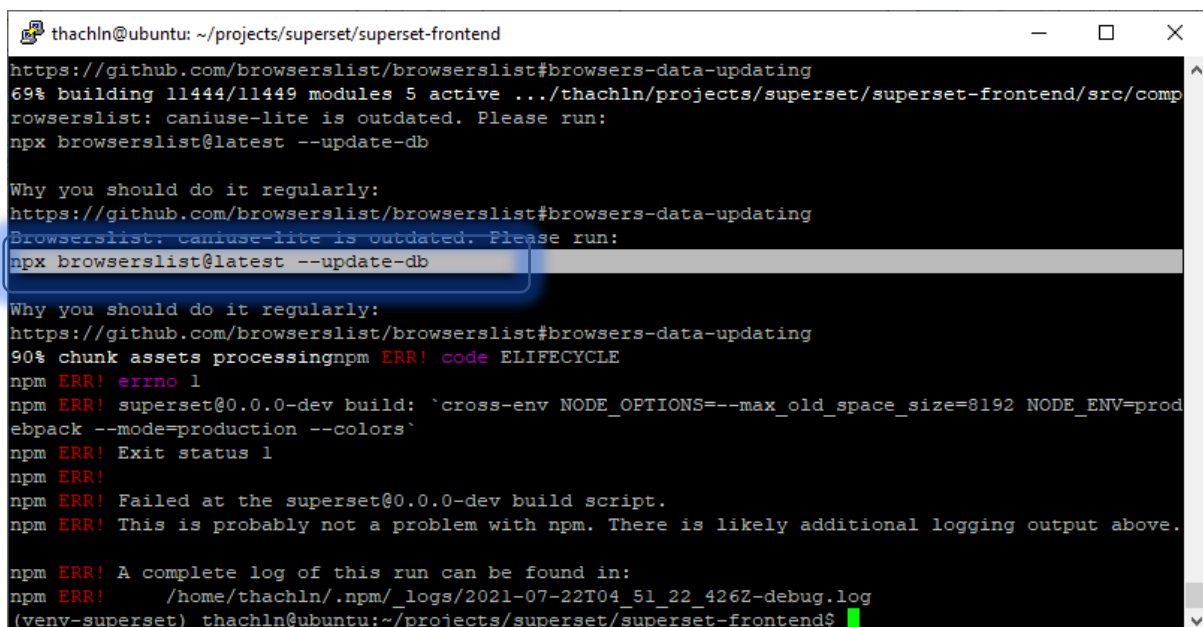
Cài đặt thư viện

```
cd ~/projects/superset/superset-frontend
# Install dependencies from `package-lock.json`
npm ci
```

Build assets

```
npm run build
```

Có thể bạn gặp lỗi như bên dưới: phán đoán là vì lý do nào đó mà gói [superset@0.0.0-dev](https://github.com/superset/superset/releases/tag/0.0.0-dev) không được download.



```
thachln@ubuntu: ~/projects/superset/superset-frontend
https://github.com/browserslist/browserslist#browsers-data-updating
69% building 11444/11449 modules 5 active .../thachln/projects/superset/superset-frontend/src/comp
rowserslist: caniuse-lite is outdated. Please run:
npx browserslist@latest --update-db

Why you should do it regularly:
https://github.com/browserslist/browserslist#browsers-data-updating
Browserslist: caniuse-lite is outdated. Please run:
npx browserslist@latest --update-db

Why you should do it regularly:
https://github.com/browserslist/browserslist#browsers-data-updating
90% chunk assets processingnpm ERR! code ELIFECYCLE
npm ERR! errno 1
npm ERR! superset@0.0.0-dev build: `cross-env NODE_OPTIONS=--max_old_space_size=8192 NODE_ENV=prod
ebpack --mode=production --colors`
npm ERR! Exit status 1
npm ERR!
npm ERR! Failed at the superset@0.0.0-dev build script.
npm ERR! This is probably not a problem with npm. There is likely additional logging output above.

npm ERR! A complete log of this run can be found in:
npm ERR!   /home/thachln/.npm/_logs/2021-07-22T04_51_22_426Z-debug.log
(venv-superset) thachln@ubuntu:~/projects/superset/superset-frontend$
```

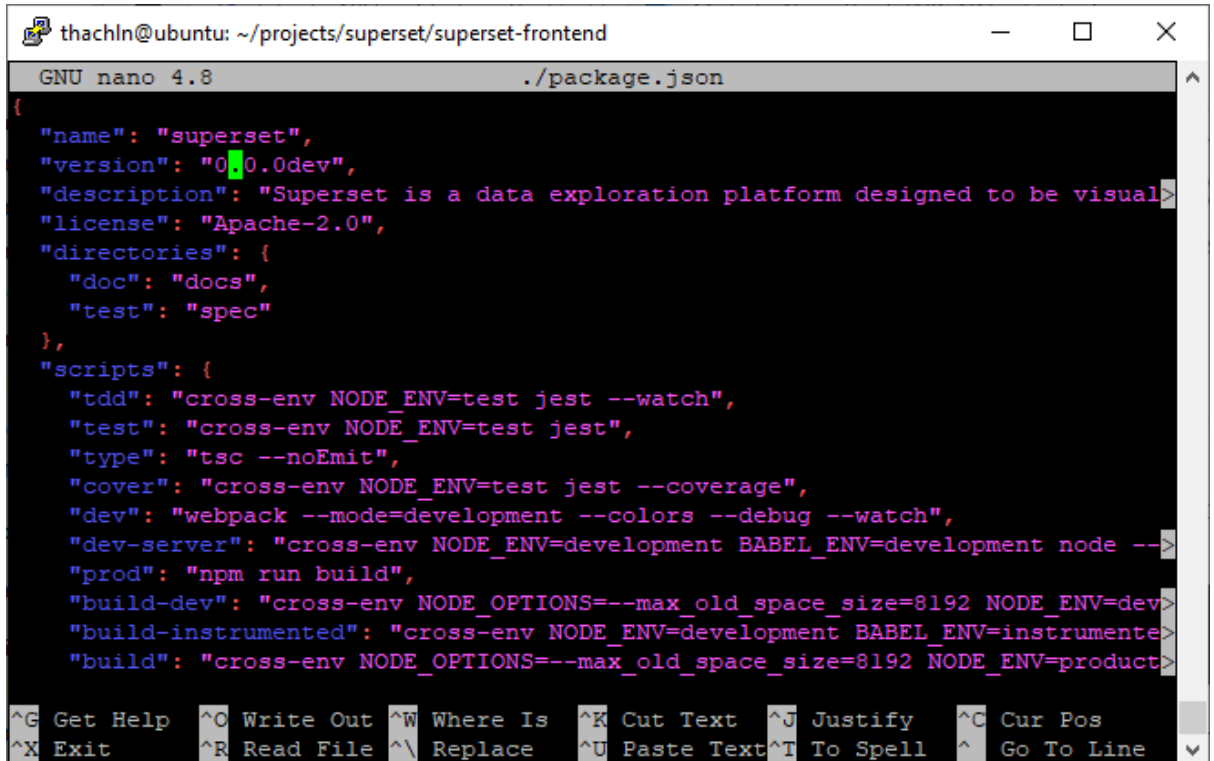
Thử khắc phục theo cách sau:

npx browserslist@latest --update-db

- Mở file package.json bằng lệnh:

```
nano ./package.json
```

Sửa lại version của superset 0.0.0dev thành 0.999.0dev



```
thachln@ubuntu: ~/projects/superset/superset-frontend
GNU nano 4.8 ./package.json
{
  "name": "superset",
  "version": "0.999.0dev",
  "description": "Superset is a data exploration platform designed to be visual",
  "license": "Apache-2.0",
  "directories": {
    "doc": "docs",
    "test": "spec"
  },
  "scripts": {
    "tdd": "cross-env NODE_ENV=test jest --watch",
    "test": "cross-env NODE_ENV=test jest",
    "type": "tsc --noEmit",
    "cover": "cross-env NODE_ENV=test jest --coverage",
    "dev": "webpack --mode=development --colors --debug --watch",
    "dev-server": "cross-env NODE_ENV=development BABEL_ENV=development node --",
    "prod": "npm run build",
    "build-dev": "cross-env NODE_OPTIONS=--max_old_space_size=8192 NODE_ENV=dev",
    "build-instrumented": "cross-env NODE_ENV=development BABEL_ENV=instrumente",
    "build": "cross-env NODE_OPTIONS=--max_old_space_size=8192 NODE_ENV=product"
  }
}
```

Sau đó build lại module frontend:

```
npm ci
npm run build
```

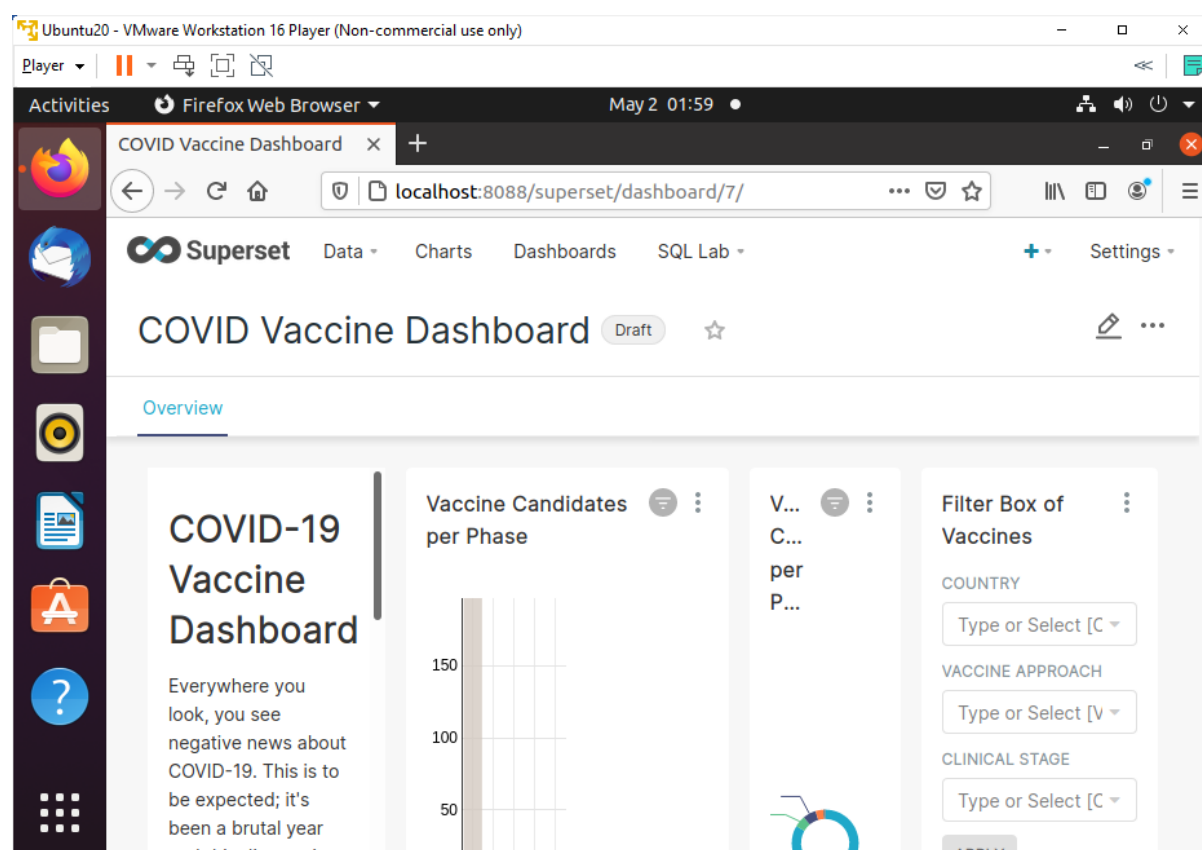
Sau khi biên dịch xong hãy khởi động lại superset bằng lệnh sau:

```
FLASK_ENV=development superset run -p 8088 --with-threads --
reload --debugger
```

Chú ý 2 dấu gạch gạch -- khi copy từ tài liệu thì có thể bị thay thế bằng 1 dấu gạch dài, hãy tự gỡ lại cho chắc.

```
thachln@ubuntu: ~/projects/superset
* Environment: development
* Debug mode: on
2021-05-02 01:56:00,281:INFO:werkzeug: * Running on http://127.0.0.1:8088/ (
Press CTRL+C to quit)
2021-05-02 01:56:00,283:INFO:werkzeug: * Restarting with stat
2021-05-02 01:56:04,622:INFO:superset.utils.logging_configurator:logging was
configured successfully
2021-05-02 01:56:04,640:INFO:root:Configured event logger of type <class 'su
perset.utils.log.DBEventLogger'>
/home/thachln/projects/superset/venv-superset/lib/python3.8/site-packages/fl
ask_caching/__init__.py:191: UserWarning: Flask-Caching: CACHE_TYPE is set t
o null, caching is effectively disabled.
  warnings.warn(
2021-05-02 01:56:08,547:WARNING:werkzeug: * Debugger is active!
2021-05-02 01:56:08,549:INFO:werkzeug: * Debugger PIN: 206-426-709
```

Nếu Ubuntu của bạn cài đặt ở chế độ có giao diện thì hãy đăng nhập và trải nghiệm superset với link <http://localhost:8088> ngay trên Ubuntu. Thông tin đăng nhập do bạn thiết lập trong quá trình cài đặt back-end.



Khởi động lại superset

Vì tôi chạy Ubuntu với VMware trên Windows khi cần sử dụng superset thì bật máy lên và thực thi các lệnh sau:

```
cd ~/projects/superset
export VENV_NAME=venv-superset
source $VENV_NAME/bin/activate
FLASK_ENV=development superset run -p 8088 --with-threads --
reload --debugger
```

Hoặc chạy với chế độ production

```
cd ~/projects/superset
export VENV_NAME=venv-superset
source $VENV_NAME/bin/activate
FLASK_ENV=production superset run -p 8088 --with-threads
```

Hoặc

```
cd ~/projects/superset
export VENV_NAME=venv-superset
source $VENV_NAME/bin/activate
export FLASK_ENV=production
superset run -p 8088 --with-threads
```

Bạn có thể tạo script để khởi động bằng cách tạo file .sh với nội dung trên như sau:

```
cd ~
./start-superset.sh
sudo nano start-superset.sh
```

Copy nội dung ở trên vào file rồi lưu và thoát ứng dụng nano.

Tiếp theo thêm thuộc tính executable cho file.

```
chmod +x ./start-superset.sh
```

Mỗi lần sử dụng thì chạy lệnh sau sau khi login:

```
./start-superset.sh
```

Chương trình superset sẽ khởi động và đứng chờ ở cửa sổ terminal.

Nếu bạn muốn khởi động superset và không ở chế độ đứng chờ thì thực hiện lệnh sau:

```
nohup ./start-superset.sh &
```

Cài đặt và cấu hình Nginx

Phần trước bạn đã biết cách cài đặt và khai thác superset qua URL <http://localhost:8088> ngay trên máy Ubuntu đã cài superset. Tình huống đặt ra cho bạn là:

① Nếu bạn muốn truy ứng dụng superset thông qua web server như nginx với URL như <http://localhost> thì phải làm sao?

② Mở rộng một chút ý ở trên: Nếu bạn muốn truy cập từ một máy khác thông qua địa chỉ IP của máy đang chạy Ubuntu thì làm sao? Cụ thể trong trường hợp này là nếu dùng trình duyệt từ máy Windows để truy cập vào superset trên máy chạy Ubuntu thì làm sao?

Một trong cách giải quyết cho hai tình huống trên là sử dụng Nginx (như đã gợi ý trong tình huống ①).

Cài đặt Nginx

Chạy lệnh sau trong Ubuntu:

```
sudo apt install nginx
```

Cấu hình Nginx

Chỉnh sửa file cấu hình mặc định của nginx bằng lệnh sau:

```
sudo nano /etc/nginx/sites-available/default
```

```
#         location / {
#             # First attempt to serve request as file, then
#             # as directory, then fall back to displaying a 404.
#             try_files $uri $uri/ =404;
#         }
#         location / {
#             proxy_pass http://127.0.0.1:8088;
#             include /etc/nginx/proxy_params;
#         }
```

Phần **màu đỏ** là chú thích lại (thêm dấu # phía trước mỗi dòng).

Phần **màu xanh** là thêm mới.

Nhấn phím Ctrl+O để lưu file. Ctrl+X để thoát trình soạn thảo nano.

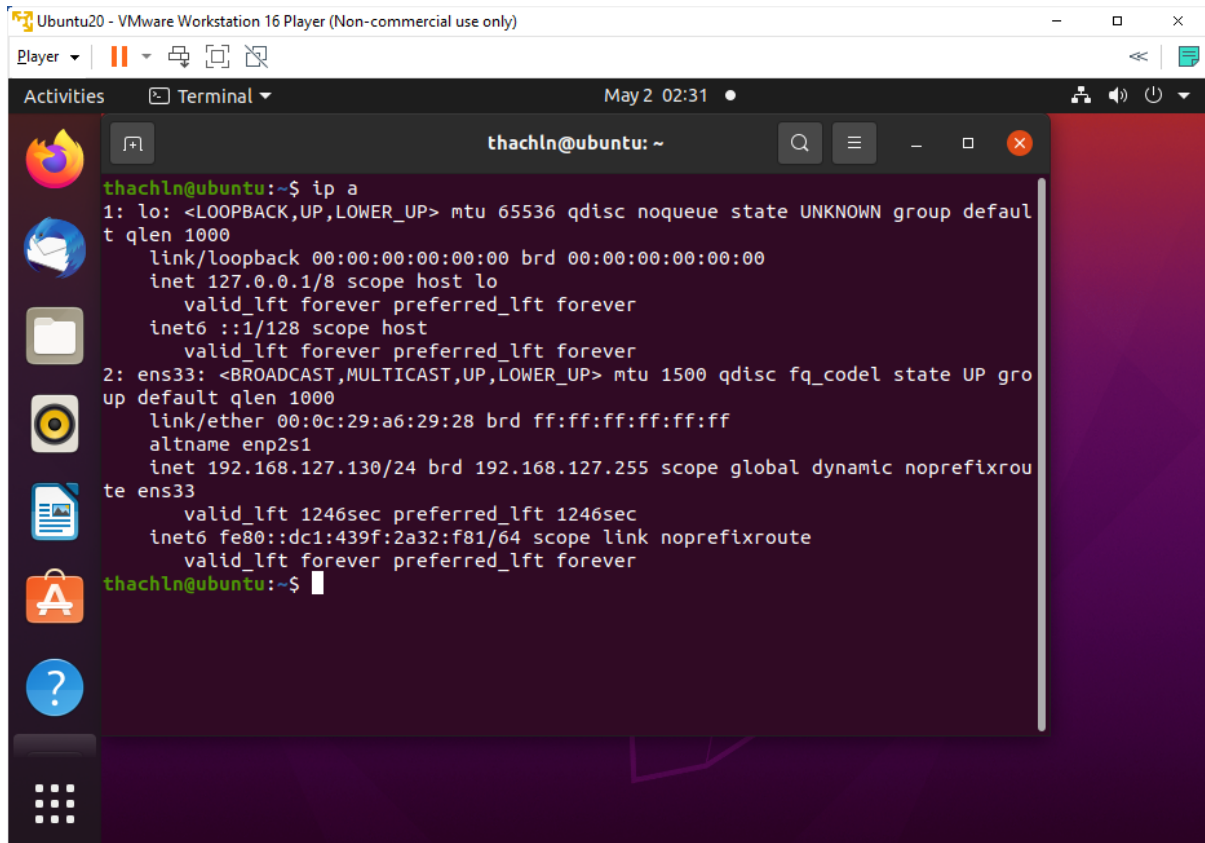
Khởi động lại nginx

```
sudo service nginx restart
```


Truy cập nginx trên Ubuntu từ máy Windows

Xem địa chỉ IP của máy Ubuntu bằng lệnh sau:

```
ip a
```

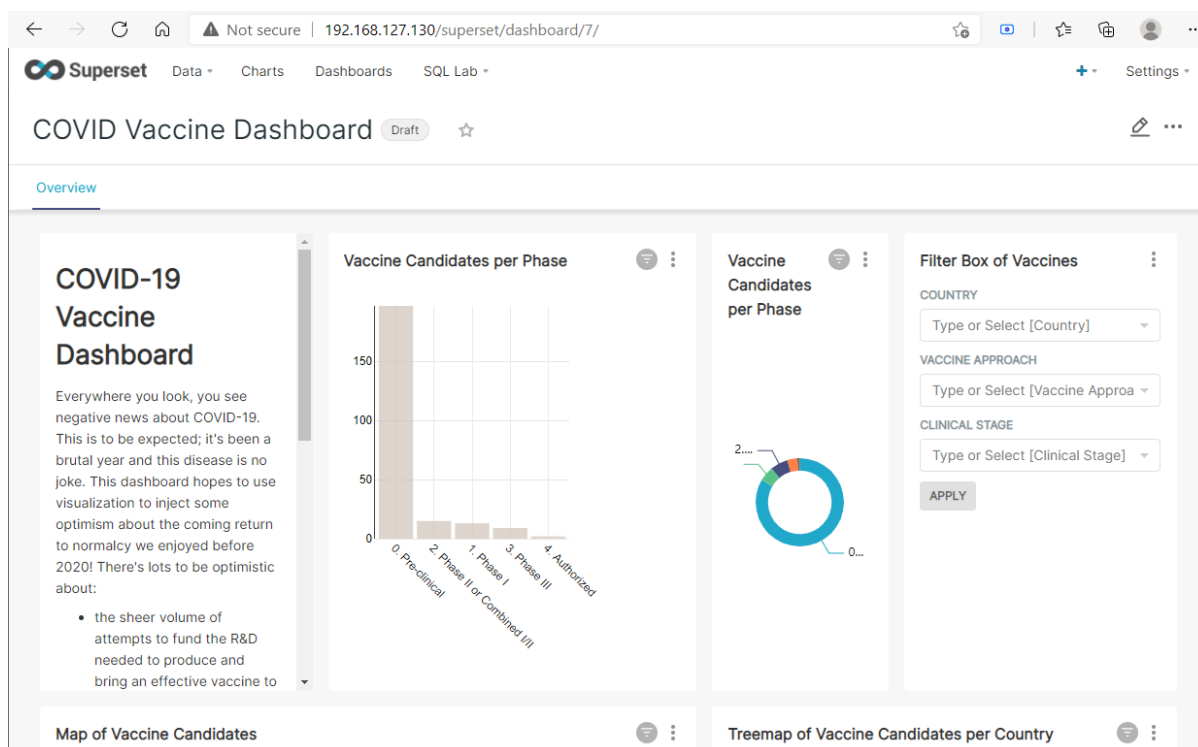


The screenshot shows a terminal window titled 'thachln@ubuntu: ~' within a VMware Workstation 16 Player. The terminal displays the output of the command 'ip a'. The output shows details for the loopback interface 'lo' (127.0.0.1) and the ethernet interface 'ens33' (192.168.127.130). The IP address 192.168.127.130 is highlighted in green. The terminal window is part of a desktop environment with various application icons on the left and a system tray on the right.

```
thachln@ubuntu:~$ ip a
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN group default qlen 1000
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
        valid_lft forever preferred_lft forever
    inet6 ::1/128 scope host
        valid_lft forever preferred_lft forever
2: ens33: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc fq_codel state UP group default qlen 1000
    link/ether 00:0c:29:a6:29:28 brd ff:ff:ff:ff:ff:ff
    altname enp2s1
    inet 192.168.127.130/24 brd 192.168.127.255 scope global dynamic noprefixroute ens33
        valid_lft 1246sec preferred_lft 1246sec
    inet6 fe80::dc1:439f:2a32:f81/64 scope link noprefixroute
        valid_lft forever preferred_lft forever
thachln@ubuntu:~$
```

Ubuntu trên máy tôi có địa chỉ IP là: 192.168.127.130.

Vì vậy đường dẫn để trải nghiệm superset là <http://192.168.127.130>



Khai thác superset

Cài đặt MySQL

[Tham khảo cài đặt MySQL trên Ubuntu tại đây.](#)

Tạo database

Thực hiện 3 lệnh sau trong cửa sổ lệnh MySQL:

```
CREATE DATABASE mysuperset DEFAULT CHARACTER SET 'UTF8MB4';  
CREATE USER 'mysuperset_user'@'%' IDENTIFIED BY  
'Mysuperset#123';  
GRANT ALL PRIVILEGES ON mysuperset.* TO 'mysuperset user'@'%'  
WITH GRANT OPTION;
```

Thử kết nối database bằng mysqlclient

Thực hiện lệnh sau trong cửa sổ lệnh của hệ điều hành (Ubuntu):

```
mysql -u mysuperset_user -pMysuperset#123 mysuperset
```

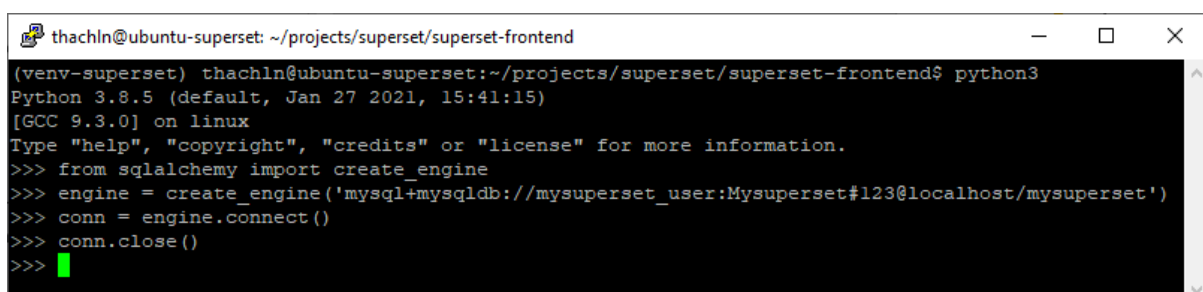
Thử test nối database bằng Python ngay trên máy Ubuntu

Mở cửa sổ lệnh Python bằng lệnh sau:

```
python3
```

Thực hiện các lệnh sau trong cửa sổ lệnh Python:

```
from sqlalchemy import create_engine  
engine =  
create_engine('mysql+mysqldb://mysuperset_user:Mysuperset#123@192.168.  
127.131/mysuperset?charset=utf8mb4')  
conn = engine.connect()  
conn.close()
```



```
thachln@ubuntu-superset: ~/projects/superset/superset-frontend  
(venv-superset) thachln@ubuntu-superset:~/projects/superset/superset-frontend$ python3  
Python 3.8.5 (default, Jan 27 2021, 15:41:15)  
[GCC 9.3.0] on linux  
Type "help", "copyright", "credits" or "license" for more information.  
>>> from sqlalchemy import create_engine  
>>> engine = create_engine('mysql+mysqldb://mysuperset_user:Mysuperset#123@localhost/mysuperset')  
>>> conn = engine.connect()  
>>> conn.close()  
>>>
```

Thử kết nối database bằng Python trên máy Windows

Cấu hình MySQL Server

Mặc định MySQL được cấu hình chỉ chạy trên localhost, tức địa chỉ IP là 127.0.0.1. Để MySQL được truy cập từ máy khác trong mạng thì sửa file cấu hình bằng lệnh sau:

```
sudo nano /etc/mysql/mysql.conf.d/mysqld.cnf
```

Tìm dòng **bind-address** sửa lại như sau:

```
bind-address            = 0.0.0.0
```

Sau đó khởi động lại mysql server bằng lệnh:

```
sudo service mysql restart
```

Cài thư viện cho môi trường python (trên client)

```
pip install mysqlclient
```

Mã nguồn Python

```
from sqlalchemy import create_engine
engine =
create_engine('mysql+mysqldb://mysuperset_user:Mysuperset#123@192.168.
127.131/mysuperset')
engine.connect()
engine.close()
```

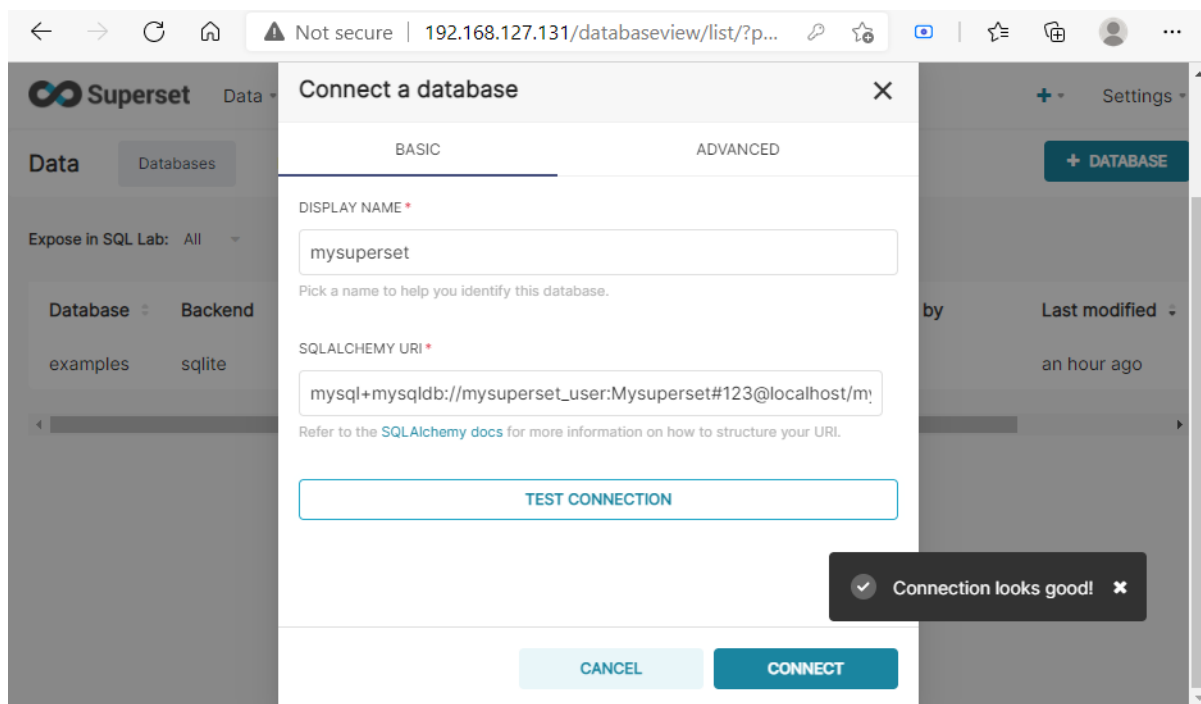
(Hãy thay địa chỉ IP bằng địa chỉ IP đúng trên máy Ubuntu của bạn)

Tạo Database trong superset

Tạo database với 2 thông số sau:

DISPLAY NAME	mysuperset (<i>bạn có thể đặt tên khác</i>)
SQL ALCHEMY URI	mysql+mysqldb://mysuperset_user:Mysuperset#123@localhost/mysuperset

Bấm nút “TEST CONNECTION” để kiểm tra kết nối.



Thử bậc “Allow data upload”

Vào tab “ADVANCED”, đánh dấu vào mục “Allow data upload”.

Edit database

×

BASIC

ADVANCED

SQL Lab

Configure how this database will function in SQL Lab.

▼

Performance

Adjust settings that will impact the performance of this database.

▼

Security

Add connection information for other systems.

▼

Other

Additional settings.

▲

☐ Impersonate Logged In User (Presto & Hive) ⓘ

☒ Allow data upload ⓘ

EXTRA

1 {

2 "metadata_params": {},

3 "engine_params": {},

4 "metadata_cache_timeout": {},

5 "schemas_allowed_for_csv_upload": []

6 }

7

JSON string containing extra configuration elements.

1. The engine_params object gets unpacked into the sqlalchemy.create_engine call,

CANCEL

SAVE

Nếu bạn không gặp lỗi gì thì coi như đã rất may mắn 😊!

Cài đặt service cho superset

Sử dụng systemd trong Ubuntu

```
sudo nano /etc/systemd/system/superset.service
```

```
# Metadata and dependencies section
[Unit]
Description=Apache Superset service
After=network.target

# Define users and app working directory
[Service]
User=thachln
Group=www-data
WorkingDirectory=/home/thachln/projects/superset
Environment="FLASK_ENV=production"
ExecStart=/home/thachln/projects/superset/venv-superset/bin/superset run -p 8088 --with-threads
# Link the service to start on multi-user system up
[Install]
WantedBy=multi-user.target
```

Bật và khởi động service superset

```
sudo systemctl enable superset
sudo systemctl start superset
```

Kiểm tra trạng thái service

```
sudo systemctl status superset
```

```
thachln@ubuntu-superset: ~
• superset.service - Apache Superset service
  Loaded: loaded (/etc/systemd/system/superset.service; enabled; vendor
  Active: active (running) since Mon 2021-07-19 14:34:35 UTC; 11min ago
  Main PID: 827 (superset)
  Tasks: 3 (limit: 9415)
  Memory: 238.8M
  CGroup: /system.slice/superset.service
          └─827 /home/thachln/projects/superset/venv-superset/bin/python>

Jul 19 14:43:04 ubuntu-superset superset[827]: 2021-07-19 14:43:04,801:INFO>
Jul 19 14:43:05 ubuntu-superset superset[827]: 2021-07-19 14:43:05,323:INFO>
Jul 19 14:43:05 ubuntu-superset superset[827]: 2021-07-19 14:43:05,378:INFO>
lines 1-12/19 43%
```


Sử dụng gunicorn (not OK)

```
cd ~/projects/superset
export VENV_NAME=venv-superset
source $VENV_NAME/bin/activate
```

Nâng cấp pip3

```
$VENV_NAME/bin/python3 -m pip install --upgrade pip
```

Cài gunicorn bằng lệnh sau:

```
$VENV_NAME/bin/pip3 install gunicorn
```

Tạo file

```
sudo nano /etc/systemd/system/superset.service
```

```
# Metadata and dependencies section
[Unit]
Description=Apache Superset service
After=network.target

# Define users and app working directory

[Service]
PIDFile = /home/thachln/projects/superset/superset-
webserver.PIDFile
User=thachln
Group=www-data
Environment=SUPERSET_HOME=/home/thachln/projects/superset
Environment=PYTHONPATH=/home/thachln/projects/superset/venv-
superset/bin
WorkingDirectory=/home/thachln/projects/superset
ExecStart =/home/thachln/projects/superset/venv-
superset/bin/python3 /home/thachln/projects/superset/venv-
superset/bin/gunicorn --workers 8 --worker-class gevent --bind
0.0.0.0:8088 --pid /home/thachln/projects/superset/superset-
webserver.PIDFile superset:app
ExecStop = /bin/kill -s TERM $MAINPID

[Install]
```

```
WantedBy=multi-user.target
```

```
sudo systemctl daemon-reload
sudo systemctl start superset
sudo systemctl enable superset
```

Sử dụng uWSGI

uWSGI is a [software application](#) that "aims at developing a full stack for building [hosting services](#)".^[3] It is named after the [Web Server Gateway Interface](#) (WSGI), which was the first plugin supported by the project.^[3]

uwsgi (all lowercase) is the native binary protocol that uWSGI uses to communicate with other servers.^[4]

uWSGI is often used for serving [Python web applications](#) in conjunction with [web servers](#) such as [Cherokee](#) and [Nginx](#), which offer direct support for uWSGI's native uwsgi protocol.^[5] For example, data may flow like this: HTTP client ↔ Nginx ↔ uWSGI ↔ Python app.^[6]

(<https://en.wikipedia.org/wiki/UWSGI>)

Đăng nhập lại máy ảo và thực hiện lệnh sau:

```
cd ~/projects/superset
export VENV_NAME=venv-superset
source $VENV_NAME/bin/activate
```

Nâng cấp pip3

```
$VENV_NAME/bin/python3 -m pip install --upgrade pip
```

Cài uWSGI bằng lệnh sau:

```
$VENV_NAME/bin/pip3 install uwsgi
```

Tạo file superset.ini với nội dung sau:

```
nano ~/superset.ini
```

```
[uwsgi]
module = run:application
master = true
processes = 5
socket = superset.sock
chmod-socket = 660
vacuum = true
```

```
die-on-term = true
```

```
# Metadata and dependencies section
[Unit]
Description=Apache Superset service
After=network.target

# Define users and app working directory
[Service]
User=thachln
Group=www-data
WorkingDirectory=/home/thachln/projects/superset
Environment="FLASK_ENV=production"
ExecStart=/home/thachln/projects/superset/venv-
superset/bin/uwsgi --ini /home/thachln/superset.ini

# Link the service to start on multi-user system up
[Install]
WantedBy=multi-user.target
```

Khởi động

```
sudo systemctl start superset
```

Tạo user

Lệnh sau tạo user superset không thể login, không có thư mục home.

```
sudo useradd superset -s /bin/false -M
```

Tạo file

```
sudo nano /etc/systemd/system/superset.service
```

Với nội dung:

```
[Unit]
Description=Gunicorn instance to serve superset
```

```
After=network.target

[Service]
User=superset
Group=www-data
WorkingDirectory=/opt/superset
Environment="PATH=/opt/superset/venv/bin"
ExecStart=/opt/superset/venv/bin/gunicorn --workers 3 --bind
unix:superset.sock -m 007 wsgi:app
```

Tham khảo:

<https://www.digitalocean.com/community/tutorials/how-to-serve-flask-applications-with-gunicorn-and-nginx-on-ubuntu-20-04>