# Hướng dẫn dùng docker nộp bài Zalo AI Challenge 2020



## Giới thiệu về Docker

Docker là một nền tảng giúp người dùng đóng gói và chạy chương trình của mình trên các môi trường khác nhau một cách nhanh nhất dựa trên các container.

Docker image là một dạng tập hợp các tệp của ứng dụng, được tạo ra bởi Docker engine. Nội dung của các Docker image sẽ không bị thay đổi khi di chuyển. Docker image được dùng để chạy các Docker container.

Docker Container là một dạng runtime của các Docker image, dùng để làm môi trường chạy ứng dụng.

Hướng dẫn chi tiết các bạn có thể tham khảo tại:

https://docs.docker.com/get-started/

## Cài đặt Docker

Trên Ubuntu

\$ sudo apt-get install docker.io

Đối với các hệ điều hành khác, tham khảo cách cài đặt tại đây (https://docs.docker.com/install/overview/)

Test docker

Kiểm tra phiên bản docker

\$ sudo docker --version

Docker version 18.06.1-ce, build e68fc7a

Chay hello-world

S sudo docker run hello-world

Unable to find image 'hello-world:latest' locally

latest: Pulling from library/hello-world

9db2ca6ccae0: Pull complete

Digest: sha256:4b8ff392a12ed9ea17784bd3c9a8b1fa3299cac44aca35a85c90c5e3c7afacdc

Status: Downloaded newer image for hello-world:latest

Hello from Docker!

This message shows that your installation appears to be working correctly.

...

Liệt kê các docker image đang có \$ sudo docker image Is

REPOSITORY	TAG	IMAGE ID	CREATED	SIZE	
hello-world	latest	2cb0d9787c4d	6 weeks ago	1.85kB	

Liệt kê các docker container đang có

\$ sudo docker container Is --all

CONTAINER ID	IMAGE	COMMAND	CREATED	STATUS	
1cb1aefa1f44	hello-world	"/hello"	9 minutes ago	Exited (0) 9 minutes ago	

#### Nvidia docker

Để sử dụng được GPU trên docker, bạn cần cài đặt Nvidia docker.

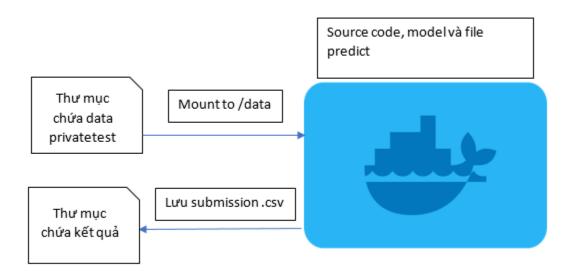
Hướng dẫn cài đặt Nvidia docker tại đây:

https://github.com/NVIDIA/nvidia-docker

### Nộp bài cho Zalo AI Challenge

Các bước để build một docker image cho 1 task là:

- Tạo một docker container mới (hoặc sử dụng docker có sẵn).
- Đưa model và source code đang có vào docker.
- Install các packages và libraries cần thiết mà solution của bạn sử dụng để chạy.
- Viết một file script tên predict.sh. File này chứa command các bước để chạy test.
   Nhận input từ /data và output ra /result/submission.jsonl.
- Tập tin private test là: private\_test.jsonl (trong thư mục data)
- Commit các thay đổi trong docker container
- Save docker container thành 1 file image và nộp lên website cuộc thi.



Ví dụ model và source code của bạn gồm các file

```
|---- predict.py
|---- preprocessing.py
|---- saved_models
| |---- models.h5
|---- train.py
|---- utils.py
```

Khởi động docker container \$ sudo docker run –it [image name]

user:~/ZA\$ sudo docker run -it zaloai/news

Lấy CONTAINER ID \$ sudo docker ps

> user:~/ZA\$ sudo docker ps CONTAINER ID IMAGE 0678bc69efd zaloai/news

1. Copy source code và model vào container sudo docker cp [source path] [CONTAINER ID]:[destination path]

user:~/ZA\$ sudo docker cp /home/ml/ZA/zaloai\_news/ 0678cb69efed:/model

2. Cài đặt package còn thiếu vào container.

Ví du

\$ pip install [python-package-name]

hoăc

\$ apt-get install [package-name]

(base) root@0678cb69efed:/notebooks# pip install imgaug
Collecting imgaug
Downloading https://files.pythonhosted.org/packages/ad/2e
ae9b585f448569ae520031932687761165419a2/imgaug-0.2.6.tar.gz
100% | 634kB 152kB/s

3. Chỉnh sửa file /model/predict.sh

Ví dụ có 2 bước để chạy bộ test ra kết quả.

python /model/preprocessing.py python /model/predict.py

4. Lưu lại thay đổi trên container (terminal 2)\$ sudo docker commit [CONTAINER ID] [image name]

user:~/ZA\$ sudo docker commit 0678cb69efed zaloai\_news:version2 sha256: 6d63dd18d69849f9785fc675a7bbe3de93feed8c41116e6e0fbd053849ce60dc

5. Verify lại docker

 $\$  sudo docker run –v [path to test data] /data –v ./ /result [docker name] /bin/bash /model/predict.sh

Kết quả sẽ được trả về file submission. jsonl ở thư mục hiện tại nếu cài đặt đúng.

6. Export file image và nộp bài\$ docker save -o [output file] [image id]file [output file] là file bạn cần nộp.

## **Upload docker**

- 1. Kiểm tra content của docker:
  - ✓ Toàn bô source code
  - ✓ Model checkpoints
  - ✓ Training & inference documents.
- 2. Nén docker của các bạn ở các định dạng sau: .tar.gz | .tar.bz2 | .7z | .rar | .zip
- 3. Upload docker:

BTC sẽ download docker của các bạn về máy chủ để tiến hành bước kiểm tra kết quả cuối cùng.

- ✓ Upload lên dịch vụ storage của Google drive
- ✓ Chỉnh sang chế độ share "Anyone can view".
- ✓ Gửi URL download cho BTC

