



TẬP ĐOÀN BƯU CHÍNH VIỄN THÔNG VIỆT NAM
CÔNG TY CÔNG NGHỆ THÔNG TIN VNPT

SUBMISSION GUIDELINE



TẬP ĐOÀN BƯU CHÍNH VIỄN THÔNG VIỆT NAM

CÔNG TY CÔNG NGHỆ THÔNG TIN VNPT

1. Giới thiệu

Tài liệu này quy định các chuẩn mực về cấu trúc Repository của cuộc thi **VNPT AI - Age of AInicorns - Track 2 The Builder**, quy trình xử lý dữ liệu và môi trường Docker để đảm bảo hệ thống chấm điểm tự động hoạt động chính xác.

Các đội tham gia vui lòng đọc kỹ và tuân thủ các yêu cầu dưới đây.

2. Yêu cầu về Dockerfile

Để đảm bảo tính nhất quán môi trường giữa máy của thí sinh và máy chấm của Ban Tổ Chức (BTC), các đội thi phải cung cấp Dockerfile để build image cho giải pháp của mình.

2.1. Nguyên tắc chung:

- Dockerfile phải build được từ một **Base Image rỗng/sạch**.
- Các tài nguyên cần thiết (model weights, database indexes) nên được tải về hoặc khởi tạo trong quá trình build.

2.2. Cấu hình CUDA

Đối với các giải pháp sử dụng GPU, bắt buộc sử dụng Base Image hỗ trợ **CUDA version 12.2** để tương thích với phần cứng của BTC.

VD: nvidia/cuda:12.2.0-devel-ubuntu20.04

2.3. Dockerfile Template

Dưới đây là mẫu Dockerfile tiêu chuẩn mà các đội có thể tham khảo và tùy chỉnh cho phù hợp với source code của mình:

Dockerfile

```
# BASE IMAGE
```



TẬP ĐOÀN BƯU CHÍNH VIỄN THÔNG VIỆT NAM
CÔNG TY CÔNG NGHỆ THÔNG TIN VNPT

```
# Lưu ý: Sử dụng đúng phiên bản CUDA 12.2 để khớp với Server BTC
# -----
FROM nvidia/cuda:12.2.0-devel-ubuntu20.04

# -----
# SYSTEM DEPENDENCIES
# Cài đặt Python, Pip và các gói hệ thống cần thiết
# -----
RUN apt-get update && apt-get install -y \
    python3 \
    python3-pip \
    git \
&& rm -rf /var/lib/apt/lists/*

# Link python3 thành python nếu cần
RUN ln -s /usr/bin/python3 /usr/bin/python

# -----
# PROJECT SETUP
# -----
# Thiết lập thư mục làm việc
WORKDIR /code

# Copy toàn bộ source code vào trong container
COPY #PATH_TO_REPO /code

# -----
# INSTALL LIBRARIES
# -----
# Nâng cấp pip và cài đặt các thư viện từ requirements.txt
RUN pip3 install --no-cache-dir --upgrade pip && \
    pip3 install --no-cache-dir -r requirements.txt

# -----
# EXECUTION
# Lệnh chạy mặc định khi container khởi động
```



TẬP ĐOÀN BƯU CHÍNH VIỄN THÔNG VIỆT NAM

CÔNG TY CÔNG NGHỆ THÔNG TIN VNPT

```
# Pipeline sẽ đọc private_test.json và xuất ra submission.csv.  
submission_time.csv  
# -----  
CMD [ "bash", "inference.sh" ]
```

2.4. Checklist kiểm tra trước khi nộp

Trước khi gửi link repo, hãy tự kiểm tra bằng cách build và run docker container trên máy local:

1. sudo docker build -t team_submission .
2. sudo docker run --gpus all -v /path/to/data:/app/data team_submission
3. Kiểm tra xem file **submission.csv**, **submission_time.csv** có được sinh ra sau khi container chạy xong hay không.
4. Đẩy image đã được tạo lên Dockerhub.

Lưu ý: Các đội sẽ submit link Github Repository (đã bao gồm Dockerfile), cùng với đó là tên Image đã được đẩy lên Docker Hub. Đội với các Image được đẩy lên quá thời gian submit **23:59 (UTC+7) ngày 19/12/2025** thì sẽ coi là không hợp lệ.

2. Yêu cầu về GitHub Repository

Mỗi đội thi cần cung cấp link GitHub Repository chứa toàn bộ mã nguồn của giải pháp. Repository cần đảm bảo các yếu tố sau:

2.1. README.md

File README.md là bắt buộc và phải chứa các thông tin mô tả chi tiết:

1. **Pipeline Flow:** Mô tả rõ ràng luồng xử lý của hệ thống (có sơ đồ minh họa là một điểm cộng).
2. **Data Processing:** Các bước thu thập, làm sạch và xử lý dữ liệu.
3. **Resource Initialization:** Hướng dẫn cách khởi tạo Vector Database, Indexing, các tài nguyên cần thiết để chạy được pipeline đã xây dựng.

2.2. Quản lý thư viện



TẬP ĐOÀN BƯU CHÍNH VIỄN THÔNG VIỆT NAM

CÔNG TY CÔNG NGHỆ THÔNG TIN VNPT

- Liệt kê đầy đủ các thư viện và phiên bản cụ thể trong file requirements.txt
- Đảm bảo không xảy ra xung đột phiên bản khi cài đặt

2.3. Tổ chức mã nguồn

Repository **bao gồm** file predict.py đóng vai trò là entry-point của pipeline. File này sẽ thực hiện cơ chế End-to-End:

1. Đọc dữ liệu từ file /code/private_test.json (File này sẽ được hệ thống mount vào khi chấm điểm).
2. Chạy toàn bộ pipeline (Retrieval, Generation, etc.).
3. Xuất kết quả ra file **submission.csv**, **submission_time.csv** (xem hướng dẫn bên dưới) theo định dạng quy định của BTC.

Đối với các file khác bao gồm xử lý dữ liệu, khởi tạo vector database, các đội vui lòng viết một bash scripts chạy pipeline từ đầu tới cuối, ví dụ:

```
# inference.sh
#!/bin/bash
python data_process.py
python predict.py
```

Đối với phần chấm điểm thời gian inference, đề nghị các đội chạy theo vòng for theo từng item trong pipeline để có thể lấy được thời gian chạy của từng sample, từ đó BTC có căn cứ chấm điểm thời gian inference của các đội một cách công bằng nhất.

Các đội có thể tham khảo code dưới đây để lấy thời gian chạy của từng câu trong tập test:

```
for item in test:    # test là list các câu test
    start = time.time()
    # --- chạy pipeline của đội tại đây ---
    # ví dụ:
    output = model.inference(item)
    # -----
    end = time.time()
```



TẬP ĐOÀN BƯU CHÍNH VIỄN THÔNG VIỆT NAM
CÔNG TY CÔNG NGHỆ THÔNG TIN VNPT

```
time_infer_sample = end - start
```

Sau đó, các đội tiến hành ghi thêm file **submission_time.csv** có chứa thông tin của cột time cho từng sample.

submission.csv	submission_time.csv
qid,answer	qid,answer,time
test_0001,A	test_0001,A,1.2345
test_0002,B	test_0002,B,2.9087
test_0003,C	test_0003,C,1.0021
test_0004,D	test_0004,D,3.4509

3. Phương thức nộp bài thi cuối cùng

Trong ngày 19/12/205, BTC sẽ mở PORT cho các thí sinh nộp bài thi, các thông tin thí sinh cần nộp bao gồm:

1. Link Github Repository - được public và không được chỉnh sửa sau thời gian submit.
2. Tên Docker Image đã được push lên DockerHub. (theo như quy định tại [mục 2.4](#)).

Chúc các đội thi thành công!