**BÁO CÁO CRAWLER & DOCKER**

**Môn: Nhập môn dữ liệu lớn**

Tên: Đỗ Thành Minh Phú

MSSV: 22701391

Link source code: [Bigdata-Major/Week1 at main · Esonip/Bigdata-Major](https://github.com/Esonip/Bigdata-Major/tree/main/Week1)

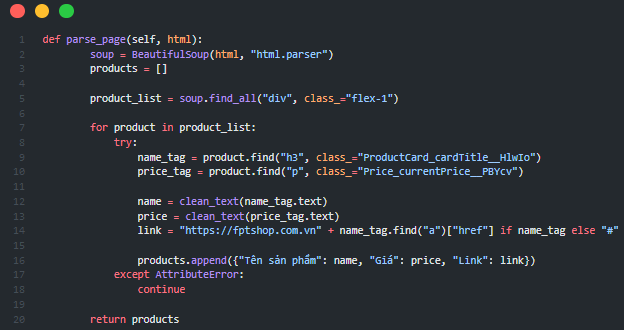
# **1. Crawl và lưu trữ dữ liệu**

## Trang FPTShop

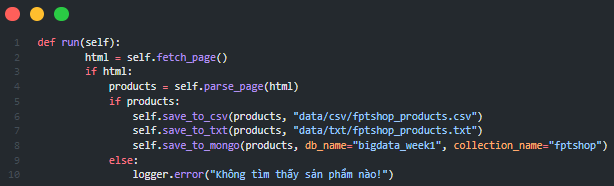
* Mục tiêu: Thu thập sản phẩm(điện thoại), giá tiền và link sản phẩm

### Crawl dữ liệu:

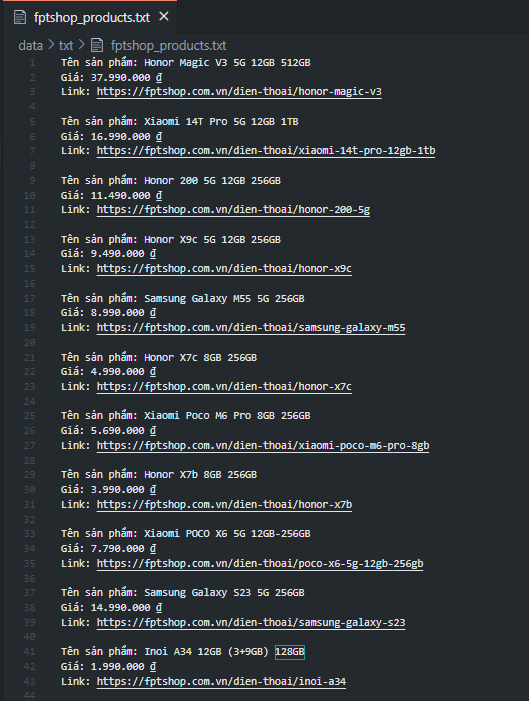
* + Phân tích HTML và lấy thông tin sản phẩm

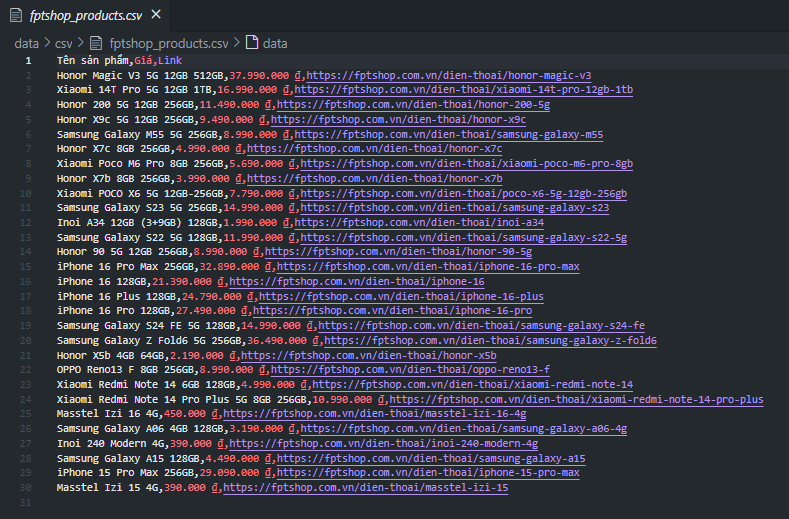


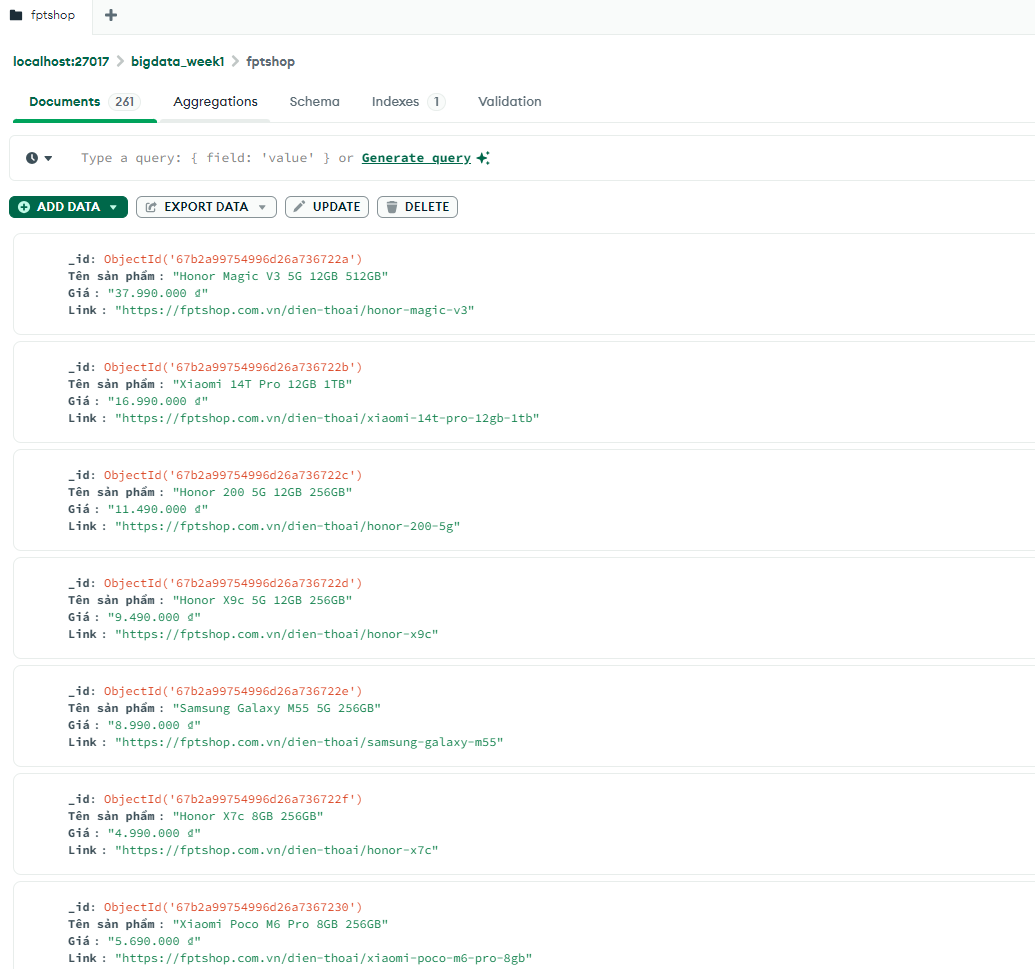
* + Chạy và lưu thông tin vào mongoDB, csv và txt.



### Lưu trữ vào mongoDB, csv, txt.







B. Trang Nhaccuatui

## Trang Nhaccuatui

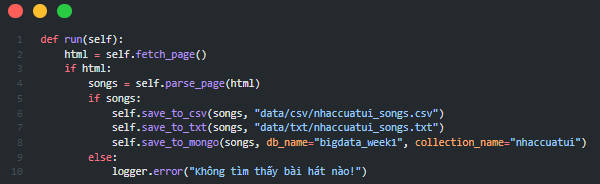
* Mục tiêu: Thu thập tên bài hát, nghệ sĩ và link bài hát

### Crawl dữ liệu:

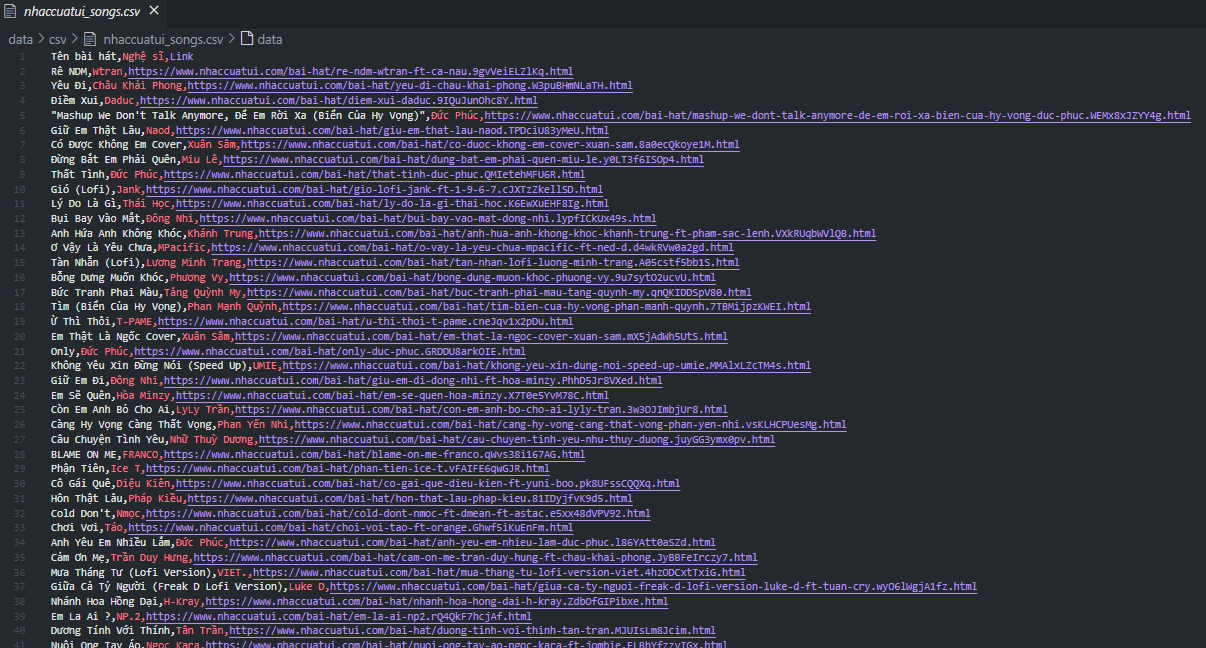
* 1. Phân tích HTML và lấy thông tin bài hát

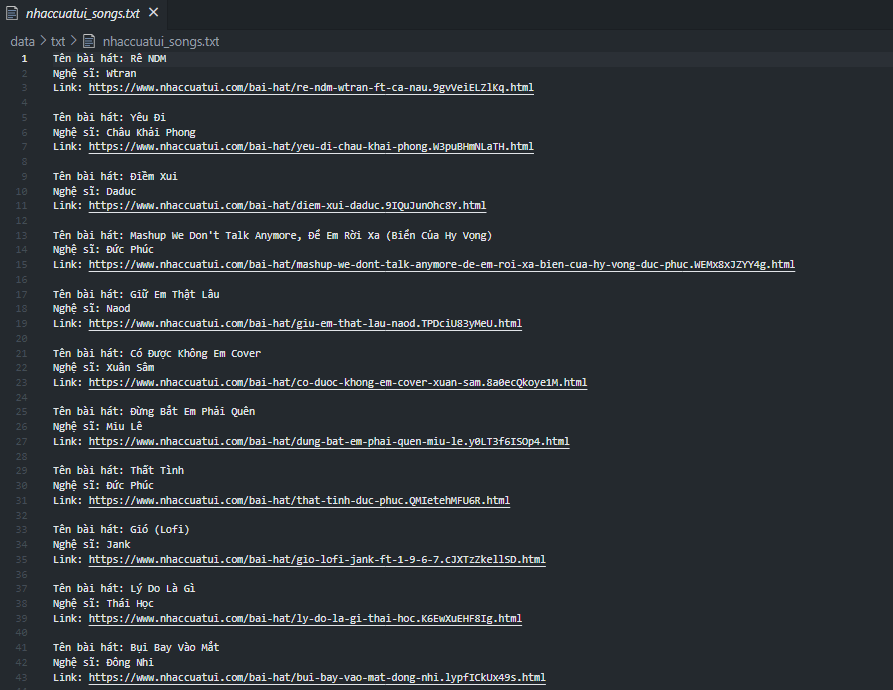


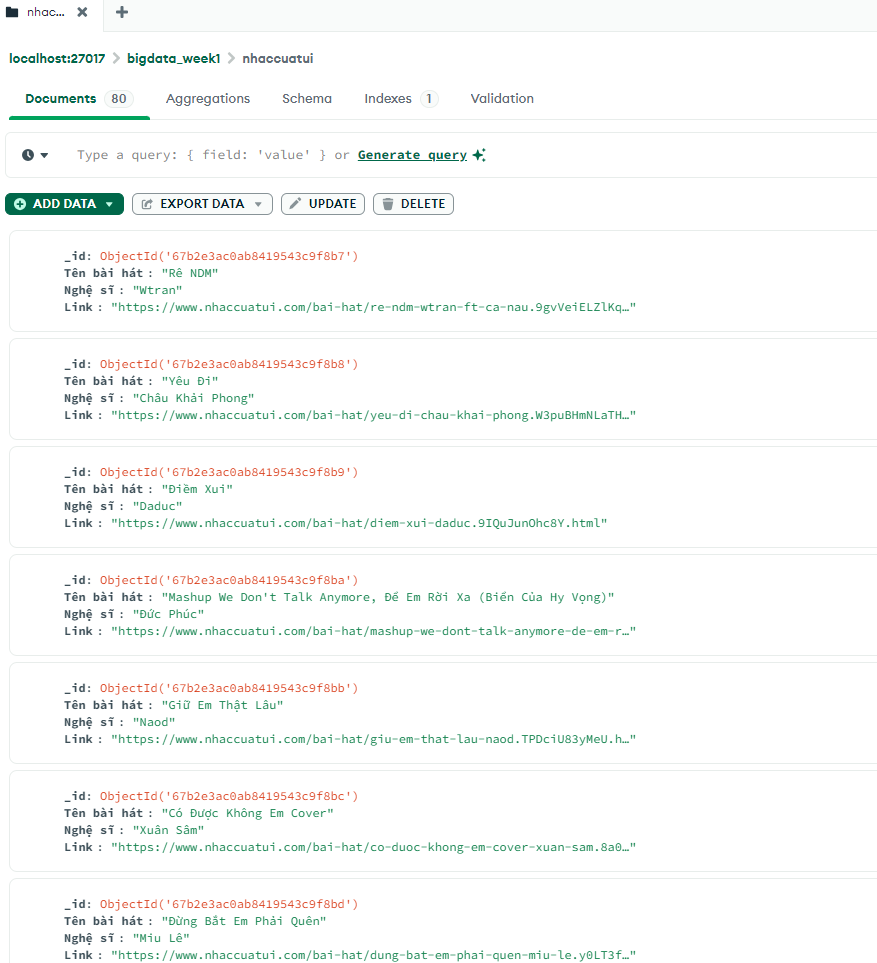
* 1. Chạy và lưu thông tin vào mongoDB, csv và txt.



### Lưu trữ vào mongoDB, csv, txt

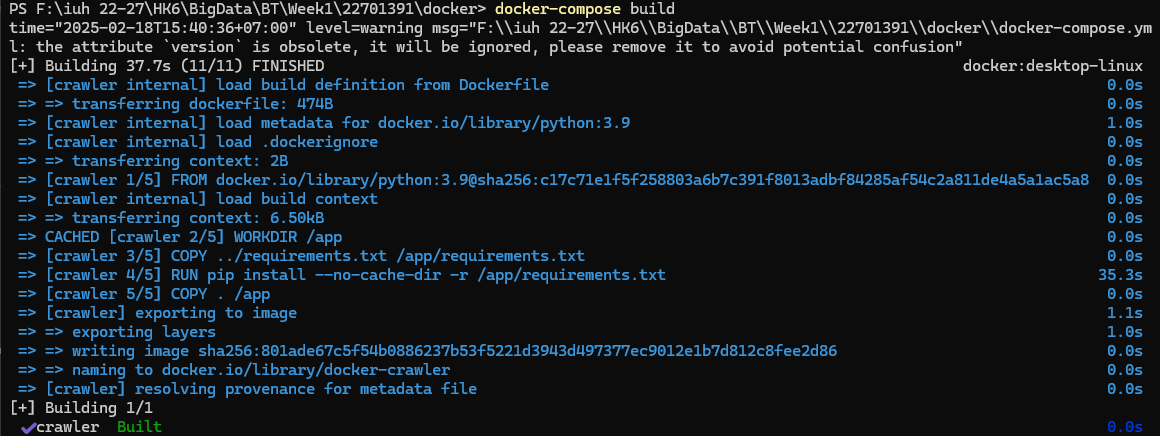




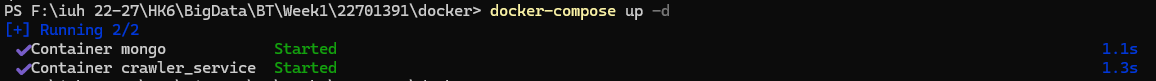


# **2. Đóng gói và triển khai bằng Docker**

* Xây dựng Docker Image:  **docker-compose build**



* Chạy docker container: **docker-compose up -d**



* Kiểm trả container hoạt động: **docker ps**

