

PySpark

Giảng viên: Tiến sĩ Bùi Thanh Hùng
Email: tg_buithanhhung@tdtu.edu.vn
Website: <https://sites.google.com/site/hungthanhbui1980>

Dữ liệu về chiến tranh Việt Nam ở đường link:

<https://drive.google.com/drive/folders/1vJHhgMdcE9MyIBJw149E4oQX7tCnL05m>

Dữ liệu gồm 2 files:

Bombing_Operations .json.gz

Aircraft_Glossary.json.gz

Mô tả chi tiết về dữ liệu gồm có các thông tin sau:

File Bombing_Operations.json.gz

- AirCraft: *Aircraft model*
- ContryFlyingMission: *Country*
- MissionDate: *Date of the mission*
- OperationSupported: *Supported War operation*
- PeriodOfDay: *Day or night*
- TakeoffLocation: *Take off airport*
- TimeOnTarget
- WeaponType
- WeaponsLoadedWeight

File Aircraft_Glossary.json.gz

- AirCraft: *Aircraft model*
- AirCraftName
- AirCraftType

Hãy sử dụng PySpark giải quyết các câu hỏi sau:

- 1- Những quốc gia nào tham gia và thực hiện bao nhiêu phi vụ trong cuộc chiến?
- 2- Hiện thị số lượng phi vụ theo thời gian cho mỗi quốc gia liên quan?
- 3- Ai đã đánh bom địa điểm này?



Bức ảnh này là một cơ sở ở Hanoi POL (Miền Bắc Việt Nam) đang bốc cháy sau khi bị Không quân Hoa Kỳ tấn công vào ngày 29 tháng 6 năm 1966 trong bối cảnh chiến dịch Rolling Thunder.

- 4- Hãy cho biết loại máy bay được sử dụng nhiều nhất trong chiến tranh Việt Nam (loại máy bay thực hiện số lượng phi vụ nhiều nhất) là gì?