Báo cáo giữa kì Project 1

**Yêu cầu:**

Xây dựng công cụ thu thập dữ liệu về các cloud services / cloud provider

**Thực hiện:**

**- Tìm hiểu về các Cloud Provider**

+ Tìm và liệt kê các Cloud provider phổ biến rồi tìm những dịch vụ / tùy chọn phổ biến của các Cloud Provider đã liệt kê

+ Đưa ra check list các dịch vụ / tùy chọn của các Cloud Provider

Graphical user interface, table

Description automatically generated

<https://docs.google.com/spreadsheets/d/1xHjT7R3sLXNq3wEinUSKC6FYG7ARK5b-/edit#gid=1497487099>

**- Phân tích các loại dịch vụ Cloud:**

+ Cloud Hosting

+ Cloud Storage

+ Cloud Virtual Machine

**- Phân tích các yếu tố đánh giá chung cho Cloud provider và Cloud option dựa trên SMI(Service measurement index)**

+ Accountability: các chỉ số liên quan đến nhà cung cấp dịch vụ đám mây

+ Agility: tính linh hoạt của nhà cung cấp

+ Assurance: độ bảo đảm của nhà cung cấp

+ Financial: tài chính mà khách hang bỏ ra

+ Performance: thông số của tùy chọn

+ Security and Privacy: Các tùy chọn bảo mật

+ Usability: tính dễ sử dụng của ứng dụng

**- Các công nghệ sử dụng cho việc thiết kế, xây dựng công cụ thu thập dữ liệu về các cloud services / cloud provider:**

**+ Python3: ngôn ngữ nền tảng để xây dựng crawler và các công cụ xử lí dữ liệu**

**Dành cho linux:**

Trong terminal gõ “**sudo apt-get install python3-pip**” để install pip

Install các thư viện bằng cách gõ trong terminal “**pip install <Tên thư viện**>”

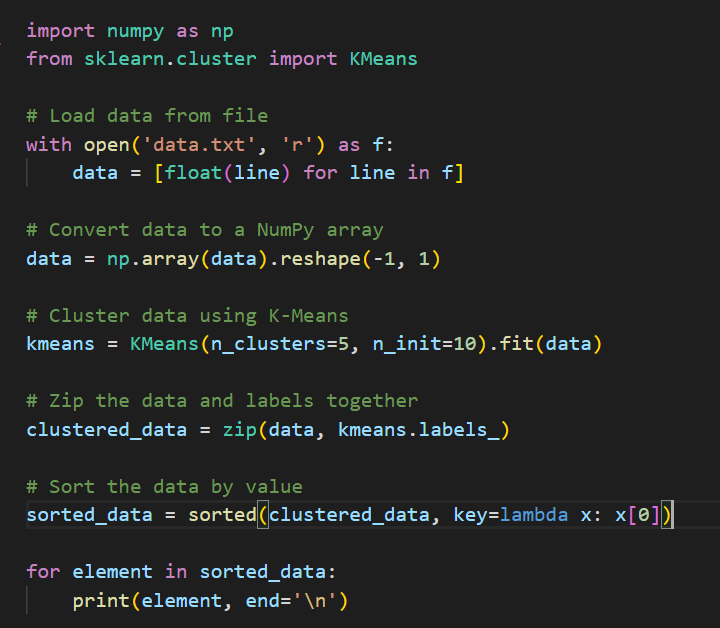
**\* Các thư viện sử dụng:**

**<beautifulsoup> : Thư viện dành cho công cụ crawler**

**<numpy>: Thư viện dành cho việc tính toán, sắp xếp dữ liệu**

**<** **scikit-learn>: Thư viện để phân cụm, phân lớp dữ liệu để tạo thang điểm đánh giá**

**<pyodbc>: để kết nối code python với SQL Server**

****

Giả xử các dữ liệu cần phân cụm đang được để tạm ở file data.txt, đoạn code dưới giúp phân cụm các dữ liệu để có thể tạo ra được 1 thang điểm đánh giá

\*Vẫn chưa thể sắp xếp sao cho điểm cũng như chỉ số dữ liệu đồng thời tăng

**Text

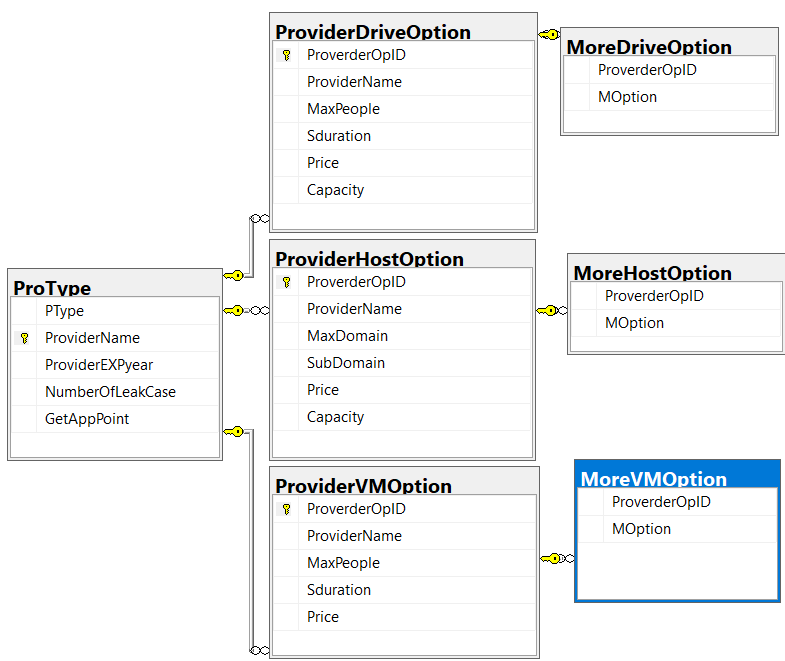
Description automatically generated**

Kết quả trả về của đoạn thuật toán phân cụm dựa trên dữ liệu thửu nghiệm cho trước, bên trái là giá trị thật, bên phải là label, cùng label tức cùng số điểm. Hiện tại chỉ mới có thể sort hoặc value, hoặc label.

**+ Microsoft SQL Server**

**- Hiện tại chưa thành công cài đặt cho máy chạy linux do yếu tố khách quan**

**- Phiên bản DEMO sử dụng trên Window để hình dung cơ sở dữ liệu:**

****

**Bảng Protype: Bao gồm tên và các thuộc tính để đánh giá 1 provider**

\* Trong đó ProType là loại dịch vụ provider đó cùng cấp: “Host,Drive,Virtual Machine”

Từ đó với mỗi **PType** sẽ có 1 bảng tương ứng bao gồm các thuộc tính chung nhất, cơ bản nhất để đánh giá 1 option mà provider cung cấp, các thuộc tính khác của option sẽ được đẩy sang các bảng **MoreOption** tương ứng ở bên cạnh để Users có thể dung để search 1 hoặc vài các tính năng đặc biệt mà chỉ 1 vài option cung cấp

Công việc đánh giá và đưa ra lựa chọn tốt nhất cho user sẽ phụ thuộc vào thang điểm sau khi chạy hệ thống nhân với hệ số được đưa ra dựa trên yêu cầu của Users.