**ĐẠI HỌC BÁCH KHOA HÀ NỘI**

**TRƯỜNG CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**

**------------\*------------**



**Báo cáo**

**Project I**

**Sinh viên thực hiện: Trịnh Phú Quang**

**MSSV: 20205219**

**Giáo viên hướng dẫn:** TS. Vũ Thị Hương Giang

**HÀ NỘI, 02/2023**

**Mục Lục**

[**Chương 1. Giới thiệu đề tài** 3](#_Toc129636277)

[**1.1 Đặt vấn đề** 3](#_Toc129636278)

[**1.2 Mục tiêu và phạm vi đề tài** 3](#_Toc129636279)

[**1.3 Định hướng giải pháp** 4](#_Toc129636280)

[**1.4 Bố cục báo cáo** 4](#_Toc129636281)

[**Chương 2 Thiết kế** 5](#_Toc129636282)

[**2.1 Khảo sát hiện trạng** 5](#_Toc129636283)

[**2.2 Ý tưởng thực hiện** 5](#_Toc129636284)

[**2.3 Tổng quan các chức năng** 6](#_Toc129636285)

[**2.4 Đặc tả chức năng** 6](#_Toc129636286)

[**Chương 3. Công nghệ sử dụng** 7](#_Toc129636287)

[**3.1 Các công nghệ cần cài đặt** 7](#_Toc129636288)

[**3.2 Hướng dẫn sử dụng công nghệ** 9](#_Toc129636289)

[**3.2.1 Hướng dẫn sử dụng MySQL** 9](#_Toc129636290)

[**Chương 4. Thực nghiệm và đánh giá** 16](#_Toc129636291)

[**4.1 Thiết kế** 16](#_Toc129636292)

[**4.1.1 Thiết kế cơ sở dữ liệu** 16](#_Toc129636293)

[**4.1.1 Xử lí dữ liệu trong Database** 20](#_Toc129636294)

[**4.1.3 Tương tác người dùng và ứng dụng** 23](#_Toc129636295)

[**4.2 Kiểm thử** 28](#_Toc129636296)

[**4.2 Đánh giá** 30](#_Toc129636297)

[**Chương 5. Kết luận** 30](#_Toc129636298)

# 

# **Chương 1. Giới thiệu đề tài**

## **1.1 Đặt vấn đề**

Trong thời đại số hóa, việc sử dụng các dịch vụ đám mây (cloud services) đang trở nên ngày càng phổ biến. Các dịch vụ này cung cấp khả năng lưu trữ, xử lý, truyền thông, phân tích dữ liệu, và nhiều hơn nữa, thông qua mạng internet. Các nhà cung cấp đám mây (cloud providers) như Amazon Web Services (AWS), Microsoft Azure, và Google Cloud Platform (GCP) cung cấp nhiều tùy chọn khác nhau cho khách hàng của họ để tùy chỉnh và sử dụng các dịch vụ này.

Tuy nhiên, với sự phát triển nhanh chóng của các công nghệ và dịch vụ đám mây, việc chọn lựa và so sánh các tùy chọn của các nhà cung cấp đám mây đã trở nên phức tạp và khó khăn hơn. Việc mỗi nhà cung cấp đám mây đưa ra 1 tiêu chuẩn riêng và cách gọi tên riêng cho dịch vụ của họ đã đem lại nhiều khó khăn cho người dung trong việc lựa chọn nhà cung phù hợp khi không có 1 tiêu chuẩn chung để đánh giá các nhà cung cấp dịch vụ đám mây này.

## **1.2 Mục tiêu và phạm vi đề tài**

Mục tiêu của đề tài này là xây dựng một hệ thống tự động crawl các nhà cung cấp đám mây và đánh giá các tùy chọn của các nhà cung cấp, để giúp người dùng có thể tìm kiếm và so sánh các dịch vụ đám mây dễ dàng hơn. Cụ thể, đề tài sẽ tập trung vào các nhà cung cấp đám mây hàng đầu và đưa ra đánh giá dựa trên các tiêu trí như giá cả, tính năng, hiệu suất, bảo mật, độ tin cậy, và khả năng hỗ trợ khách hàng. Ngoài ra, hệ thống cũng sẽ cung cấp gợi ý các dịch vụ đám mây phù hợp với nhu cầu và yêu cầu của người dùng. Với đề tài này, tôi hy vọng sẽ giúp người dùng tiết kiệm thời gian và nâng cao hiệu quả tìm kiếm và sử dụng các dịch vụ đám mây.

Đề tài nghiên cứu này hiện này có vẻ vẫn còn mới, vậy nên tính tới thời điểm làm nghiên cứu này, tôi chưa tìm thấy nghiên cứu hay ứng dụng nào liên quan để có thể tham khảo hay đánh giá qua về đề tài này.

Đề tài sẽ tập trung vào việc phân tích dữ liệu từ các nguồn dữ liệu có thể crawl về được và đưa ra những đánh giá cơ bản về nhà cung cấp, đề tài này sẽ tập trung chủ yếu và 2 mảng lớn của việc cung cấp dịch vụ đám mây đó là cung cấp Cloud Storage: dịch vụ lưu trữ đám mây và VPS host: dịch vụ máy chủ ảo để host các dịch vụ trực tuyến, các website hay các ứng dụng web

## **1.3 Định hướng giải pháp**

Một hệ thống thống kê và tổng hợp các tiêu chí của các nhà cung cấp đám mây rồi đưa ra những lựa chọn về nhà cung cấp tốt nhất bằng những thông tin người dùng yêu cầu và đối chiếu vào trong hệ thống cơ sở dữ liệu. Hệ thống đánh giá được đưa ra những tiêu chí đánh giá dựa trên Service Measurement Index (SMI).

## **1.4 Bố cục báo cáo**

Phần còn lại của báo các gồm các phần như sau:

Chương 2 bao gồm việc khảo sát các công cụ tương tự trong việc giải quyết bài toán đề ra, chương trình để giải quyết bài toán của đề tài cùng với các đặc tả và biểu đồ usecase của chương trình.

Chương 3 bao gồm các công nghê được sử dụng cho chương trình, link hướng dẫn cài đặt, một số lưu ý khi cài đặt, các thư việc được sử dụng và hướng dẫn sử dụng MySQL.

Chương 4 gồm ý tưởng thiết kế chương trình, phân tích chương trình, một số kiểm thử trong khi chạy chương trình, 1 số lỗi thường gặp trong khi chạy và đánh giá tổng quan về kết quả của chương trình

# **Chương 2 Thiết kế**

## **2.1 Khảo sát hiện trạng**

Hiện nay, theo tôi tìm hiểu, chưa thấy có sản phẩm nào trên thị trường đáp ứng được việc đánh giá chính xác và xếp hạng các nhà cung cấp đám mây. Theo tôi đây là một đề tài còn mới và có nhiều hạn chế. Việc đánh giá chính xác nhà cung cấp nào tốt hơn nhà cung cấp nào tốt hơn hiện nay là một việc không dễ dàng. Đối với sản phầm hiện nay của tôi cũng chỉ mang tính đánh giá khách quan, đưa ra gợi ý cho người dùng chứ không thể hoàn toàn thay thế việc đánh giá của con người bởi vì có rất nhiều tính năng, thông số không rõ rang dẫn tới việc đánh giá chỉ mang tính chất tương đối.

Trong đồ án này của tôi, tôi phụ trách việc tính toán và phân tích dữ liệu sau khi nhận được dữ liệu từ bên phía crawler. Các chức năng được chia ra đợt trong 2 file đó là **ClusterPoint\_3.py**  với chức năng phân tích và cho điểm các thông số đã crawl về được và file manipulation\_3.py với các chức năng tương tác với người dùng. File manipulation\_3.py có chức năng tương tác với người dùng, hỏi người dùng các câu hỏi để xác định các yêu cầu của người dùng hoặc để mặc định theo hệ thống nếu người dùng không nhập và trả ra top 5 hoặc ít hơn các tùy chọn và nhà cung cấp tốt nhất mà đáp ứng được yêu cầu người dùng đưa ra.

## **2.2 Ý tưởng thực hiện**

Để đánh giá điểm của 1 tùy chọn, tôi sử dụng điểm đánh giá nhà cung cấp đó được lấy từ trang **getapp.com**, nơi đây có dữ liệu đánh giá về nhiều nhà cung cấp và tôi thấy là 1 nguồn dữ liệu có thể tham khảo để sử dụng được cùng với điểm trung bình của tùy chọn, lấy trung bình điểm của cả 2 để ra điểm trung của tùy chọn. Hiểu đơn giản là điểm trung bình giựa trên thông số của tùy chọn lấy trung với điểm đánh giá của nhà cung cấp. Điểm đánh giá của nhà cung cấp dựa trên trang **getapp.com** không mang tính tuyệt đối nhưng có nói lên 1 phần về độ mức độ tin cậy của nhà cung cấp cũng như đưa ra cái nhìn chung nhất về nhà cung cấp. Đối với những nhà cung cấp không tìm thấy trên trang, điểm số sẽ được tính theo trung bình cộng của tất cả các nhà cung cấp khác

Để tính điểm trung bình của mỗi tùy chọn được nói tới ở trên, tôi đánh giá các tiêu chí dựa trên Service Mesurement Index(SMI) như giá thành, thông số, tính linh hoạt của tùy chọn. Mỗi trường thông tin sẽ được đánh giá theo thang điểm từ 1 tới 5 bằng cách phân cụm các thông tin có thể tìm được từ nhà cung cấp. Các giá trị gần nhau trong cùng 1 trường dữ liệu sẽ được đánh theo cùng thang điểm tự tự động bằng việc sử dụng giải thuật phân cụm. Sau đó với mỗi trường đã được đánh giá điểm số, ta lấy trung bình cộng của tất cả các điểm số để ra được điểm trung bình của các tùy chọn

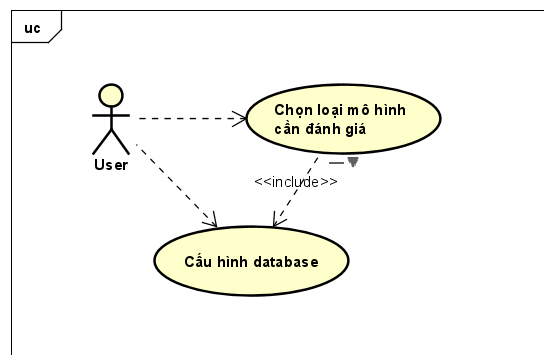
## **2.3 Tổng quan các chức năng**

Để chạy chương trình, có 2 theo tác chính. Thứ nhất là thao tác với MySQL và thứ hai là thao tác với python trên Visual Studio Code.

+ Trên MySQL cần cài đặt, thiết lập database cũng như nạp dữ liệu crawl được từ phần crawler vào trong database.

+ Trên Visual Studio Code, các code python cần được sửa lại để phù hợp với cấu hình của database. Đối với source code python gồm 2 file **ClusterPoint\_3.py**  và **manipulation\_3.py.** Cần sửa lại bên trong code ở cả 2 file để kết nối với data base. Trong đó file **ClusterPoint\_3.py**  có tác dụng đánh giá thang điểm của các nhà cung cấp, các tùy chọn và xử lí những dữ liệu chưa rõ ràng, file **manipulation\_3.py** để người dùng thao tác với chương trình, chọn loại nhà cung cấp cần theo nhu cầu, đặt ra các yêu cầu cần thiết để tìm nhà cung cấp phù hợp trong database

Biểu đồ usecase tổng quan cho chương trình trong **manipulation\_3.py**



Biểu đồ usecase của cả 2 chức năng tìm kiếm tùy chọn của lọc top 5 Cloud Storage option và Cloud Hosting option

Diagram

Description automatically generated

## **2.4 Đặc tả chức năng**

Cả hai chức năng tìm kiếm nhà cung cấp Cloud storage và cloud hosting đều có đắc tả giống nhau

Cả 2 chức năng của chương trình sẽ đều hỏi người dùng về những thông số kĩ thuật cơ bản mà người dùng cần thiết. Người dùng có thể bỏ qua phần này bằng cách bấm enter để bỏ quả câu hỏi và chương trình sẽ lấy giá trị mặc định được thiết lập sắn đễ sử dụng. Sau các thông số cơ bản, chương trình sẽ hỏi người dùng có muốn thêm thông số đặc biệt nào hay không, nếu có chương trình sẽ in ra bảng các loại thông số đặc biệt có trong cơ sở dữ liệu và dữ liệu in ra sẽ có các thông số đặc biệt được yêu cầu. Trường hợp không có tùy chọn nào phù hợp, thỏa mãn yêu cầu, chương trình sẽ in ra không có nhà cung cấp nào phù hợp.

***Đặc tả use case UC001 “***top 5 Cloud Storage option***”***

|  |  |  |  |
| --- | --- | --- | --- |
| **Mã Use case** | UC001 | **Tên Use case** | top 5 Cloud Storage option |
| **Tác nhân** | End user | | |
| **Mô tả** | Tìm ra top 5 Cloud Storage option từ database dựa trên tiêu chí của user | | |
| **Tiền điều kiện** | Kết nối thành công với database | | |
| **Luồng sự kiện**  **chính**  **(Thành công)** | **STT Thực hiện bởi** | **Hành động** | |
| 1. User | **Nhập thông số cấu hình yêu cầu hoặc nhấn Enter để lấy bỏ quả** | |
| 2. Hệ thống | **Trả về top 5 hoặc ít hơn số lượng tùy chọn đáp ứng yêu cầu hoặc báo người dùng không có tùy chọn nào đáp ứng cấu hình** | |

***Đặc tả use case UC002 “***top 5 Cloud Hosting option***”***

|  |  |  |  |
| --- | --- | --- | --- |
| **Mã Use case** | UC001 | **Tên Use case** | top 5 Cloud Hosting option |
| **Tác nhân** | End user | | |
| **Mô tả** | Tìm ra top 5 Cloud Hosting option từ database dựa trên tiêu chí của user | | |
| **Tiền điều kiện** | Kết nối thành công với database | | |
| **Luồng sự kiện**  **chính**  **(Thành công)** | **STT Thực hiện bởi** | **Hành động** | |
| 1. User | **Nhập thông số cấu hình yêu cầu hoặc nhấn Enter để lấy bỏ quả** | |
| 2. Hệ thống | **Trả về top 5 hoặc ít hơn số lượng tùy chọn đáp ứng yêu cầu hoặc báo người dùng không có tùy chọn nào đáp ứng cấu hình** | |

# **Chương 3. Công nghệ sử dụng**

## **3.1 Các công nghệ cần cài đặt**

Để sử dụng ứng dụng, trước tiên máy người dùng cần cài đặt những phần mềm này trước:

+ MySQL Server

+ MySQL Workbench

+ Visual studio code

+ Python3

\* Link hướng dẫn cài đặt:

MySQL và Workbench: <https://youtu.be/BYwb50Xbf8s>

Visual studio code: <https://youtu.be/YWODbdejecU>

Python3: <https://youtu.be/tYu8oHG3vqk>

\* Các thư viện kèm theo cần thiết để chạy chương trình trong python3:

+ mysql.connector – Để kết nối với MySQL database

+ sklearn -- Để phân cụm dữ liệu đầu vào làm điểm số đánh giá

+ numpy -- Để sắp xếp dữ liệu

Có thể cài đặt các thư việc trên bằng cách sử dụng trên **Visual studio code** bằng việc dùng câu lệnh: pip install <tên thư viện>

\* Có thể chạy câu lệnh sau thay thế để tải hết 1 lượt tất cả các thư viện

pip install mysql-connector-python sklearn numpy

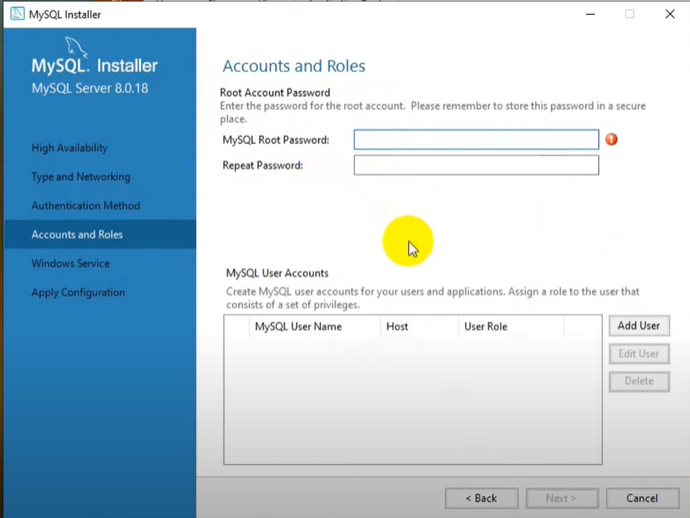
Trong quá trình cài đặt có thể xảy ra nhiều vấn đề lỗi phát sinh trong quá trình cài đặt các ứng dụng cần thiết để chạy chương trình, tùy thuộc vào các hệ điều hành, hướng dẫn trong Project 1 này được sử dụng trên hệ điều hành **Window**.

\*Lưu ý: Do hệ thống Crawler của Project 1 được cài đặt trên **Linux** trong khi hệ thống phân tích, đánh giá được cài trên **Window**, vậy nên file dữ liệu sau khi crawl được cần được gửi lên Cloud hoặc copy qua máy chạy hệ điều hành Window để thực hiện. Điều này xảy ra vì lí trong thời gian xây dựng và phát triển phần mềm đã có 1 số lỗi phát sinh trong quá trình cài đặt các ứng dụng nền tảng để phát triển phần mềm.

## **3.2 Hướng dẫn sử dụng công nghệ**

### **3.2.1 Hướng dẫn sử dụng MySQL**

\*Lưu ý: khi cài đặt cần nhớ rõ mật khẩu root lúc cài đặt server



Sau khi cài đặt xong workbench, mở file init.sql để chạy

Graphical user interface, text, application

Description automatically generated

Bôi đen dòng lệnh và chạy từng bước bằng cách bấm biểu tượng tia chớp theo từng bước như sau

B1: khởi tạo database

Graphical user interface, application

Description automatically generated

B2: chọn database crawler

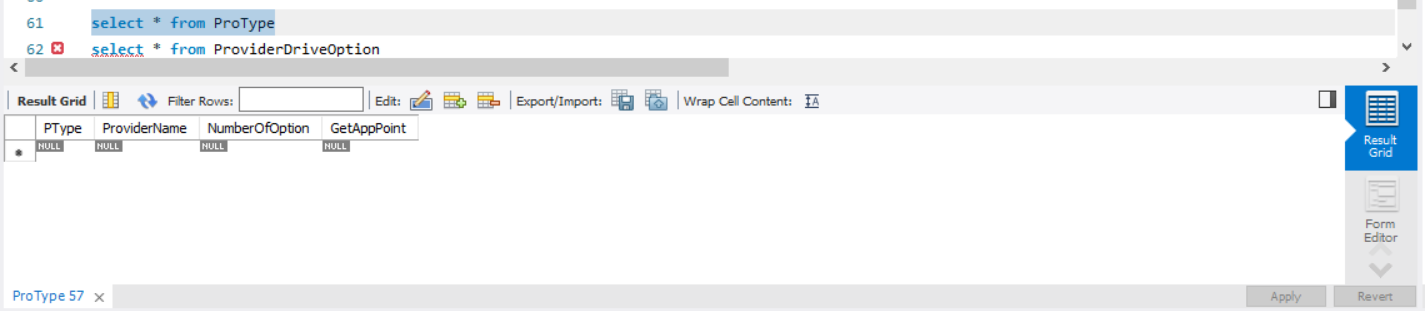
Graphical user interface, text, application

Description automatically generated

B3: Bôi đen hết cho tới trước phần select rồi chạy bên dưới để tạo bảngGraphical user interface, text, application

Description automatically generated

B4: select từng bảng, rồi chạy sau đó import file dữ liệu vào bằng nút 



Lựa chọn file tương ứng để importGraphical user interface

Description automatically generated

\*Lưu ý: có dữ liệu thử nghiệm mẫu để test chương trình trong thư mục test\_data, tên file là bảng tương ứng

Chọn bảng tương ứng

Graphical user interface, text, application, email

Description automatically generated

Kiểm tra lại hàng, cột tương ứng đúng hay chưa

Graphical user interface

Description automatically generated

Các bảng còn lại làm tương tự

Mẫu file .csv để import, cần theo đúng format của mỗi bản

A picture containing text, crossword puzzle

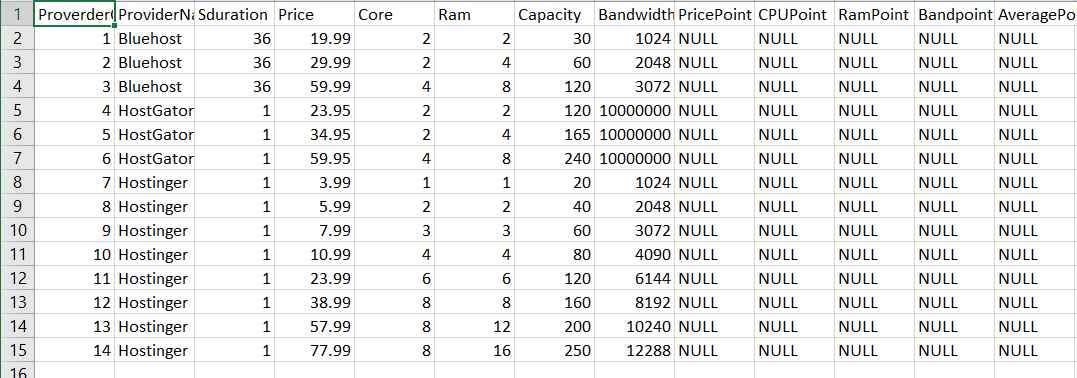
Description automatically generated

Bảng ProType

Table

Description automatically generated

Bảng InsertDrive



Bảng InsertHost

Table

Description automatically generated

Bảng MoreDriveOption

Table

Description automatically generated

Bảng MoreHostOption

# **Chương 4. Thực nghiệm và đánh giá**

## **4.1 Thiết kế**

### **4.1.1 Thiết kế cơ sở dữ liệu**

Diagram

Description automatically generated

**Biểu đồ EER của cơ sở dữ liệu**

Mô tả cơ sở dữ liệu:

Cơ sở dữ liệu gồm 5 bảng là

- **ProType**: lưu trữ tên của các nhà cung cấp đám và loại dịch vụ của nhà cung cấp

+ Ptype: loại cung dịch vụ của nhà cung cấp đó (Drive với Cloud Storage và host với web hosting

+ Providername: Tên nhà cung cấp

+ GetAppPoint: Điểm số của nhà cung cấp lấy được từ trang getapp.com

- **ProviderDriveOption**: Các tùy chọn của các nhà cung cấp cho Cloud Storage

+ ProverderOpID: Mã số định danh của tùy chọn trong cơ sở dữ liệu

+ Providername: Tên nhà cung cấp chiếu lên bảng **ProType**

+ MaxPeople: Số lượng người tối đa trong cùng lúc có thể sử dụng của tùy chọn

+ Sduration: Thời gian sử dụng tùy chọn, thời gian càng dài thông thường tùy chọn sẽ càng rẻ khi tính giá thành trên mỗi tháng

+ Price: giá thành của tùy chọn (tính theo $)

+ Capacity: dung lượng lưu trữ của tùy chọn

+ GbPerPrice: dung lượng lưu trữ tính trên 1$ của tùy chọn, mặc định bằng Capacity/Price

+ GbPerPricePoint: điểm số của trường GbPerPrice

+ MaxPeoplePoint: điểm số của trường MaxPeoplePoint

+ AveragePoint: Điểm trung bình của tùy chọn, tính bằng: (GbPerPricePoint+ MaxPeoplePoint)/2

- **MoreDriveOption**: những đặc chưng đặc biết của tùy chọn mà không theo khuôn mẫu chung

+ MoreOptID: Id của tùy chọn đặc biệt cho cloud storage

+ ProverderOpID: ID của tùy chọn có chứa đặc chưng đặc biệt này chiếu lên bảng **ProviderDriveOption**

+ Moption: đặc tả đặc chưng đặc biệt của tùy chọn

- **ProviderHostOption**: Các tùy chọn của các nhà cung cấp cho web hosting

+ ProverderOpID: Mã số định danh của tùy chọn trong cơ sở dữ liệu

+ Providername: Tên nhà cung cấp chiếu lên bảng **ProType**

+ Sduration: Thời gian sử dụng tùy chọn, thời gian càng dài thông thường tùy chọn sẽ càng rẻ khi tính giá thành trên mỗi tháng

+ Price: giá thành của tùy chọn (tính theo $)

+ Core: số lượng CPU core

+ Capacity: dung lượng lưu trữ trên máy server của tùy chọn

+ Bandwidth: dung lượng băng thông mỗi tháng của tùy chọn

+ PricePoint: điểm số của trường Price

+ CorePoint: điểm số của trường Core

+ RamPoint: điểm số của trường Ram

+ Bandwidthpoint: điểm số của trường Bandwidth

+ AveragePoint: Điểm trung bình của tùy chọn, tính bằng: (PricePoint+ CorePoint+ RamPoint +Bandwidthpoint)/4

- **MoreHostOption**: những đặc chưng đặc biết của tùy chọn mà không theo khuôn mẫu chung

+ MoreOptID: Id của tùy chọn đặc biệt cho hosting

+ ProverderOpID: ID của tùy chọn có chứa đặc chưng đặc biệt này chiếu lên bảng **Provider** chiếu lên bảng **ProType Option**

+ Moption: đặc tả đặc chưng đặc biệt của tùy chọn

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text

Description automatically generated

Câu lệnh tạo database, tạo bảng cần thiết ban đầu để khởi tạo database nằm trong file **init.sql. C**ần thiết chạy file này trước khi làm thêm bất cứ bước kế tiếp nào. Thực hiện bôi đen toàn file rồi bấm biểu tượng 

File **config.txt** bao gồm các cấu hình cần để chạy chương trình. File gồm 4 dòng cần được thêm vào trước khi chạy chương trình theo format:

Graphical user interface, text, application

Description automatically generated

Dòng 1: địa chỉ ip của database, thông thường để là local host khi server ở trên chính máy được cài đặt

Dòng 2: tên user, mặc định để là root

Dòng 3: mật khẩu của user, nếu user là root thì đây là mật khẩu cài đặt lúc cài đặt databasr

Dòng 4: tên của database xử dụng, mặc định là cralwer khi chạy theo câu lệnh trong **createtable.sql**

### **4.1.1 Xử lí dữ liệu trong Database**

Code trong file **ClusterPoint\_3.py :**

**Thư viện sử dụng và hàm kết nối với database:**

**Text

Description automatically generated**

Import các thư viện sử dụng và hàm kết nối với database (Thay thế user và password ở đây tùy theo cấu hình của server).

**Hàm Đánh giá điểm các tùy chọn:**

Hàm này có mục đích là gom dữ liệu đầu vào thành các cụm, những liệu trong cùng một cụm sẽ có chung điểm số. Các cụm này sẽ có dữ liệu được sắp xếp theo chiều tăng dần hoặc giảm dần và điểm của dữ liệu sẽ tăng dần từ 1 tới 5Text

Description automatically generated

Hàm **cluster\_and\_update** có dữ liệu đầu vào là: **tên trường của dữ liệu trong bảng, số lượng cluster hay thang điểm** (mặc định là 5), **thứ tự sắp** xếp (True nếu muốn điểm tăng dần theo giá dữ liệu và False nếu ngược lại**), kiểu dữ liệu** (0 nếu là float và 1 nếu là int), **tên bảng trong database.**

Hàm gồm các tính năng theo trình tự là:

+ Sắp xếp lại dữ liệu theo cách sắp xếp được lựa chọn, việc này quan trong trước khi tạo cluster bởi các cluster được tạo ra có label không được xếp theo trật tự.

+ Tạo các cluster dữ liệu, dù dữ liệu đã được sắp xếp nhưng cluster label của chúng thì không theo trật tự. Rồi cả giá trụ của dữ liệu và label của cluster sẽ được đóng gói vào chung trong 1 list trong **clustered\_data**.

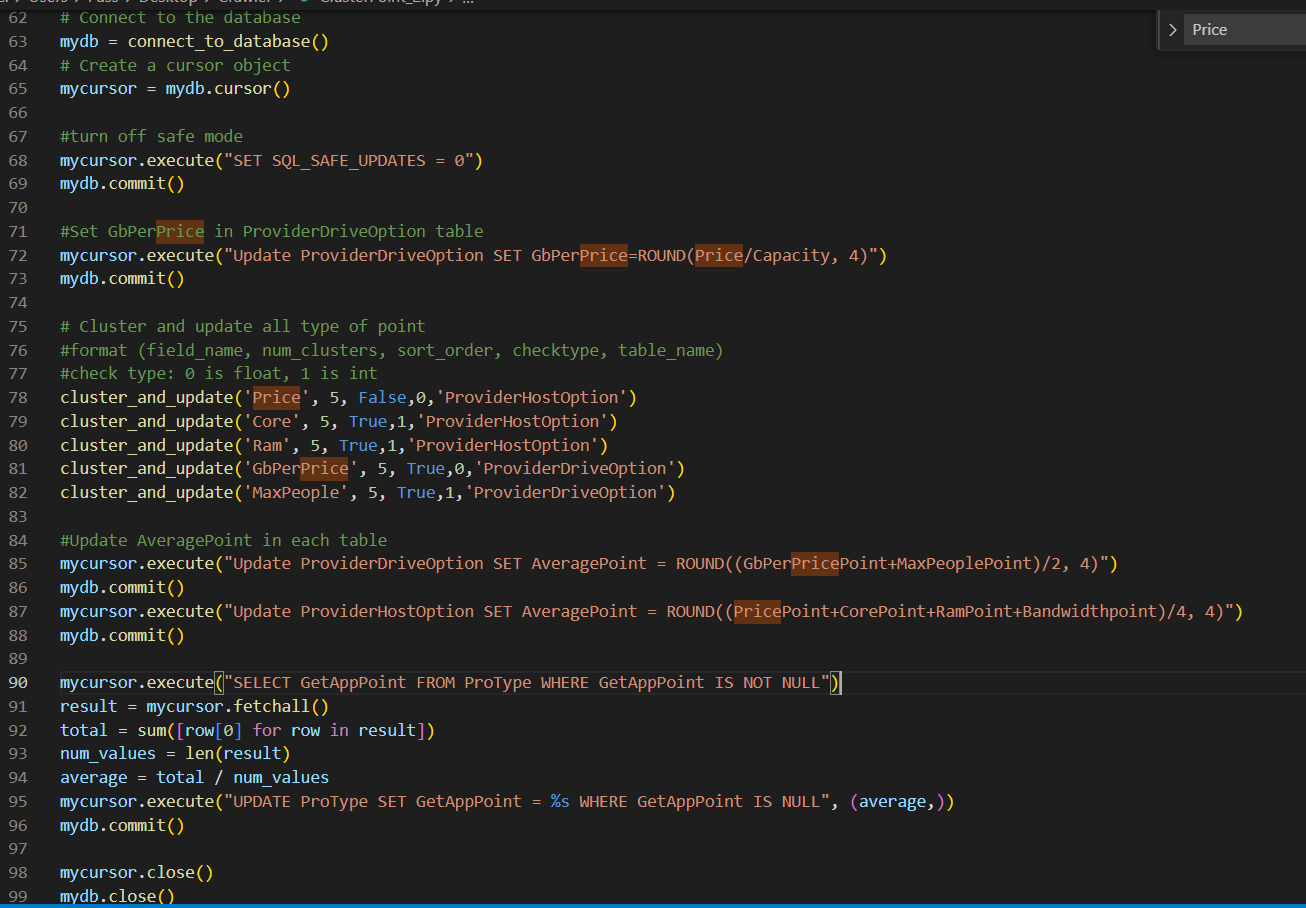
+ Đánh lại label trong **clustered\_data** , giá trị các dữ liệu đầu vào đã được sắp xếp theo 1 chiều nhất định, việc kế đến là sắp xếp label của các giá trị đó tăng dần. Ví dụ các label từ label đầu tiên đang là [0 0 0 3 3 2 2 2 4 4 6] sẽ được chuyển lại thành [0 0 0 1 1 2 2 2 3 3 4] và bây giờ label chính là điểm đã được đánh giá cho dữ liệu của trường dữ liệu đó.

Text

Description automatically generated

+ Cập nhật điểm đã đánh giá vào trong bảng, cần chú ý checktype đối với kiểu dữ liệu đầu vào tương ứng. 0 là float và 1 là int

**Hàm chính:**



Hàm chính gồm theo đúng tuần tự:

+ Gọi hàm kết nối với database, tạo cursor để đọc câu lệnh cho MySQL

+ Tắt safemode để có thể update được bình thường (nếu không tắt safemode thì sẽ xảy ra lỗi trong lần đầu tiên chạy chương trình và không update được)

+ Gọi hàm tính toán điểm cho các tùy chọn của các nhà cung cấp bằng hàm **cluster\_and\_update**

+ Tính điểm trung bình của các tùy chọn cho mỗi bảng

+ Thay thế những điểm là NULL trong bảng ProType bằng giá trị trung bình

+ Đóng kết nối

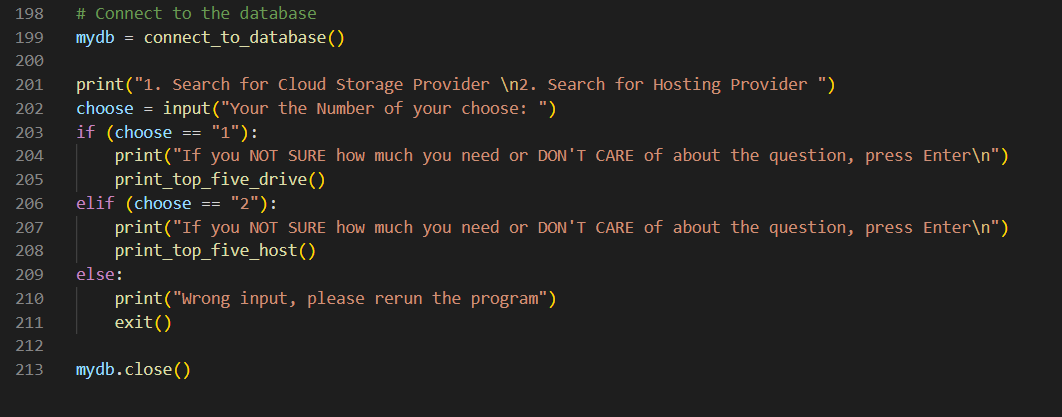
### **4.1.3 Tương tác người dùng và ứng dụng**

Code trong file **manipulation\_3.py:**

**Thư viện sử dụng:**



**Hàm chính:**



Các tính năng trong hàm chính:

+ Gọi hàm kết nối code với MySQL ( cần thay đổi user và password tùy theo cấu hình server)

+ In ra cho người dùng lựa chọn loại hình Cloud người dùng cần và chạy hàm tương ứng, nếu sai tùy chọn thì thoát ra ngoài

\* Nhắc nhở người dùng nếu có tùy chọn nào họ không quan tâm hoặc không rõ thì bấm phím Enter.

**Hàm kết nối MySQL:**

**Text

Description automatically generated**

**Hàm khi người dùng chọn tùy chọn là Cloud storage:**

**Text

Description automatically generated**

Các tính năng có trong hàm lần lượt là:

+ Hỏi người dùng những câu hỏi về những thông số cơ bản mà họ cần cho dịch vụ, nếu bấm enter, thông số sẽ tự cài theo default.

+ Hỏi người dùng có tùy chọn đặc biệt nào muốn hay không, nếu có, cho họ xem những tùy chọn đặc biệt có trong database và bảo người dùng nhập vào 1 tùy chọn có trong số tùy chọn được nêu ra

**Text

Description automatically generated**

+ Khai triển câu lệnh SQL tương ứng với việc người dùng có chọn thêm tùy chọn hay không, tìm kiếm những tùy chọn đáp ứng các điều kiện người dùng đặt ra và đưa ra các thông số của tùy chọn đó

Text

Description automatically generated

+ Trả kết quả, nếu không có kết quả nào phù hợp, quay về menu tùy chọn

+ Đếm xem trong database có tổng bao nhà cung cấp vào bao nhiều tùy chọn, và đưa dữ liệu này cho người dùng

+ Đưa ra không nhiều hơn 5 tùy chọn tốt nhất mà đáp ứng nhu cầu của người dùng, mỗi tùy chọn sẽ bao gồm tên nhà cung cấp, ID của tùy chọn, giá thành, dung lượng và điểm trung bình của tùy chọn dựa trên cách đánh giá của hệ thống

**Hàm khi người dùng chọn tùy chọn là Cloud hosting:**

**Text

Description automatically generated**

A screenshot of a computer

Description automatically generated

Các tính năng có trong hàm lần lượt là:

+ Hỏi người dùng những câu hỏi về những thông số cơ bản mà họ cần cho dịch vụ, nếu bấm enter, thông số sẽ tự cài theo default.

+ Hỏi người dùng có tùy chọn đặc biệt nào muốn hay không, nếu có, cho họ xem những tùy chọn đặc biệt có trong database và bảo người dùng nhập vào 1 tùy chọn có trong số tùy chọn được nêu ra

Text

Description automatically generated

+ Khai triển câu lệnh SQL tương ứng với việc người dùng có chọn thêm tùy chọn hay không, tìm kiếm những tùy chọn đáp ứng các điều kiện người dùng đặt ra và đưa ra các thông số của tùy chọn đó

Text

Description automatically generated

+ Trả kết quả, nếu không có kết quả nào phù hợp, quay về menu tùy chọn

+ Đếm xem trong database có tổng bao nhà cung cấp vào bao nhiều tùy chọn, và đưa dữ liệu này cho người dùng

+ Đưa ra không nhiều hơn 5 tùy chọn tốt nhất mà đáp ứng nhu cầu của người dùng, mỗi tùy chọn sẽ bao gồm tên nhà cung cấp, ID của tùy chọn, giá thành, số CPU core, ram, băng thôngdung lượng và điểm trung bình của tùy chọn dựa trên cách đánh giá của hệ thống

## **4.2 Kiểm thử**

Chương trình thực hiện trên terminal của Visual studio code

Chạy file **ClusterPoint\_3.py**  bằng câu lện python ClusterPoint\_3.py

Nếu chương trình báo connection fail, tức có lỗi kết nối code, cần kiểm tra lại phần này ở cả 2 file sorce code hoặc kiểm tra lại cấu hình của MySQL:

Text

Description automatically generated

Nếu connection successful nhưng báo lỗi như sau tức thiếu dữ liệu trong bảng hoặc thiếu bảng

Text

Description automatically generated

Chương trình khu chạy thành công:

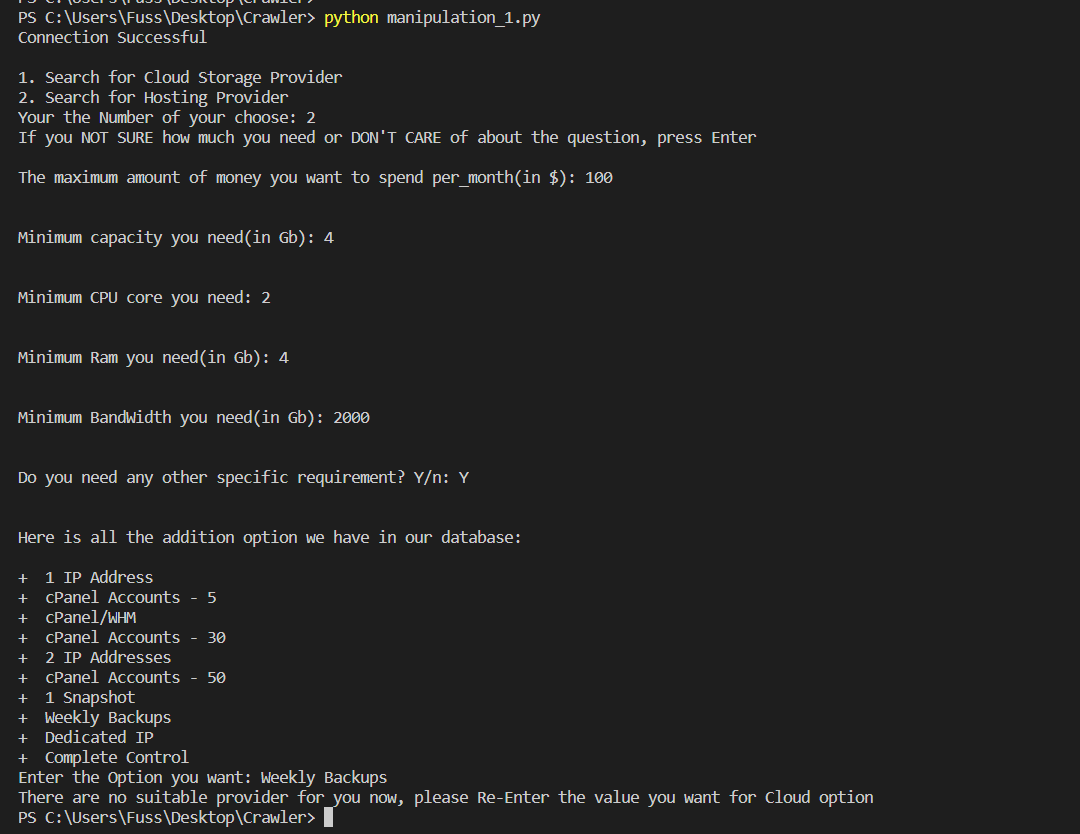
Graphical user interface, text

Description automatically generated

Chạy file **manipulation\_3.py** để tìm tùy nhà cung cấp và dịch vụ cung cấp đám mây người dùng cần:

Text

Description automatically generatedví dụ mẫu khi người dùng cần tìm nhà cung cấp lưu trữ đám mây mà không có điều kiện đi kèm



ví dụ mẫu khi người dùng cần tìm nhà cung cấp Host đám mây có điều kiện đi kèm

## **4.3 Đánh giá**

Chương trình có thể hoạt động được dưới những điều kiện nhất định, có thể có nhiều lỗi nảy sinh do người dùng không làm theo đúng các bước cần thực hiện. Chương trình này chỉ mang tính chất tham khảo để cho người dùng có thể có một cơ sở để đánh giá tham khảo xem nên lựa chọn nhà cung cấp nào tùy thuộc theo như cầu của người dùng chứ hoàn toàn thay thế được việc đánh giá thay thế con người vì vẫn còn nhiều biến số chưa rõ ràng. Cũng như còn có nhiều điều kiện khác chưa được đề cập đến trong chương trình.

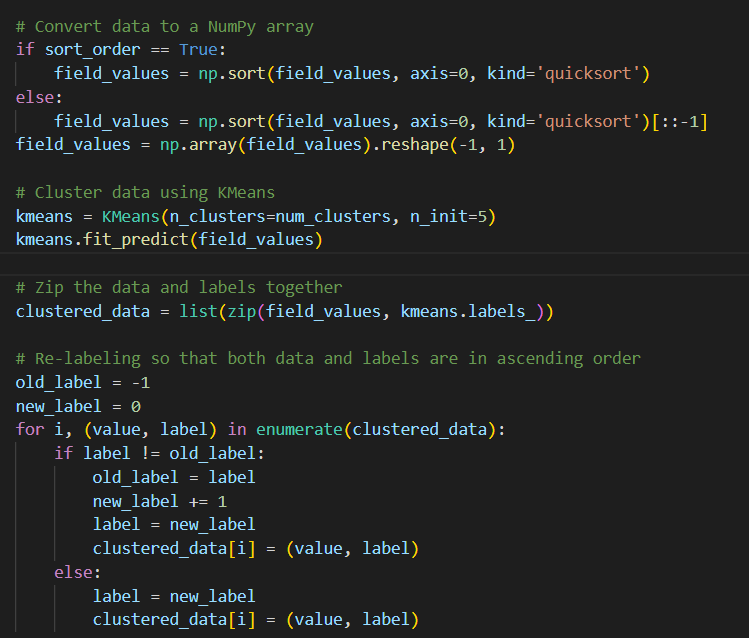
# **Chương 5. Các giải pháp và đóng góp nổi bật**

## **5.1.1. Đặt vấn đề**

Hiện tại, các dữ liệu sau khi được crawl về chưa có một hệ thống đánh giá, các Cloud Provider có nhưng tùy chọn, tính năng đặc biệt của riêng họ. Các dữ liệu đầu vào chưa đồng nhất, có những thông số phần lớn các nhà cung cấp đều có nhưng một vài nhà cung cấp thì lại không.

## **5.1.2. Giải pháp**

Sử dụng thư viện **numpy** của python để sort lại dữ liệu đầu vào, việc này giúp cho việc khi phân cụm theo **K-Means** bằng thư viện **sklearn**, các cluster khi phân cụm cạnh nhau sẽ có giá trị tăng dần hoặc giảm dần, khi này ta đánh lại label theo chiều tăng dần vì khi tạo cluster thì label của mỗi cluster là ngẫu nhiên.



Đối với các giá trị đầu vào không thể tìm thấy của Cloud Provider thì được lấy bằng giá trị trung bình của các Cloud Provider và các tùy chọn của các Cloud Provider khác.

# **Chương 6. Kết luận và hướng phát triển**

Đề tài này là một đề tài còn mới, quá trình phát triển còn nhiều khó khăn với lí do chính là về nhân lực. Nguồn nhân lực còn thiếu kinh nghiệm cùng với các nguồn tài liệu để tham khảo còn ít dẫn tới việc chất lượng sản phẩm còn thiếu tính chính xác, có thể phát sinh nhiều lỗi phát sinh trong quá trình cài đặt và sử dụng. Giao diện của chương trình còn ở mức sơ khai, chưa thân thiện với người dùng. Các tính năng vẫn còn thô, tính chính xác vẫn còn chưa cao. Đề tài vẫn còn nhiều điểm có thể khai thác nhưng vì lí do thời gian và nhân lực nên chưa thể hoàn thiện vậy nên ở đề tài này vẫn còn nhiều điểm có thể nâng cấp, cải tiến.

**​**

​

​

​

​

​

​

​

​

​

​

​

**​**

​

​

​

​

​

​

​

​

**​**

​

​

**​**

**​**