

**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**  
**KHOA HỌC MÁY TÍNH**



**KHAI THÁC DỮ LIỆU VÀ ỨNG DỤNG – CS313.N22**

*Bài tập 1: Tìm hiểu ngôn ngữ lập trình Python và môi trường phát triển Google Colab*

Nhóm 3:

Bùi Nguyễn Anh Trung (NT)	20520332
Hồ Thanh Tịnh	20520813
Nguyễn Trần Minh Anh	20520394
Lê Nguyễn Bảo Hân	20520174
Nguyễn Văn Đức Ngọc	20521666

*Hồ Chí Minh, 13 tháng 03 năm 2023*

## Nội dung

<b>I. Ngôn ngữ lập trình Python</b>	<b>4</b>
1. Khái niệm	4
2. Ứng dụng	5
3. Hạn chế	6
4. Lợi thế ( so với các ngôn ngữ khác )	7
5. Cài đặt	8
6. Sử dụng	9
<b>II. Môi trường phát triển Google Colab</b>	<b>10</b>
1. Khái niệm	10
2. Chức năng	10
3. Lợi thế khi so sánh với các môi trường khác	11
4. Hạn chế	11
5. Sử dụng	12
<b>III. Tham khảo</b>	<b>15</b>

Bảng phân công

<div>Thành viên</div> <div>Công Việc</div>	Bùi Nguyễn Anh Trung	Hồ Thanh Tịnh	Nguyễn Trần Minh Anh	Lê Nguyễn Bảo Hân	Nguyễn Văn Đức Ngọc
Tạo sườn báo cáo		✓			
Viết báo cáo			✓		
Chỉnh sửa nội dung, định dạng báo cáo	✓				
Làm slide				✓	
Thuyết trình		✓			
Soạn chương trình chạy thử trên Colab					✓
Phân công, quản lý công việc chung	✓				
Mức độ hoàn thiện ( %)	100%	100%	100%	100%	100%

# I. Ngôn ngữ lập trình Python



*Hình 1: Ngôn ngữ lập trình Python*

## 1. Khái niệm

- ❖ Python là ngôn ngữ lập trình thông dịch, đa mục đích và đối tượng, cú pháp đơn giản, dễ đọc và dễ hiểu [1].
- ❖ Nó được phát triển vào cuối thập niên 1980 bởi Guido van Rossum tại Hà Lan.
- ❖ Python là một trong những ngôn ngữ lập trình phổ biến nhất trên thế giới và được sử dụng rộng rãi trong khoa học dữ liệu, trí tuệ nhân tạo, phát triển web, và nhiều lĩnh vực khác.

## 2. Ứng dụng



*Hình 2: Các thư viện được hỗ trợ trong Python*

Ứng dụng chính của Python trong lĩnh vực phân tích dữ liệu bao gồm:

- ❖ Xử lý dữ liệu:
  - Python có các thư viện mạnh mẽ như NumPy và Pandas
  - Để xử lý các tập dữ liệu lớn, tính toán và phân tích thống kê.
- ❖ Trực quan hóa dữ liệu:
  - Python cung cấp các thư viện như Matplotlib và Seaborn
  - Để trực quan hóa dữ liệu, giúp hiểu rõ hơn về tính chất và mối quan hệ giữa các biến.
- ❖ Học máy và khai thác dữ liệu:
  - Python có các thư viện mạnh mẽ như Scikit-Learn và TensorFlow
  - Để thực hiện các tác vụ học máy, khai phá dữ liệu và xây dựng mô hình.

### 3. Hạn chế

Những điểm hạn chế của python có thể kể đến:

❖ Tốc độ thực thi:

- Python chậm hơn các ngôn ngữ lập trình khác như C++ và Java, đặc biệt là khi xử lý các tác vụ yêu cầu tính toán nặng.
- Tuy nhiên, điều này đã được cải thiện đáng kể với sự xuất hiện của các thư viện tối ưu hóa như NumPy, SciPy, Pandas và TensorFlow.

❖ Bảo mật:

- Python không phải là ngôn ngữ lập trình được thiết kế để có tính bảo mật cao.
- Nếu không được xử lý cẩn thận, các ứng dụng Python có thể dễ dàng bị tấn công bởi các lỗ hổng bảo mật như injection SQL, cross-site scripting (XSS) và buffer overflow.

❖ Thư viện hạn chế:

- Mặc dù có nhiều thư viện Python có sẵn để giải quyết các vấn đề khác nhau, nhưng vẫn có một số lĩnh vực mà Python không có những thư viện tốt nhất để xử lý.
  - Ví dụ, Python không phải là ngôn ngữ phù hợp nhất để phát triển các ứng dụng trò chơi, ứng dụng di động, hoặc ứng dụng yêu cầu độ trễ thấp và xử lý đa luồng cao.

❖ Giao diện đồ họa:

- Mặc dù Python có thể được sử dụng để phát triển các ứng dụng desktop, nhưng nó không cung cấp nhiều công cụ để tạo giao diện đồ họa,
- Điều này làm cho việc phát triển các ứng dụng desktop với Python trở nên khó khăn hơn so với các ngôn ngữ lập trình khác như Java hoặc C#.

## 4. Lợi thế ( so với các ngôn ngữ khác )

So với các ngôn ngữ lập trình khác, Python có những ưu điểm sau trong lĩnh vực phân tích dữ liệu:

- ❖ Dễ học và đọc:
  - Cú pháp của Python đơn giản và rõ ràng
  - Giúp cho người mới bắt đầu dễ dàng tiếp cận và phát triển các ứng dụng phân tích dữ liệu.
- ❖ Tính linh hoạt:
  - Python có thể được sử dụng cho nhiều mục đích khác nhau, từ viết script đến phát triển các ứng dụng phức tạp
  - Giúp cho các nhà phân tích dữ liệu có thể sử dụng Python cho nhiều tác vụ khác nhau.
- ❖ Hỗ trợ mạnh mẽ từ cộng đồng:
  - Python có một cộng đồng lớn các nhà phát triển và chuyên gia phân tích dữ liệu, cung cấp các thư viện và công cụ mạnh mẽ để
  - Giúp cho các nhà phân tích dữ liệu thực hiện các tác vụ phân tích và khai thác dữ liệu hiệu quả hơn.



## 5. Cài đặt



*Hình 3 : Version mới nhất là Python3.11*

- ❖ Có nhiều phiên bản của Python như Python 2.x và Python 3.x
- ❖ Tuy nhiên, để tận dụng được các tính năng mới nhất và được hỗ trợ tốt nhất, bạn nên sử dụng Python 3.x.

Bạn có thể tải phiên bản mới nhất của Python tại trang chủ của Python: [Python Downloads](#) hoặc cài đặt trên MS store.

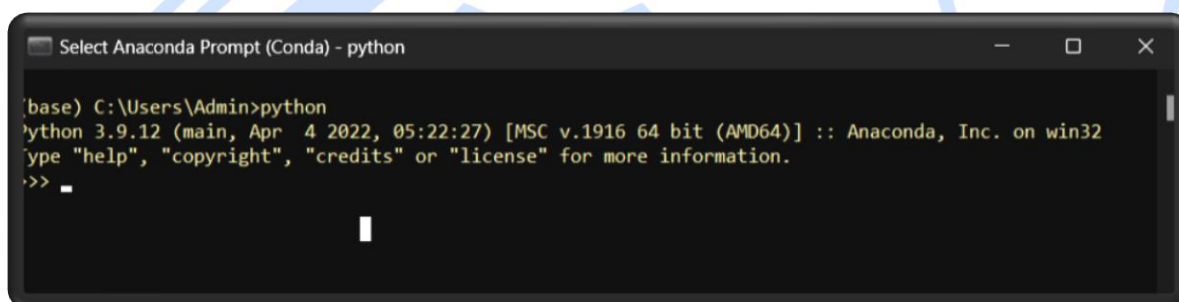


## 6. Sử dụng

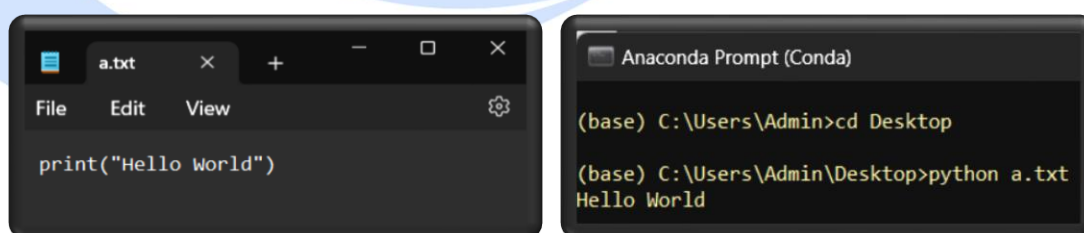
❖ Để sử dụng Python, cần thực hiện các bước sau:

- Bước 1: Kiểm tra phiên bản
- Bước 2:
  - Khởi động Terminal (cmd) gõ *Python*.
  - Hoặc sử dụng file text bất kỳ để ghi script sau đó sử dụng terminal để thực thi
- Bước 3: Trong terminal thực thi file bằng cách gõ: *python <tên file script>*

Hoặc có thể sử dụng các compiler như Visual Code, Google Colab,...



Hình 4: Khởi động Terminal và gõ `python`



Hình 5 - 6: Tạo file *a.txt*, ghi lệnh `print("Hello World")` và lưu lại. Sau đó thực thi file bằng lệnh `python a.txt`

## II. Môi trường phát triển Google Colab

### 1. Khái niệm

- ❖ Google Colab là một môi trường phát triển tích hợp (IDE) miễn phí và trực tuyến được cung cấp bởi Google.
- ❖ Cho phép người dùng tạo và chia sẻ các tài liệu Jupyter (Jupyter notebooks) chứa mã Python, phân tích dữ liệu và trực quan hóa dữ liệu, và cho phép chạy các đoạn mã trên các máy chủ được quản lý bởi Google [2] [3].

### 2. Chức năng

Một số chức năng và ứng dụng của Google Colab bao gồm:

- ❖ Hỗ trợ các thư viện Python phổ biến, bao gồm cả các thư viện phân tích dữ liệu như Pandas và NumPy.
- ❖ Cho phép lưu trữ và tải lên các tệp notebook Jupyter lên Google Drive.
- ❖ Cung cấp các tài nguyên tính toán đám mây miễn phí, bao gồm CPU, GPU và TPU, để phục vụ cho các nhu cầu phân tích dữ liệu và học máy.
- ❖ Cho phép chạy các lệnh shell và cài đặt các gói phần mềm bên ngoài, giúp người dùng mở rộng khả năng của Colab.
- ❖ Cho phép chia sẻ và làm việc đồng thời trên cùng một notebook với nhiều người dùng khác nhau.

### 3. Lợi thế khi so sánh với các môi trường khác

So với các môi trường lập trình khác trong lĩnh vực phân tích dữ liệu, Google Colab có những ưu điểm sau:

- ❖ Miễn phí và dễ sử dụng, không yêu cầu cài đặt phần mềm hoặc cấu hình môi trường phức tạp.
- ❖ Cho phép sử dụng tài nguyên đám mây mạnh mẽ để tính toán, bao gồm cả GPU và TPU, giúp tăng tốc các quá trình tính toán phức tạp.
- ❖ Cho phép chia sẻ và làm việc đồng thời trên cùng một notebook với nhiều người dùng khác nhau.
- ❖ Có sẵn các thư viện Python phổ biến trong lĩnh vực phân tích dữ liệu, giúp người dùng tiết kiệm thời gian và nỗ lực trong việc cài đặt và cấu hình.

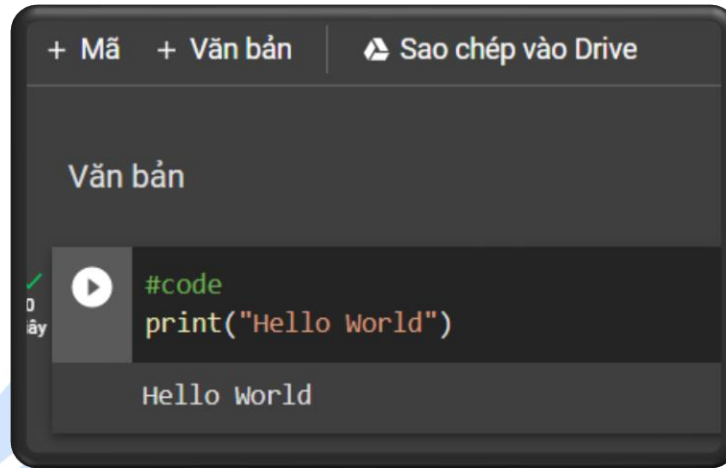
### 4. Hạn chế

Tuy nhiên, Google Colab cũng có một số hạn chế, bao gồm:

- ❖ Không được thiết kế để xử lý các tác vụ tính toán quy mô lớn hoặc tài nguyên yêu cầu cao.
- ❖ Môi trường đóng: người dùng Google Colab chỉ có thể sử dụng các thư viện có sẵn và cài đặt thêm một số thư viện cần thiết khác, nên không thể sử dụng thư viện cá nhân của người dùng hay các công cụ cần thiết cho các tác vụ mang tính chuyên môn hơn.
- ❖ Không thể truy cập trực tiếp vào các tệp dữ liệu được lưu trữ trên máy.
- ❖ Các dữ liệu, biến mà chương trình tạo ra được chỉ tồn tại trong khoảng thời gian còn kết nối với môi trường, một khi thoát kết nối thì toàn bộ dữ liệu sẽ bị xóa tự động.
- ❖ Google Colab cũng không cho phép trong một notebook có từ hơn 2 người code cùng một lúc, điều này khá bất tiện trong cho các dự án làm việc nhóm.

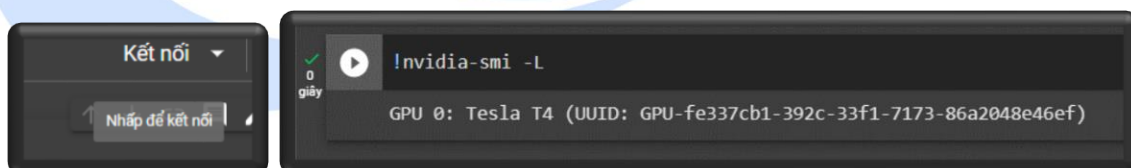
## 5. Sử dụng

- ❖ Google Colab được sử dụng như một compiler với cấu trúc là một notebook bao gồm nhiều cell chứa code hoặc văn bản.



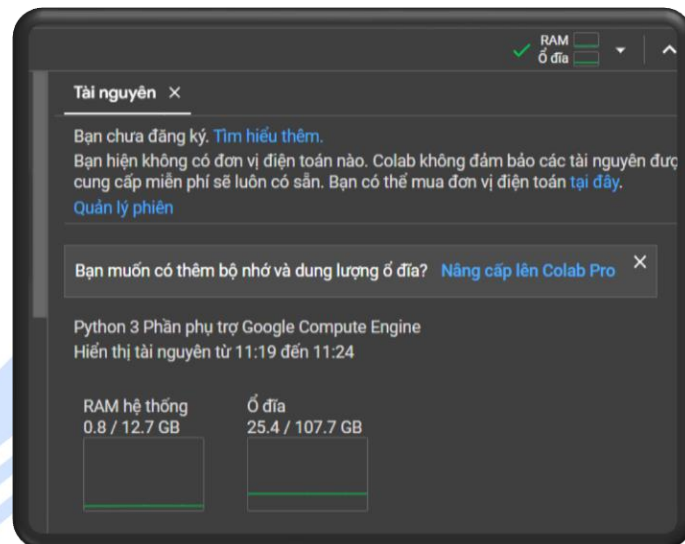
Hình 7 : Trong cell của Google Colab. Gõ lệnh `print("Hello World")` và nhấn nút biểu tượng Play để thực thi. Kết quả được hiện ra ngay dưới cell

- ❖ Google Colab sử dụng tài nguyên là GPU miễn phí có sẵn từ Google, nên sẽ có hạn chế về môi trường, tài nguyên và thời gian sử dụng đối với từng tài khoản. Để chương trình có thể thực thi, trước hết cần phải kết nối notebook đang sử dụng với môi trường để tiến hành thực thi code từ các cell.



Hình 8-9 : Kiểm tra GPU.

- ❖ Riêng RAM và dung lượng ổ đĩa sẽ được sử dụng dựa trên máy tính cục bộ .



Hình 10 : Thông số tài nguyên của máy cục bộ

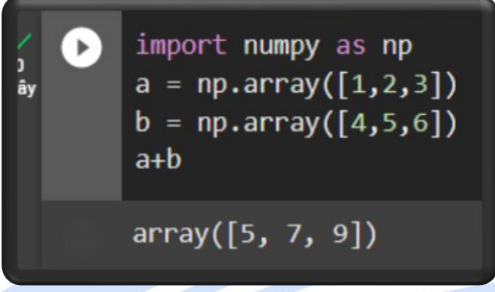
- ❖ Để có thể sử dụng các thư viện, trước hết ta cần cài đặt các thư viện đó bằng cách sử dụng terminal như pip với cú pháp !

```
!pip install tensorflow
```

```
Looking in indexes: https://pypi.org/simple, https://us-python
Requirement already satisfied: tensorflow in /usr/local/lib/p
Requirement already satisfied: keras<2.12,>=2.11.0 in /usr/lo
Requirement already satisfied: h5py>=2.9.0 in /usr/local/lib/
Requirement already satisfied: tensorflow-estimator<2.12,>=2.
Requirement already satisfied: packaging in /usr/local/lib/py
Requirement already satisfied: wrapt>=1.11.0 in /usr/local/li
Requirement already satisfied: typing-extensions>=3.6.6 in /u
Requirement already satisfied: termcolor>=1.1.0 in /usr/local
```

Hình 11 : Cài đặt thư viện bằng lệnh 'pip' và nên bắt đầu với cú pháp dấu '!'

- ❖ Và để sử dụng thư viện trong notebook. Ta buộc phải khai báo thư viện bằng lệnh **import**

A screenshot of a Jupyter Notebook cell. The code inside the cell is: 

```
import numpy as np  
a = np.array([1,2,3])  
b = np.array([4,5,6])  
a+b
```

 The output of the cell, displayed below the code, is: 

```
array([5, 7, 9])
```

 The notebook interface includes a green checkmark and a play button icon on the left side of the cell.

```
import numpy as np  
a = np.array([1,2,3])  
b = np.array([4,5,6])  
a+b  
  
array([5, 7, 9])
```

Hình 12 : Khai báo sử dụng thư viện numpy ( với cú pháp viết tắt là `np` ). Do thư viện này có sẵn trong Google Colab nên không cần phải cài đặt bằng pip từ ban đầu



### III. Tham khảo

#### Nguồn tham khảo

- [1] Python Software Foundation, 2001-2023. [Online]. Available: <https://www.python.org/doc/>.
- [2] "Colaboratory," Google, [Online]. Available: <https://research.google.com/colaboratory/faq.html>.
- [3] "Welcome to Colab!," Google, [Online]. Available: <https://colab.research.google.com/#>.

