

# Python cho Khoa học dữ liệu

## Bài 01: Tổng quan Python

Hà Minh Tuấn

Khoa Toán - Tin học  
Trường Đại học Khoa học Tự nhiên, ĐHQG-HCM

Ngày 28 tháng 8 năm 2025

# Nội dung bài học

- 1 Giới thiệu về Python
- 2 Lịch sử và Các đặc điểm chính của python
- 3 Hello World Program in Python
- 4 Giới thiệu các phiên bản mới

# Giới thiệu Python

**Python** là một trong những ngôn ngữ lập trình phổ biến nhất hiện nay. Mặc dù là ngôn ngữ lập trình đa mục đích, Python được ứng dụng rộng rãi trong *Machine Learning*, *AI*, phát triển web, IoT, và nhiều lĩnh vực khác.

Tài liệu này dành cho người mới bắt đầu, giúp bạn nắm vững từ các khái niệm cơ bản đến nâng cao của Python. Sau khi hoàn thành, bạn sẽ đạt được nền tảng vững chắc để phát triển thành một lập trình viên chuyên nghiệp.

**Phiên bản:** Python 3.13

**Python là gì?** Python là ngôn ngữ lập trình thông dịch, tương tác, hướng đối tượng, và bậc cao. Được Guido van Rossum phát triển (1985–1990), Python hỗ trợ nhiều mô hình lập trình: thủ tục, hướng đối tượng, và hàm. Triết lý thiết kế của Python nhấn mạnh tính dễ đọc với việc sử dụng thụt dòng rõ ràng.

# Ví dụ Python

## Ví dụ code Python

```
def fibonacci(n):  
    """Trả về dãy Fibonacci tới n"""  
    a, b = 0, 1  
    for _ in range(n):  
        print(a, end=" ")  
        a, b = b, a + b  
    return a,b  
  
fibonacci(10)
```

# Lịch sử Python

**Người tạo:** Guido van Rossum

**Thời gian phát triển:** 1985 – 1990

**Giấy phép:** GNU GPL

## Đặc điểm chính:

- Đa năng, thông dịch, hướng đối tượng, cấp cao.
- Thiết kế nhấn mạnh *khả năng đọc mã* (thực lẽ có ý nghĩa).
- Mã nguồn mở và miễn phí.
- Phù hợp cho người mới bắt đầu, cú pháp đơn giản.
- Hỗ trợ đa mô hình lập trình (thủ tục, OOP, hàm).

**Tóm lại:** Python là ngôn ngữ lập trình mạnh mẽ, linh hoạt và dễ tiếp cận, hiện nay được sử dụng rộng rãi trong nhiều lĩnh vực.

# Giới thiệu và các đặc điểm chính

Python là một ngôn ngữ kịch bản giàu tính năng, bậc cao, thông dịch, tương tác và hướng đối tượng. Python là một ngôn ngữ lập trình đa năng và rất phổ biến nhờ các đặc điểm như dễ đọc, đơn giản, thư viện phong phú, và nhiều hơn nữa.

## 1. Dễ Học (Easy to Learn)

Đây là một trong những lý do quan trọng nhất cho sự phổ biến của Python. Cú pháp đơn giản, việc sử dụng thụt lề để tránh sự lộn xộn của dấu ngoặc nhọn và kiểu dữ liệu động không yêu cầu khai báo biến trước giúp người mới bắt đầu học Python một cách nhanh chóng và dễ dàng.

## 2. Kiểu Dữ Liệu Động (Dynamically Typed)

Python là một ngôn ngữ lập trình có kiểu dữ liệu động. Trong Python, bạn không cần chỉ định kiểu dữ liệu tại thời điểm khai báo biến. Kiểu dữ liệu được xác định tại thời điểm chạy dựa trên giá trị được gán.

# Các đặc điểm chính

## 3. Dựa trên Trình thông dịch (Interpreter Based)

Python là một ngôn ngữ thông dịch. Trình thông dịch lấy từng lệnh từ mã nguồn, dịch nó sang mã máy và thực thi ngay lập tức. Điều này giúp gỡ lỗi chương trình dễ dàng hơn và rất hữu ích cho lập trình viên mới bắt đầu.

## 4. Tương tác (Interactive)

Python đi kèm với một môi trường dòng lệnh tương tác (shell) hoạt động theo nguyên tắc REPL (Đọc - Đánh giá - In - Lặp). Bạn có thể gõ bất kỳ biểu thức Python hợp lệ nào tại dấu nhắc `>>>` và nhận kết quả ngay lập tức.

## 5. Đa mô hình lập trình (Multi-paradigm)

Python là ngôn ngữ hoàn toàn hướng đối tượng. Tuy nhiên, nó cũng hỗ trợ lập trình mệnh lệnh, thủ tục và có các tính năng giống với lập trình hàm.

# Các đặc điểm chính

## 6. Thư viện chuẩn (Standard Library)

Python được phân phối với một thư viện chuẩn gồm rất nhiều module và gói, hỗ trợ các nhu cầu lập trình như xử lý dữ liệu Internet, nén dữ liệu, v.v. Python nổi tiếng với phương pháp "có sẵn pin" (batteries included).

## 7. Mã Nguồn Mở và Đa Nền tảng (Open Source and Cross Platform)

Python là mã nguồn mở và miễn phí. Nó là một ngôn ngữ đa nền tảng, có nghĩa là một chương trình Python có thể chạy trên nhiều hệ điều hành khác nhau như Windows, Linux, Mac OS mà không cần thay đổi.

## 8. Ứng dụng Giao diện Đồ họa (GUI Applications)

Thư viện chuẩn của Python có một thư viện đồ họa xuất sắc tên là Tkinter. Ngoài ra, nhiều bộ công cụ GUI khác như PyQt, WxWidgets cũng có thể được sử dụng để xây dựng các ứng dụng GUI thân thiện với người dùng.



## 10. Khả năng mở rộng (Extensible)

Mã nguồn Python có thể được tích hợp với các đoạn mã viết bằng ngôn ngữ khác như C hoặc C++. Điều này cho phép người dùng viết các module/thư viện hiệu năng cao bằng C và kết hợp chúng vào thư viện chuẩn.

## 11. Cộng đồng Nhà phát triển Năng động (Active Community)

Nhờ sự phổ biến và bản chất mã nguồn mở, Python có một cộng đồng nhà phát triển lớn và năng động. Các công ty lớn như Google, Microsoft và Meta cũng đóng góp rất nhiều vào hệ sinh thái này.

## Các tính năng nổi bật khác

- Hỗ trợ lập trình hàm, cấu trúc và hướng đối tượng (OOP).
- Có thể được dùng làm ngôn ngữ kịch bản hoặc biên dịch thành bytecode.
- Hỗ trợ thu gom rác (garbage collection) tự động.
- Dễ dàng tích hợp với C, C++, COM, ActiveX, CORBA, và Java.

# So Sánh Python và C++

Tiêu chí	Python	C++
Cấp độ	Ngôn ngữ bậc cao (gần với ngôn ngữ con người, trừu tượng hóa cao).	Ngôn ngữ bậc trung (gần với phần cứng, cho phép lập trình cấp thấp).
Nguồn gốc	Phát hành năm 1991 bởi Guido van Rossum.	Tạo ra trong thập niên 1980 bởi Bjarne Stroustrup.
Lĩnh vực chính	Phát triển Web, Khoa học dữ liệu, Machine Learning, tự động hóa.	Phần mềm hệ thống, phát triển game, lập trình nhúng, ứng dụng hiệu năng cao.

# So Sánh Python và C++

Tiêu chí	Python	C++
<b>Quản lý bộ nhớ</b>	Tự động. Lập trình viên không cần quan tâm nhiều đến việc cấp phát, giải phóng bộ nhớ.	Thủ công. Lập trình viên có toàn quyền và trách nhiệm kiểm soát bộ nhớ.
<b>Đối tượng</b>	Phù hợp cho người mới bắt đầu, các dự án cần ra sản phẩm nhanh.	Thường dành cho lập trình viên có kinh nghiệm, các dự án đòi hỏi hiệu năng tối đa.

*Kết luận: Mỗi ngôn ngữ có điểm mạnh riêng và lựa chọn ngôn ngữ nào phụ thuộc hoàn toàn vào yêu cầu của dự án.*

# Python Example

```
a = int(input("Enter value for a"))  
b = int(input("Enter value for b"))  
  
print("The number you have entered for a is ", a)  
print("The number you have entered for b is ", b)
```

# Bước 1: Cài đặt Python

- Để bắt đầu lập trình Python, việc đầu tiên là **\*\*cài đặt trình thông dịch Python\*\*** trên máy tính của bạn.
- **\*\*Tải Python:\*\***
  - ▶ Truy cập trang web chính thức: <https://www.python.org/downloads/>.
  - ▶ Tải xuống phiên bản Python mới nhất phù hợp với hệ điều hành của bạn (Windows, Linux, macOS, Android OS).
  - ▶ Python là ngôn ngữ **\*\*mã nguồn mở và đa nền tảng\*\***.
- **\*\*Cài đặt:\*\***
  - ▶ Chạy file cài đặt và làm theo hướng dẫn. Hãy đảm bảo bạn chọn tùy chọn "Add Python to PATH" (hoặc tương tự) trong quá trình cài đặt trên Windows để dễ dàng sử dụng từ Command Prompt.
- **\*\*Kiểm tra cài đặt:\*\***
  - ▶ Mở **\*\*Terminal\*\*** (Linux/macOS) hoặc **\*\*Command Prompt\*\*** (Windows).
  - ▶ Gõ lệnh: `python --version` hoặc `python3 --version` và nhấn Enter.
  - ▶ Nếu Python đã được cài đặt thành công, bạn sẽ thấy phiên bản Python hiện ra (ví dụ: 'Python 3.11.2').

## Bước 2: Chương trình Python đầu tiên - "Hello World"

- Chương trình "Hello World" là một chương trình cơ bản dùng để **kiểm tra cài đặt** và **làm quen với cú pháp** của một ngôn ngữ lập trình.
- Trong Python, bạn có thể dễ dàng in một thông điệp ra màn hình bằng cách sử dụng hàm `print()`.
- **Cú pháp đơn giản:**

```
##Python code to print "Hello, World!"  
print ("Hello, World!")
```

- Hàm `print()` là một hàm **có sẵn** (built-in function) của Python, được dùng để hiển thị dữ liệu ra console (đầu ra tiêu chuẩn).
- Chuỗi văn bản muốn in ra được đặt trong dấu ngoặc kép (""") hoặc dấu ngoặc đơn ('').
- Trong Python, các dòng bắt đầu bằng dấu `#` là **chú thích** và sẽ bị trình thông dịch bỏ qua.

## Bước 3: Chạy chương trình Python đầu tiên

- Có hai cách phổ biến để chạy chương trình Python:
- **1. Sử dụng chế độ tương tác (Interactive Mode):**
  - ▶ Mở Terminal/Command Prompt.
  - ▶ Gõ lệnh `python` (hoặc `python3`) và nhấn Enter để vào chế độ Python Interpreter (REPL - Read Evaluate Print Loop).
  - ▶ Tại dấu nhắc '»>', gõ lệnh: `print("Hello World")` và nhấn Enter.
  - ▶ Bạn sẽ thấy kết quả `Hello World` được hiển thị ngay lập tức.
  - ▶ Chế độ này **\*\*rất hữu ích\*\*** để thử nghiệm các đoạn mã nhỏ hoặc khám phá chức năng của thư viện.
- **2. Chạy từ file script (.py):**
  - ▶ **Tạo file:** Mở một trình soạn thảo văn bản (ví dụ: Notepad, VS Code) hoặc IDE. Gõ dòng `print("Hello World")` vào file.
  - ▶ **Lưu file:** Lưu file với tên `hello.py` (phần mở rộng '.py' là quy ước cho các file Python).
  - ▶ **Chạy chương trình:**
    - ★ Mở Terminal/Command Prompt và điều hướng đến thư mục bạn đã lưu file 'hello.py'.
    - ★ Gõ lệnh: `python hello.py` (hoặc `python3 hello.py` trên Linux) và nhấn Enter.
    - ★ Chương trình sẽ thực thi và hiển thị `Hello World` trên màn hình.

# Cập nhật & Tính năng mới của Python

## Phiên bản:

- **Ổn định hiện tại:** Python 3.13 (07/10/2024).
- **Đang phát triển:** Python 3.14.

## Tính năng mới (Python 3.14):

- PEP 649: Đánh giá trì hoãn chú thích.
- PEP 741: API C cấu hình Python.
- PEP 761: Sigstore thay PGP để xác minh.
- Cải thiện thông báo lỗi.

## Ý nghĩa trong bối cảnh rộng hơn:

- Python phát triển liên tục, duy trì vị thế hàng đầu.
- Nâng cao trải nghiệm lập trình viên, đặc biệt cho người mới.
- Tăng hiệu suất, khả năng mở rộng cho AI, ML, Khoa học dữ liệu.
- Cải thiện an toàn, bảo mật chuỗi cung ứng phần mềm.

**Tóm lại:** Các bản cập nhật mới thể hiện sự năng động, cam kết cải thiện và khả năng thích nghi của Python với công nghệ hiện đại.