

Análisis de datos con Python y Pandas

Centro de Análisis de Datos y Supercómputo

Introducción

Python

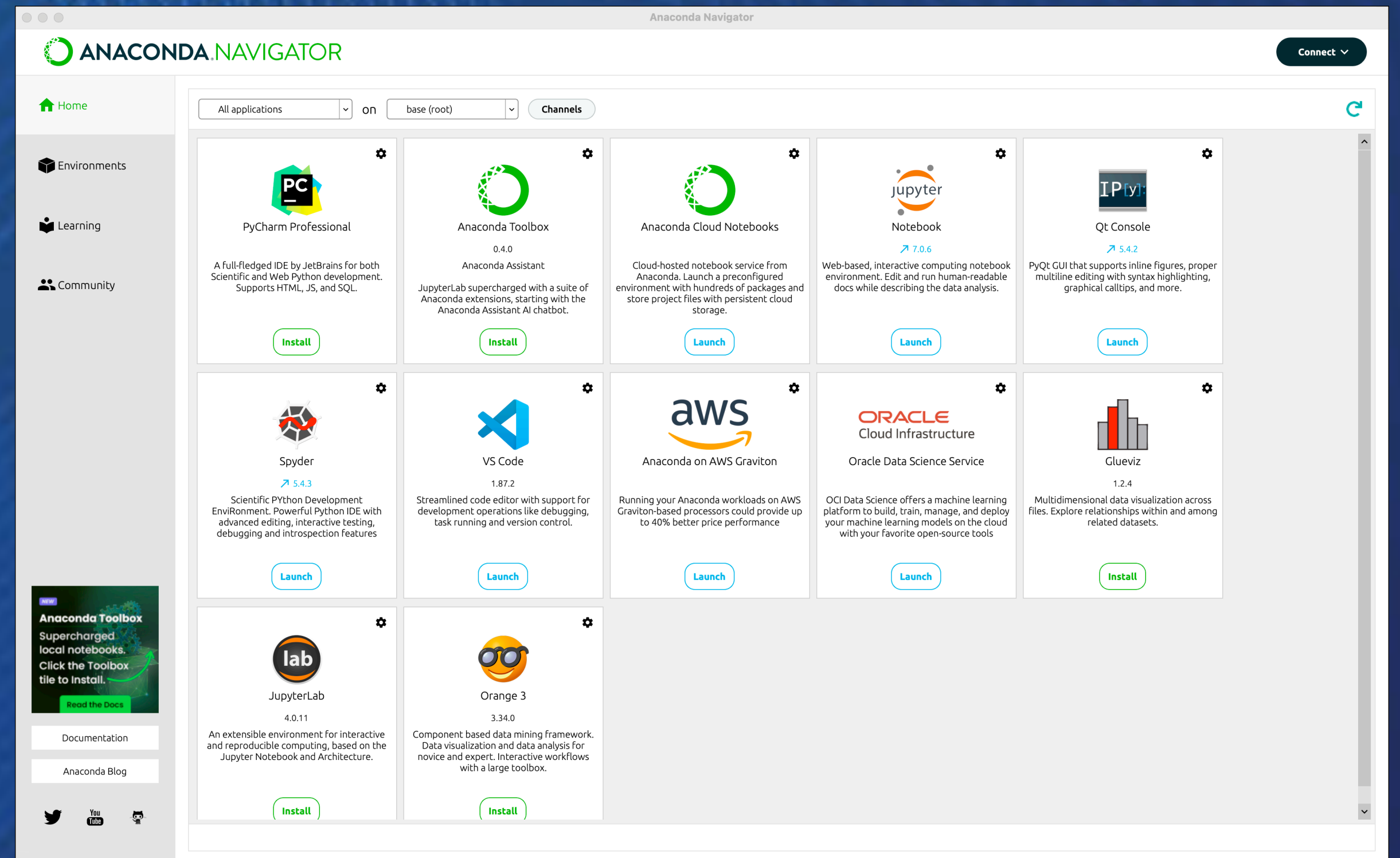


- Python es un lenguaje de programación de alto nivel, interpretado, interactivo y orientado a objetos.
- **Interpretado:** se ejecuta código línea por línea. Si existen errores en el código del programa, su ejecución se detiene.
- **Orientado a objetos:** es un paradigma de programación, donde los programas giran en torno a clases y objetos, en lugar de usar funciones y lógica. Uno se centra en manipular estos objetos, en lugar de centrarse en la lógica necesaria para esta manipulación. Un objeto se puede definir como un campo de datos con atributos y comportamientos únicos.
- **Alto nivel:** un lenguaje con fuerte abstracción de los detalles de la computadora. Uno podría no preocuparse sobre funcionalidades subyacentes, como la arquitectura o la administración de memoria.

Anaconda



- Anaconda es una distribución gratuita de Python diseñada específicamente para ciencia de datos y análisis.
- Incluye herramientas esenciales como Jupyter Notebook, un entorno interactivo para escribir y ejecutar código.
- Además, simplifica la instalación de bibliotecas como NumPy, Pandas y Matplotlib, eliminando conflictos comunes en la configuración de entornos.

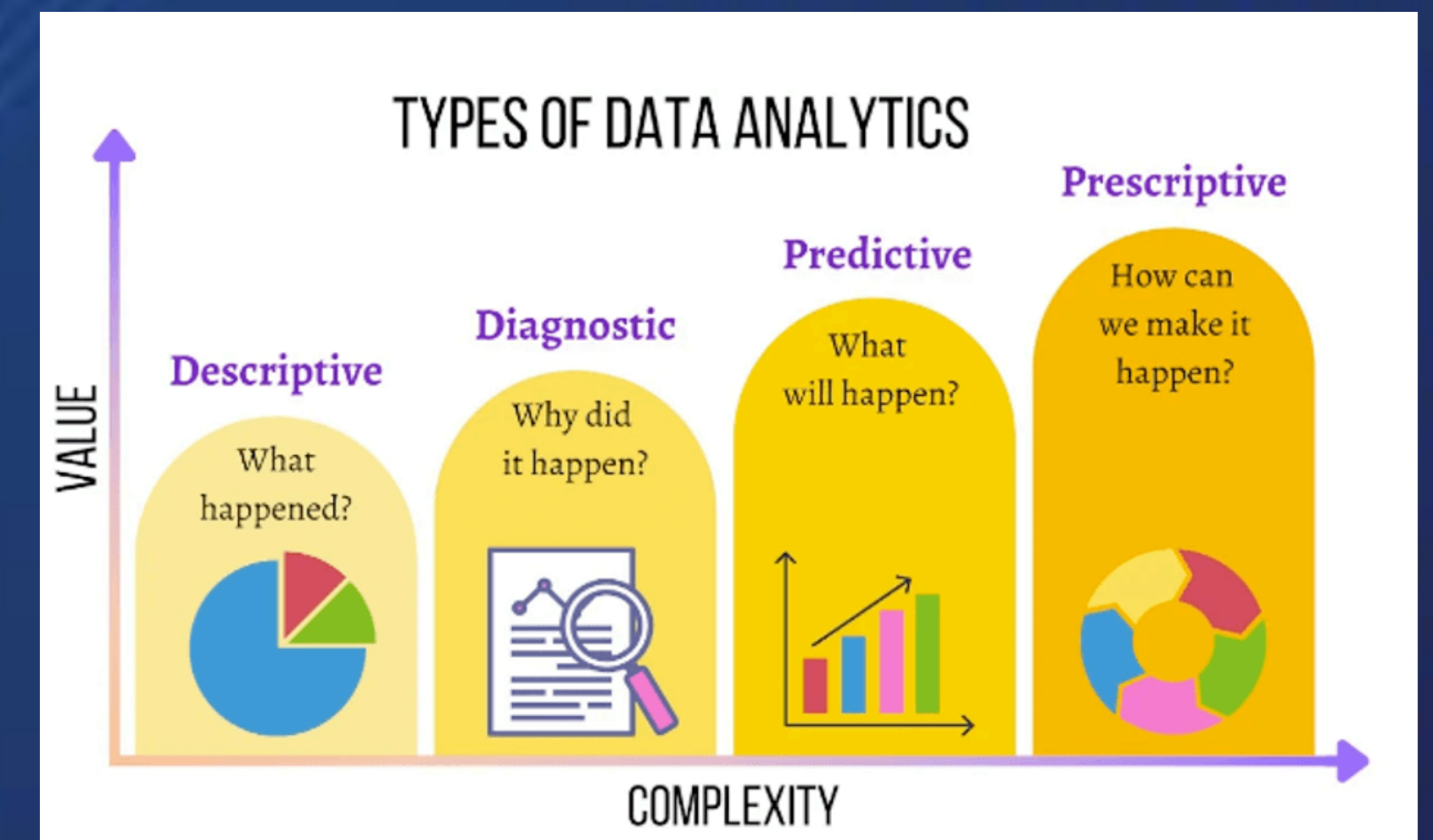
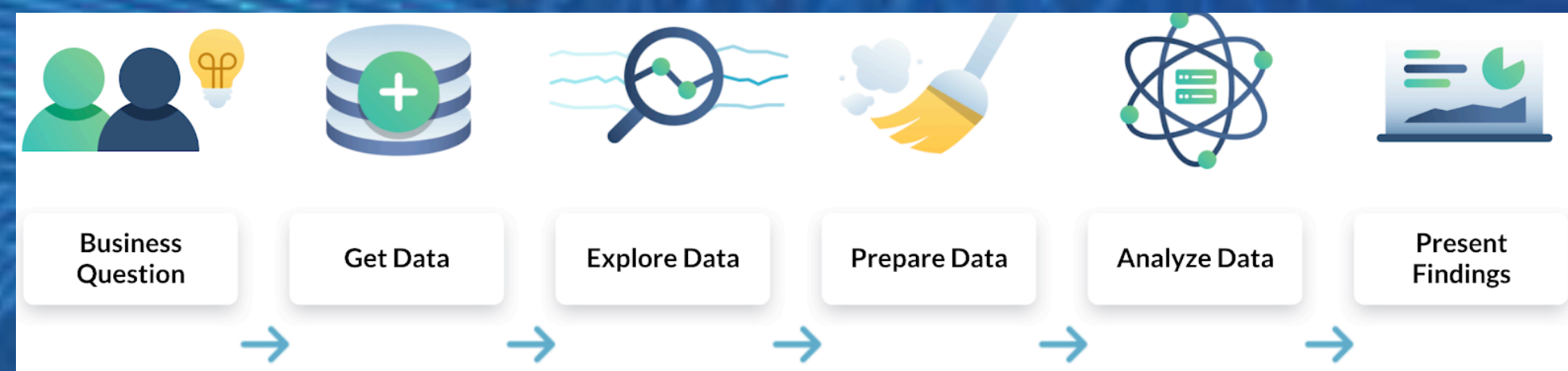


Introducción

Análisis de datos



- Proceso de obtener una perspectiva a partir de un conjunto de datos. Se busca una conclusión a partir de los mismos y se intenta identificar algún tipo de tendencia.
- La mayor parte del trabajo consiste en moldear o darle formato a la información con la que estemos trabajando.
- Los datos podrían estar dispersos, con espacios en blanco, con erratas, con valores nulos, etc. Son imperfectos así que necesitamos una herramienta que permita manipularlos y darles forma.



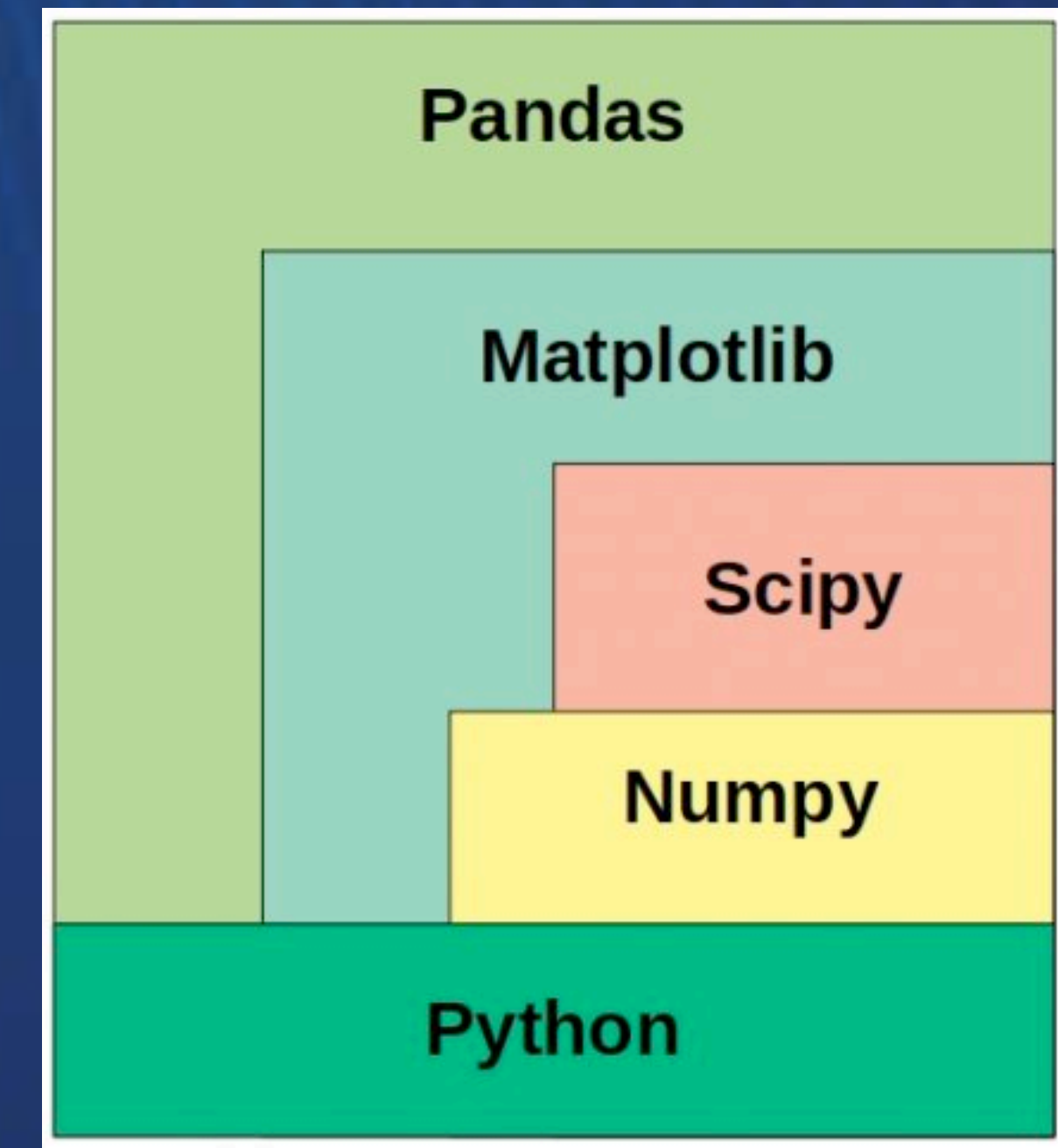
Introducción

Numpy (Numerical Python)



NumPy

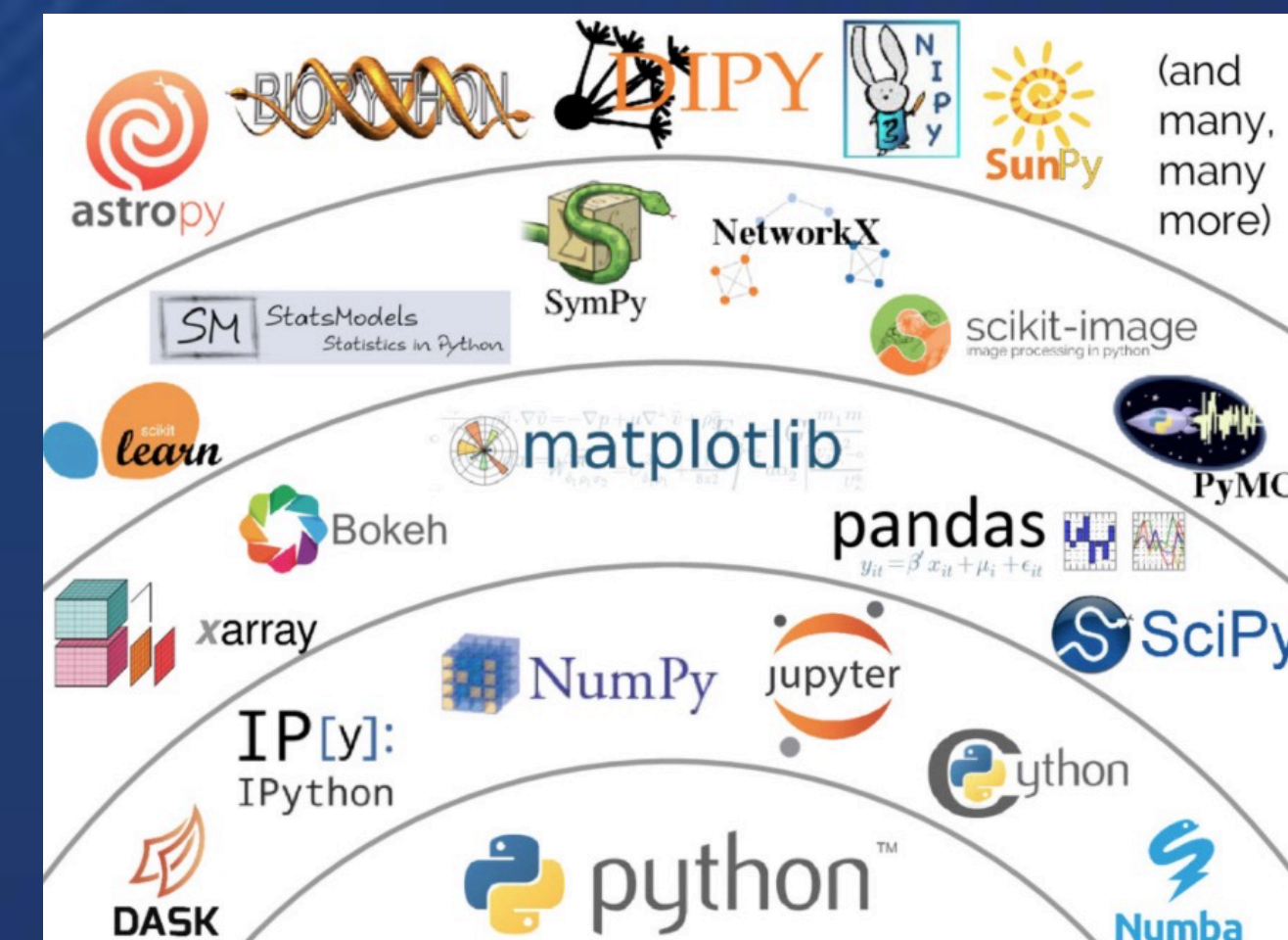
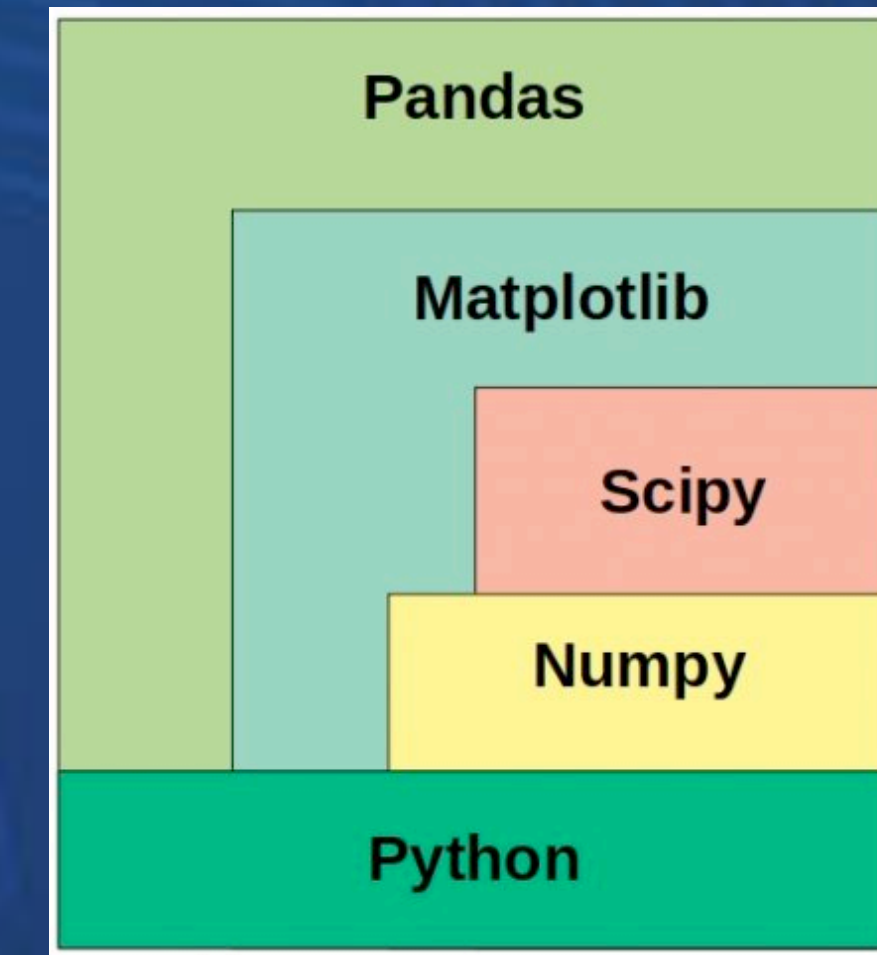
- Numpy es un paquete de Python que significa “Numerical Python”, es la biblioteca principal para la informática científica.
- Proporciona potentes estructuras de datos, implementando matrices n-dimensionales. Estas estructuras de datos garantizan cálculos eficientes con matrices (**50 veces rápido que una lista**).



Introducción

Pandas (panel data structures)

- Es una librería construida sobre Python, digamos, una extensión o complemento para el lenguaje. Un conjunto de código que desarrolladores han escrito para resolver un problema común.
- Es una herramienta robusta para analizar, filtrar, manipular, agregar, concatenar y limpiar datos.
- Por lo mismo que está escrita en python, pandas permite resolver el problema con un enfoque en programación. Es nos permite, por ejemplo, tener automatización en un proceso.

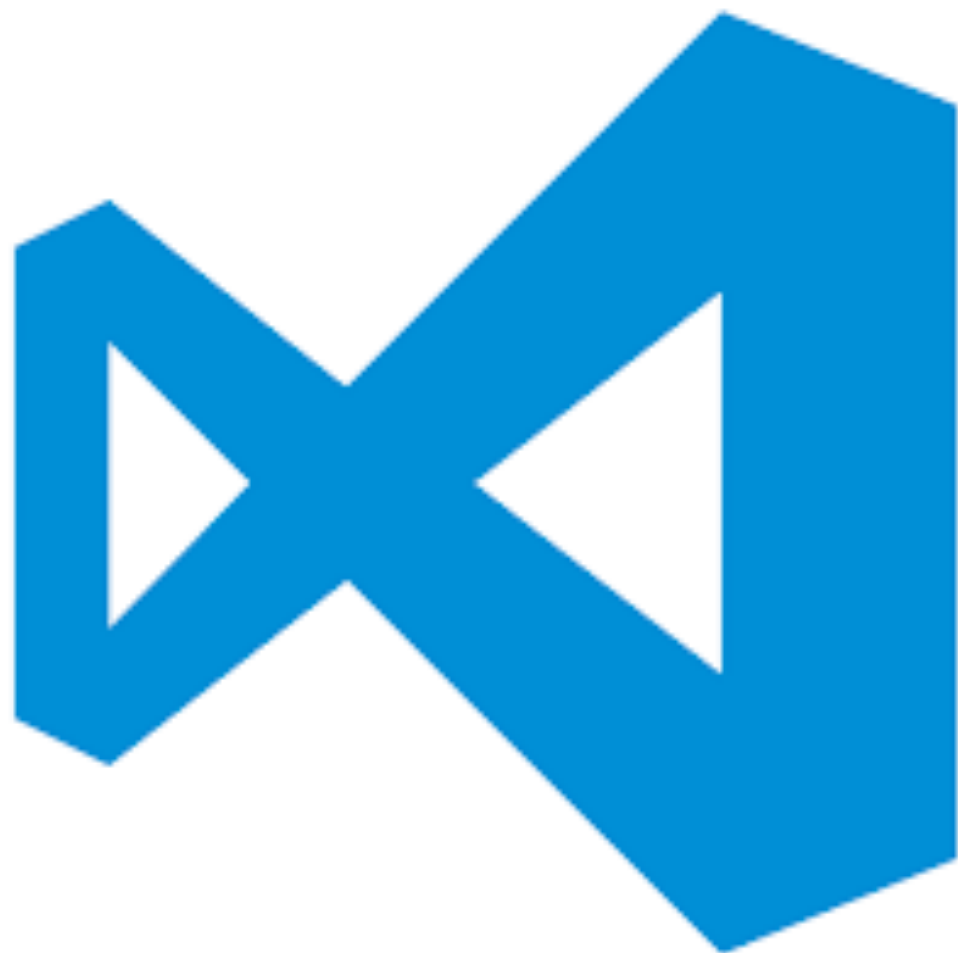




<https://github.com/lvTole/python-pandas-taller>



<https://www.anaconda.com/download>



<https://code.visualstudio.com/download>