Big Data (Análisis de grandes volúmenes de datos)

DIPLOMADO, MÓDULO III

Dr. José Luis Cendejas Valdez, UTM

Dr. Heberto Ferreira Medina, IIES-UNAM

Dr. Sergio Rogelio Tinoco Martínez, ENES-UNAM

Actividades sincrónicas:

Classroom: https://classroom.google.com/c/NzM0NTk3Mzc3MzIz?cjc=5wc4s6l

ZOOM: https://cuaieed-unam.zoom.us/j/87963800329?pwd=il8rD7Ai14yqCN4nSKxijXRvxCkwjC.1

DIPLOMADO EN APRENDIZAJE MÁQUINA Y PROFUNDO APLICADO A GRANDES VOLÚMENES DE DATOS (Big Data)
PROYECTO PAPIME PE103124

Objetivos

- **Objetivo del módulo:**
  - Desarrollar aplicaciones para la búsqueda, extracción, procesamiento, interpretación y visualización de grandes volúmenes de datos.

- **Específicos:**
  - Aportar la capacidad para procesar, interpretar y graficar datos previamente obtenidos.
  - Aplicar sus conocimientos y habilidades para análisis de grandes cantidades de datos.

Calendario:

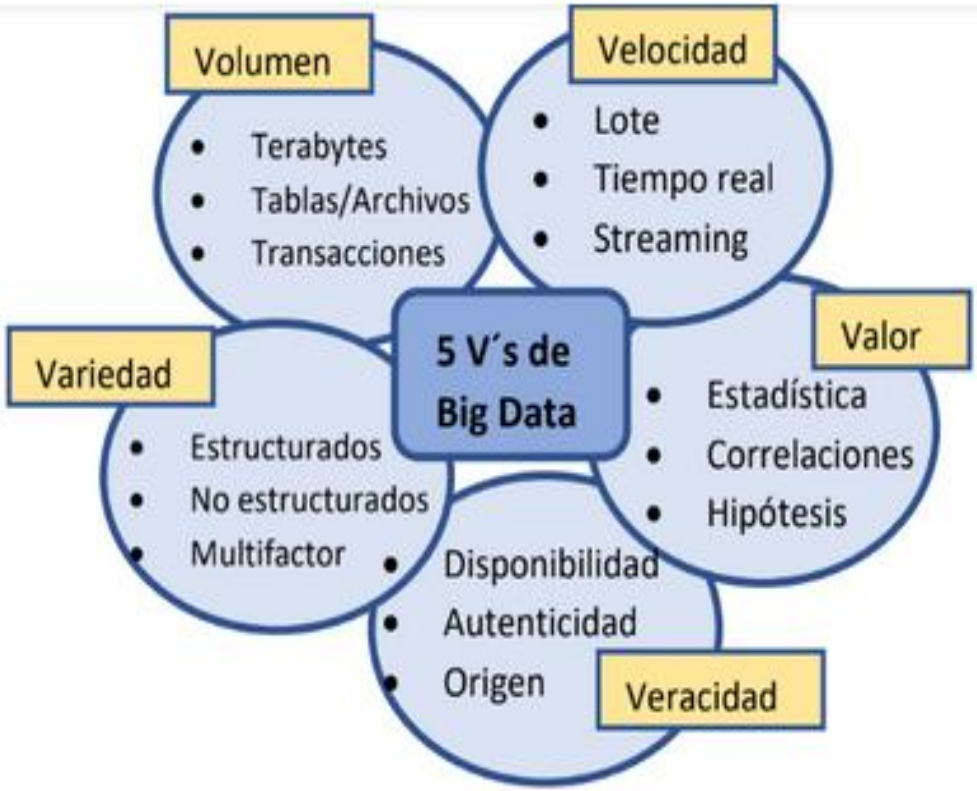
Sesiones síncronas, martes y jueves de 17:00 a 20:00 hrs.

Sesión	Fechas	Actividades	Entregas	Puntos
1	Martes 26 de noviembre	El uso del Big Data en la Ciencia de Datos.		
2	Jueves 28 de noviembre	Capítulo libro, Web Scraping (Análisis del Big Data con ML).	Tarea 1	3
3	Martes 3 de diciembre	Uso de almacenamiento en la Red (NAS).	Tarea 2	2
4	Jueves 5 de diciembre	Uso de HDFS, DASK.		
5	Martes 10 de diciembre	Análisis de grandes volúmenes de datos para el DL (Aplicación). Práctica 4.	Tarea 4	3
6	Martes 17 de diciembre	Uso de Bases de Datos (DB) no estructuradas.	Tarea 3	3
7	Jueves 19 de diciembre	Despliegue de un proyecto de Big Data con servicios Web.	Tarea 5	2
8	Jueves 9 de enero 2025	Cierre de sesiones síncronas	Capítulo de libro	3
9	Semana 13 -16 de enero 2025.	Sesiones asíncronas de apoyo **Cierre 17** de enero de 2025	Puntos	13

Actividades sesión 8.

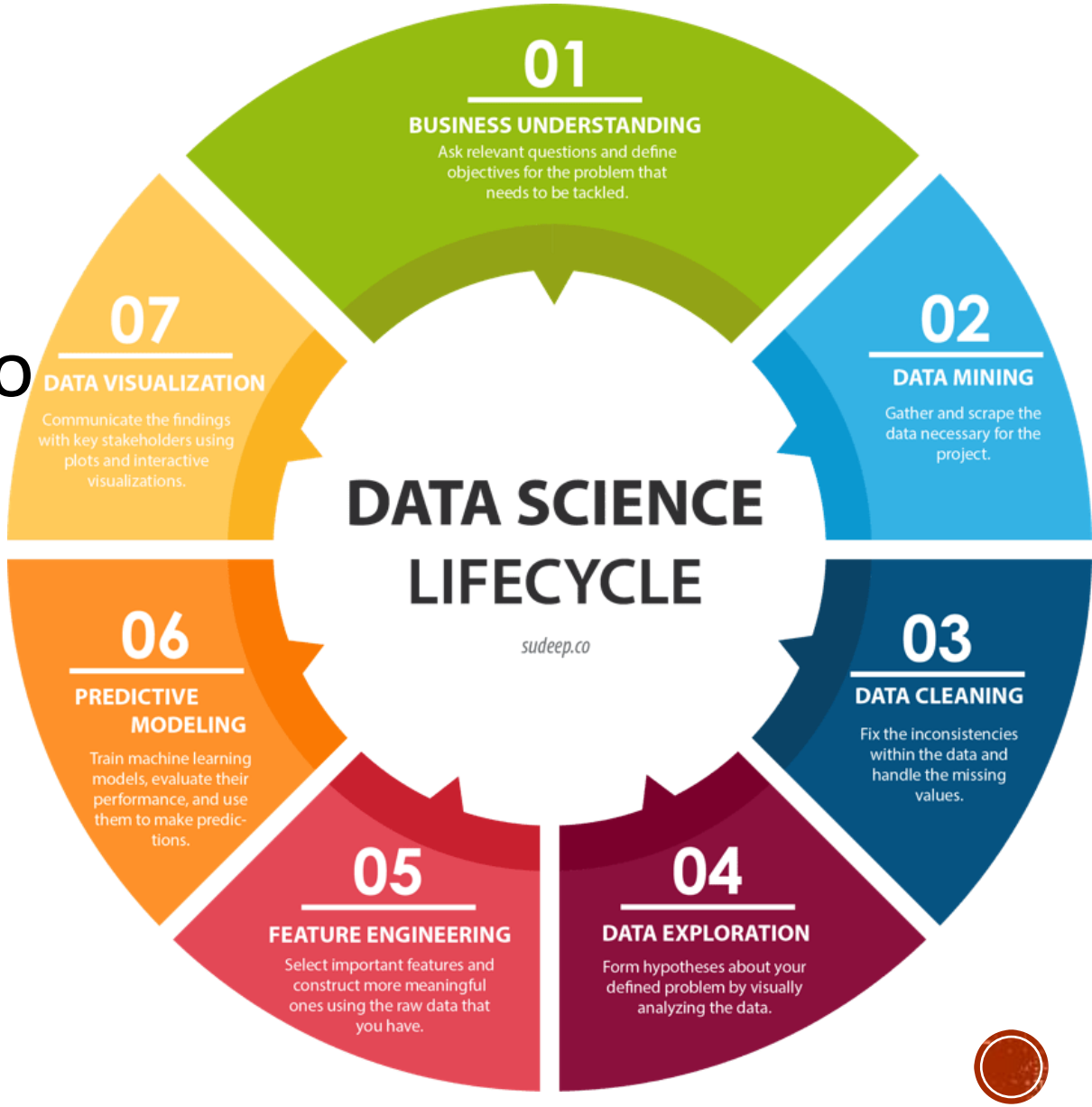
- Evaluación diagnóstica final.
- Servidor Web para proyectos de ML, DL y Big Data
- Uso de Flask
- https://flask.palletsprojects.com/en/stable/
- Cierre del módulo III.

V´s del Big Data

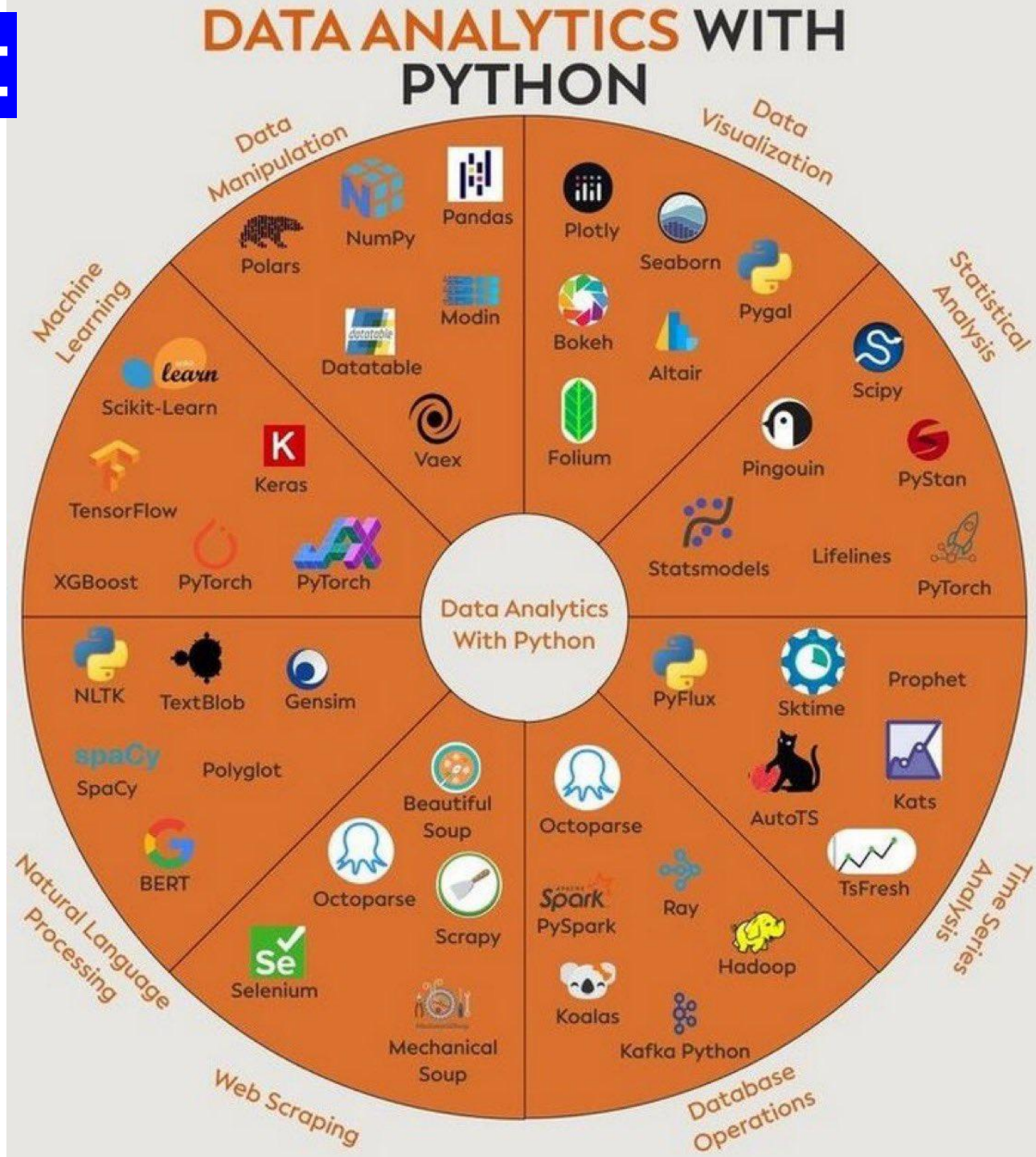


Ciclo de vida del ML,DL:

- Visualización y
- Uso del modelo



Ecosistema:



Flask

WSGI:
Web Server Gateway Interface

FLASK

- Es **marco de desarrollo Web (framework)** basado en microservicios para Python. Está diseñado de manera sencilla para comenzar rápidamente a crear una API, ya que es bastante simple y fácil.

- Se basa en el concepto de crear servicios **Restful API** con Python y Flask. Fue desarrollado en 2004 como un proyecto de software libre y se integró en 2014-2018 al marco de Boostrap.

- Flask se ha hecho popular entre los entusiastas de Python. A partir de 2020 es ampliamente utilizado en GitHub entre los marcos, y es una alternativa a Django

FLASK

- Una API REST es una aplicación mínima viable utilizando Flask-RESTful, se deben seguir 4 pasos principales para crearla:

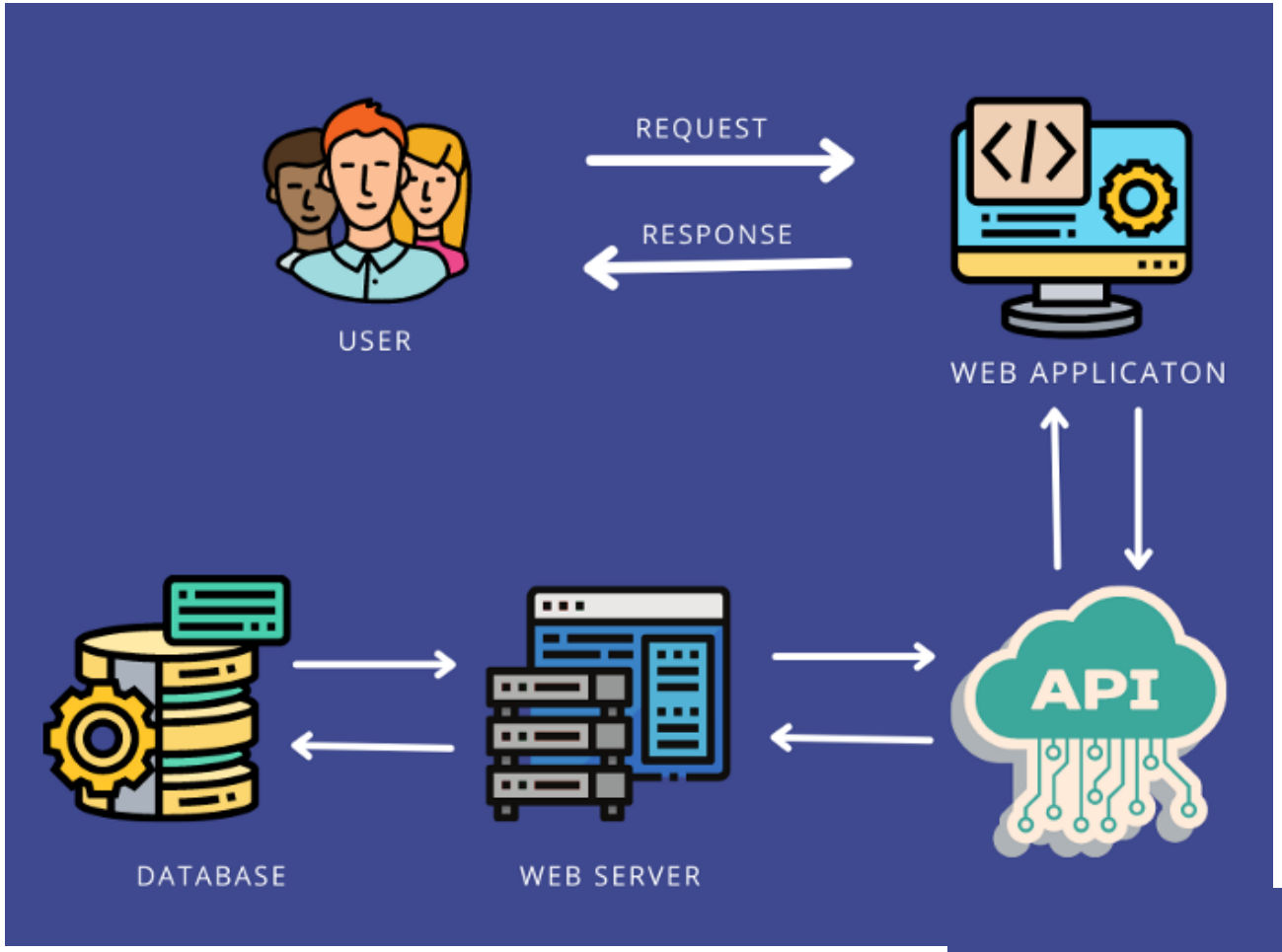
- **Paso 1:** crear un modelo: modelo que utiliza un conjunto de datos para el ML o DL

- **Paso 2:** crear una API REST: parte principal para publicación de datos. Para utilizar el modelo ML/DL
-
- **Paso 3:** Ejecutar la API REST: ejecutada en el entorno definido

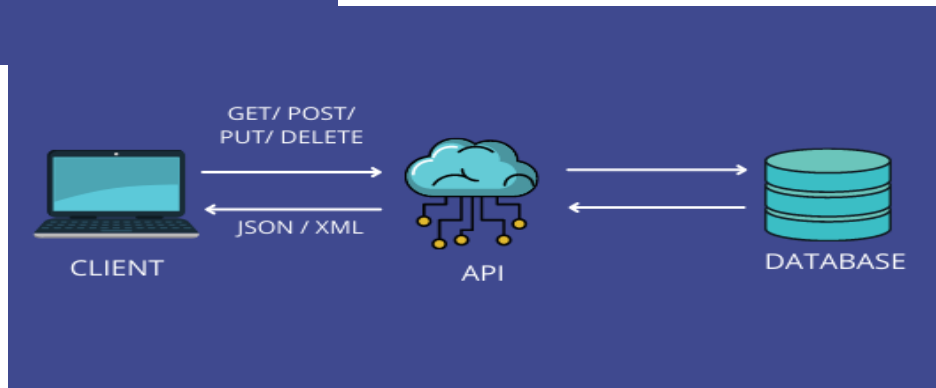
- **Paso 4**: Probar la API: usar el modelo para hacer predicciones llamando a los microservicios

Restful API:

REST o REpresentational State Transfer

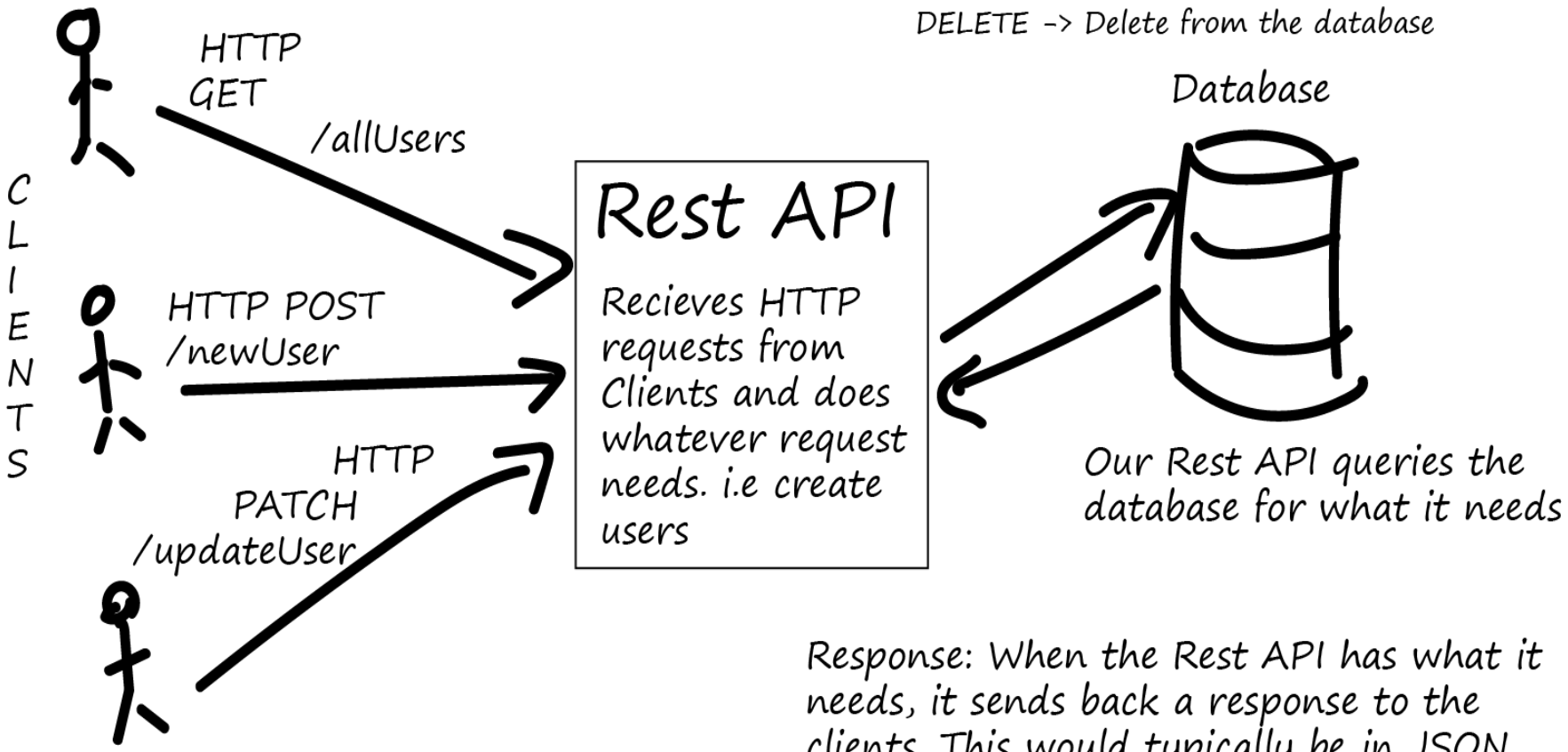


API RESTful es una interfaz cliente-servidor



Rest API Basics

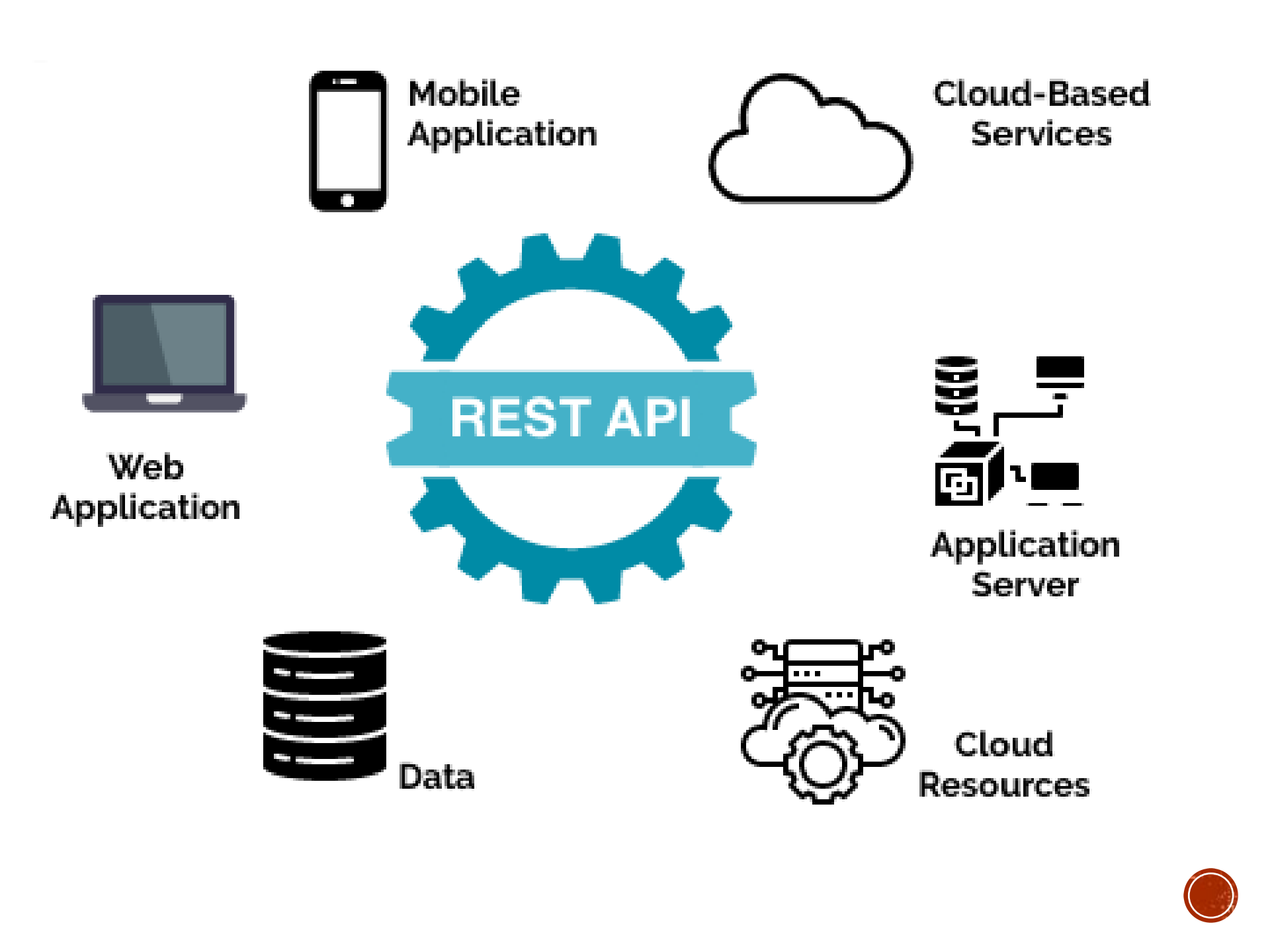
Typical HTTP Verbs:
GET -> Read from Database
PUT -> Update/Replace row in Database
PATCH -> Update/Modify row in Database
POST -> Create a new record in the database
DELETE -> Delete from the database



Our Rest API queries the database for what it needs

Our Clients, send HTTP Requests and wait for responses

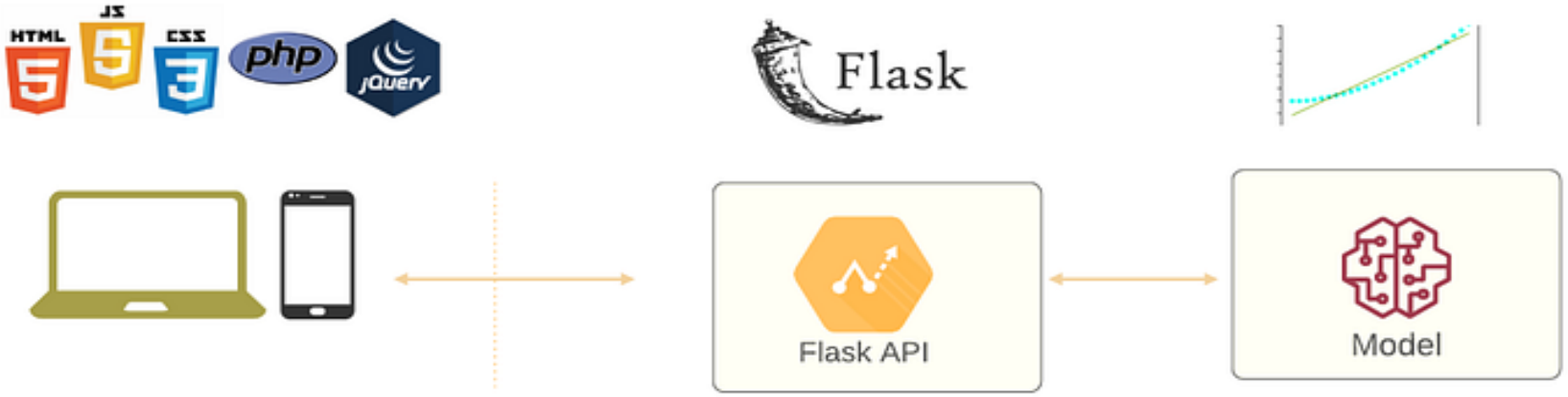
Response: When the Rest API has what it needs, it sends back a response to the clients. This would typically be in JSON or XML format.


Mobile Application
Cloud-Based Services
REST API
Web Application
Application Server
Data
Cloud Resources

Instalación y uso de Flask

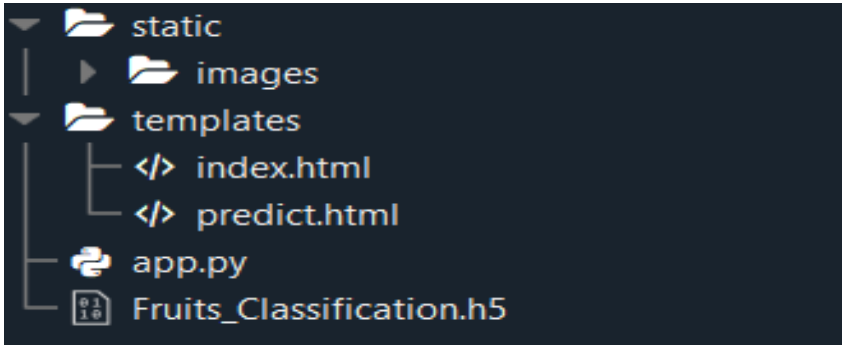
- Se deben tener un entorno previo ML, DL o Big Data
- Acceder a un nodo: ssh heberto@132.248.172.201
- $conda activate DL
- $ pip install Flask
- Revisar la instalación con:

  https://flask.palletsprojects.com/en/stable/quickstart/



Estructura de un proyecto:

- **app.py** : API Restful con métodos get/post/put implementados
- **modelo.h5** : modelo previamente entrenado (TL)
- Carpeta **static** : imágenes o archivos para pruebas o entrenamiento (train, test)
- Carpeta **templates**: archivos html para despliegue del modelo
- templates -> index.html : Página de inicio del despliegue
- templates-> predict.html : Página que ejecuta el modelo de predicción y entrega resultados en formato html



Ejecución de fruits predict:

- Para pruebas se utiliza (nodo1):
- $flask --app hello.py run –host=132.248.172.201
- Probar desde casa con: http://132.248.172.201:5000

-> “Hello world”

- Para ejecutar el modelo de TL **fruts predict** (medium.com) utilizar:
- $flask --app app.py --host=132.248.172.201
- http://132.248.172.201:5000

Ejecución multiusuario:

- Para ejecutar el modelo de TL **fruts predict con un host y puerto diferentes**:
- Editar app.py:
- app.run(debug=true, use_realoder=false, host='132.248.172.201', port=5050)
- $set FLASK_ENV=development
- $python app.py
- http://132.248.172.201:5050

Bibliografía

- Big data, machine learning y data science en python. José Manuel Ortega Candel. Ed. Ra-Ma. 2023
- Sistemas de Big Data. Victor Manuel López Fandiño. Ed. Ra-Ma. 2023
- Big Data MBA: Driving Business Strategies with Data Science, Schmarzo, Bill. Ed. Wiley
- Hands-On Machine Learning, with Scikit-Learn, Keras, and TensorFlow
- Concepts, Tools, and Techniques to Build Intelligent Systems. Aurélien Géron. O'Reylli. 2023
- Python for Data Analysis. 3a. Edición. Data Wrangling with pandas, NumPy, and Jupyter. Wes McKinney. Ed. O'Reylli. 2022
- Scaling Python with Dask. From Data Science to Machine Learning. Holden Karau and Mika Kimmins. Ed. O'Reylli. 2023
- Documentos oficiales de Anaconda Python. https://docs.anaconda.com/
- Documentos oficiales de Flask. https://flask.palletsprojects.com/en/stable/
- Medium.com (2024). https://medium.com/@draj0718/deploying-deep-learning-model-using-flask-api-810047f090ac
- Comience a utilizar boostrap. https://getbootstrap.com/docs/4.4/getting-started/introduction/