

Notas

MongoDB

Date: 2022-03-09

Gestor de bases de datos **No Relacionales**(NO-SQL). Orientada a *colecciones y documentos*, en vez de *tablas* y registros*. Dichos documentos tienen formato **JSON**.

[Documentación](#)

ETL - Extract, Transform, Load.

Date: 2022-03-09

Proceso de extraer, transformar t cargar data desde diferentes fuentes a un *data warehouse* u otros repositorios de código unificado de data. Integreción de datos.

[Fuente](#)

ELT - Extract, Load, Transform.

Date: 2022-03-09

Cargar la data cruda oara luego ser transformada.

[Fuente](#)

Pre-Scrapping

Date: 2022-03-09

¿Qué secciones y URL voy a explorar? **Crawler**

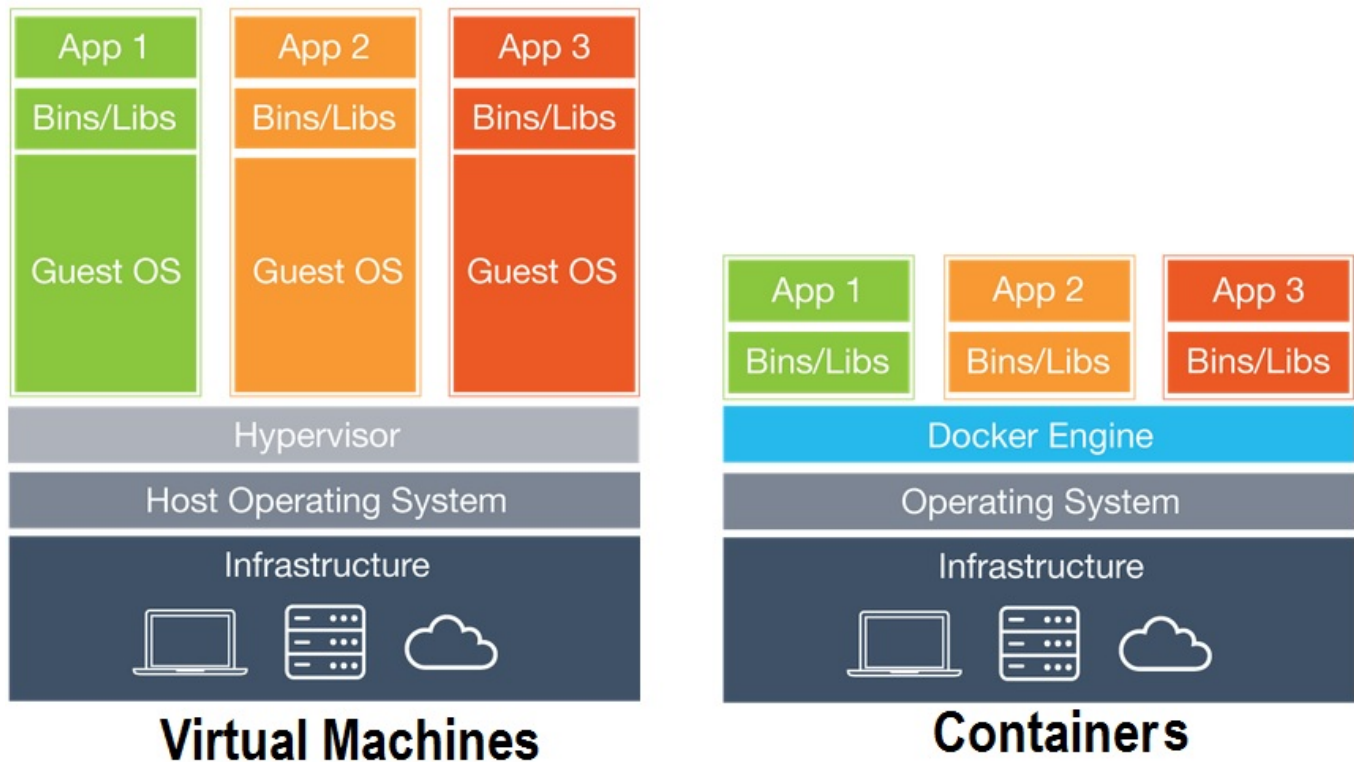
Docker

Date: 2022-03-09

Herramienta para crear, desplegar y correr apps usando **contenedores**. Los contenedores permiten desplegar una aplicación con todas las librerías, paquetes o dependencias que necesite. Así nos aseguramos que la app correrá en cualquier máquina independientemente de la configuración del sistema.

A diferencia de una máquina virtual, Docker, en vez de creat dodo un nuevo sistema de archivos con una imagen completa del SO, permite que las aplicaciones usen el mismo kernel de Linux que el sistema y sólo se le agreguen aquellas dependencias que no se encuentran en el sistema de destino. Esto implica una

mejor performance, reduce el tamaño de la aplicación, independencia, abstracción, automatización y aislar procesos.



[Docker para todos](#)

[Docker_introduccion](#)

[Instalacion](#)

Torrent

Date: 2022-03-11

Es un archivo enviado a través del protocolo **BitTorrent**. Son diferentes a las descargas regulares ya que son descargados desde varios servidores al mismo tiempo.

[Fuente](#)

BitTorrent Protocol

Date: 2022-03-11

Usa múltiples computadoras para transferir un único archivo, en consecuencia reduce el ancho de banda requerido por cada servidor. Cuando una descarga *torrent* inicia, el sistema BT localiza diferentes computadoras con el archivo completo y descarga diferentes partes del archivo desde cada computadora. Esto conlleva menor BW y mayor velocidad

[Fuente](#)

Seeders

Date: 2022-03-11

Usuarios que tienen el archivo completo y lo están compartiendo. Son quienes suben la data.

A mayor cantidad de seeders, mayor velocidad de descarga.

[Fuente](#)

Leechers

Date: 2022-03-11

Los usuarios que están bajando la data.

[Fuente](#)

YAML

Date: 2022-03-11

Lenguaje de serialización de datos a menudo usado para escribir archivos de configuración.

[yaml Fuente](#)

Install/Uninstall .deb files

Date: 2022- 03-14

1. To install a .deb file, simply Right click on the .deb file, and choose Kubuntu Package Menu->Install Package.
2. Alternatively, you can also install a .deb file by opening a terminal and typing:

```
sudo dpkg -i package_file.deb
```

3. To uninstall a .deb file, remove it using Adept, or type:

```
sudo apt-get remove package_name
```

[Fuente](#)

How to install tar.gz in ubuntu 20.04

Date: 2022-03-14

```
tar -xzf archive-name.tar.gz
cd archive-name
./configure
make
sudo make install
```

[Fuente](#)

Anaconda

Date: 2022-03-14

Verificar que esté instalado y la versión

```
conda --version
```

Listar entornos

```
conda env list
```

Eliminar entorno

```
conda-env remove -n nombre_entorno
```

Crear entorno

```
conda create -n nombre_entorno python=3.10
```

Activar entorno

```
conda activate nombre_entorno
```

Luego instalas los paquetes que desees. EJ:

```
pip install requests
```

Listar paquetes instalados en el entorno

```
pip freeze
```

```
mkdir intro_scrapping  
cd intro_scrapping  
mkdir requirements  
pip freeze > requirements/production.txt
```

Docker

Date: 2022-03-14

INSTALAR PORTAINER

[DockerHub](#)

Descargar Imagen de Mongo

```
(sudo) docker pull mongo
```

Correr Imagen

```
docker run -p 37017:27017 --name mi_mongo -d mongo:latest
```

Listar Cointariners Activos

```
docker ps
```

Frenar container

```
docker nombre_cointainer
```

Eliminar

```
docker rm nombre_cointainer
```

requests.session()

Date: 2022-03-16

El objeto *Session* nos permite persistir ciertos parametros a través de las *requests*. También persiste las *cookies* a traves de todas las requests hechas desde la misma instancia Session. Esto resulta en una significativa mejora de la performance ya que la misma coneccion TCP será reutilizada para las distintas requests hechas desde el mismo host.

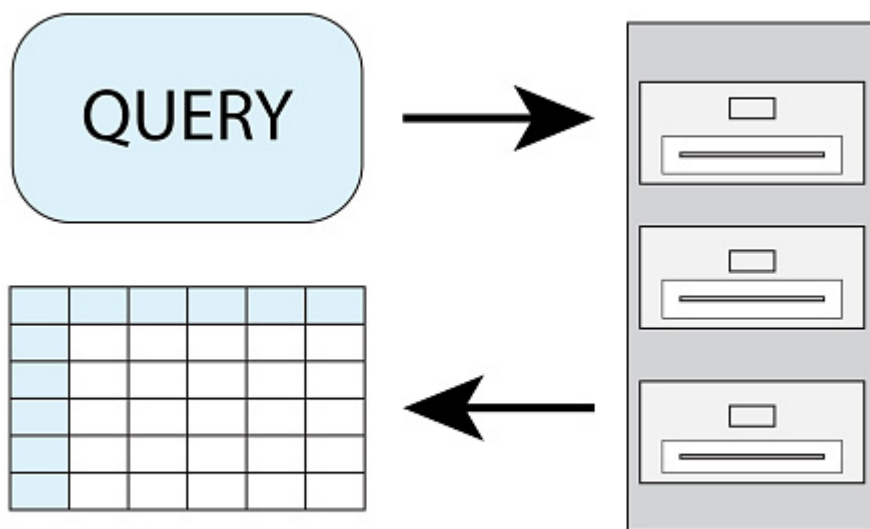
[Documentación](#)

query

Date: 2022-03-16

Una *query* es una consulta por data o informacion a una base de datos la cual devolverá una respuesta según lo consultado.

SQL = Structured **Q**uery Language



[Fuente](#)

future.result()

Date: 2022-03-17

Retorna el valor retornado por la función ejecutada por el *executor*.

[Documentación](#)

locals() - Python function

Date: 2022-03-17

Retorna un diccionario que contiene las variables definidas en el scope local donde fueron llevadas.

[Fuente](#)[Documentación](#)

argparse

Date: 2022-03-17

El modulo *argparse* hace más fácil escribir interfaces a través de líneas de comandos amigables para el usuario. Define que argumentos son requeridos, genera mensajes de ayuda y error.

[Documentación](#)

API - Application Programming Interface

Date: 2022-03-18

Es un software intermediario que permite que dos aplicaciones interactuen entre ellas.

¿SABES QUÉ ES UNA API?

Una interfaz es **una capa que conecta dos sistemas**. Una API es una interfaz que conecta aplicaciones **para que compartan información**, por ejemplo el API de mapas de Google, entre otras.

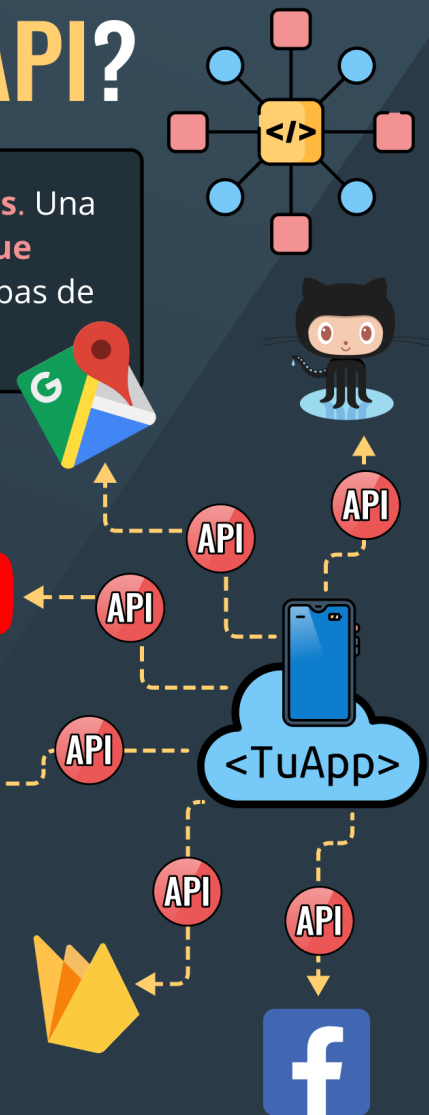
API → **INTERFACE**

→ **PROGRAMMING**

→ **APPLICATION**

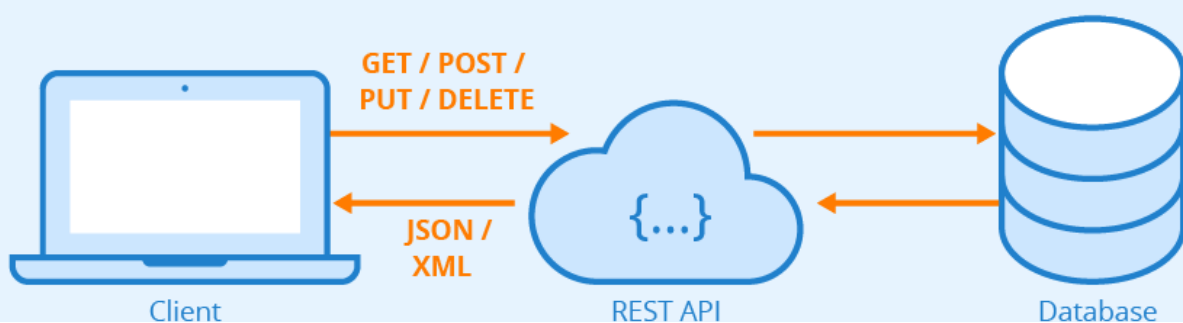
API REST

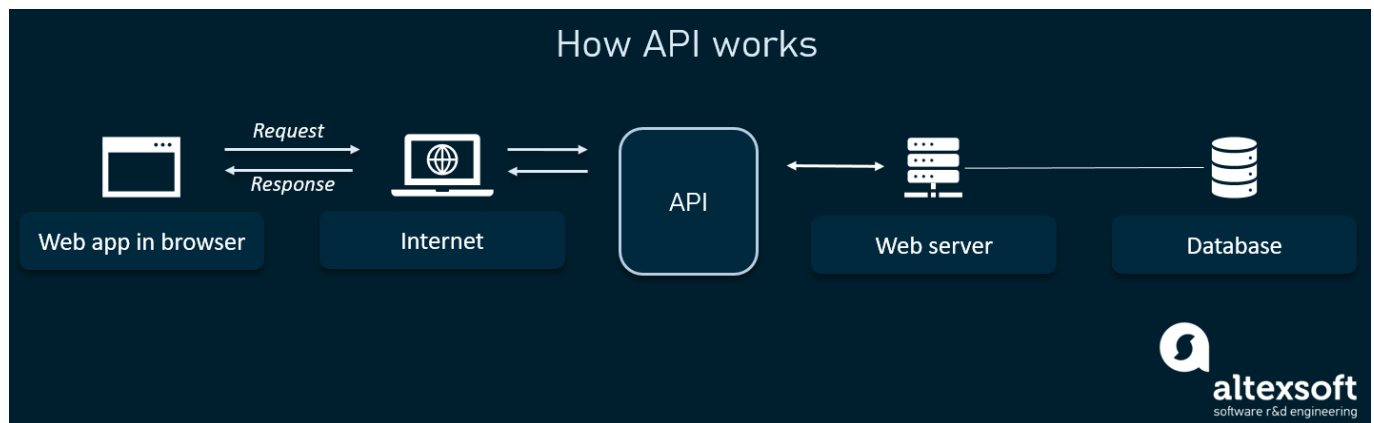
Rest es una arquitectura para APIs que se conectan **vía web con el protocolo HTTP**. Permite acceder a la información con identificadores únicos (URI).



Aprende a desarrollar APIs con buenas prácticas en:

 ed.team/cursos/api-rest





[Fuente1](#)

ORM

Date: 2022-03-18

Object-Relational mapping, o lo que es lo mismo, mapeo de objeto-relacional, es un modelo de programación que consiste en la transformación de las tablas de una base de datos, en una serie de entidades que simplifiquen las tareas básicas de acceso a los datos para el programador.

[Fuente](#)

Payload

Date: 2022-03-18

Cuando la data es enviada a través de Internet, cada unidad transmitida incluye el *header* y la *data* propiamente dicha. El *header* identifica la fuente y el destino de los paquetes, mientras que la data propiamente dicha se denomina **payload**. La informacion del header sólo es usada en el proceso de transmisión, es separada del paquete cuando llega a su destino. El **payload** es la única que recibe el sistema de destino.

[Fuente](#)

Concurrency and Parallelism

Date: 2022-03-18

Concurrency: Concurrency relates to an application that is processing more than one task at the same time. Concurrency is an approach that is used for decreasing the response time of the system by using the single processing unit. Concurrency creates the illusion of parallelism, however actually the chunks of a task aren't parallelly processed, but inside the application, there are more than one task being processed at a time. It doesn't fully end one task before it begins ensuing. Concurrency is achieved through the interleaving operation of processes on the central processing unit(CPU) or in other words by the context switching. that's rationale it's like parallel processing. It increases the amount of work finished at a time.

Parallelism: Parallelism is related to an application where tasks are divided into smaller sub-tasks that are processed seemingly simultaneously or parallel. It is used to increase the throughput and computational speed of the system by using multiple processors. It enables single sequential CPUs to do lot of things "seemingly" simultaneously.

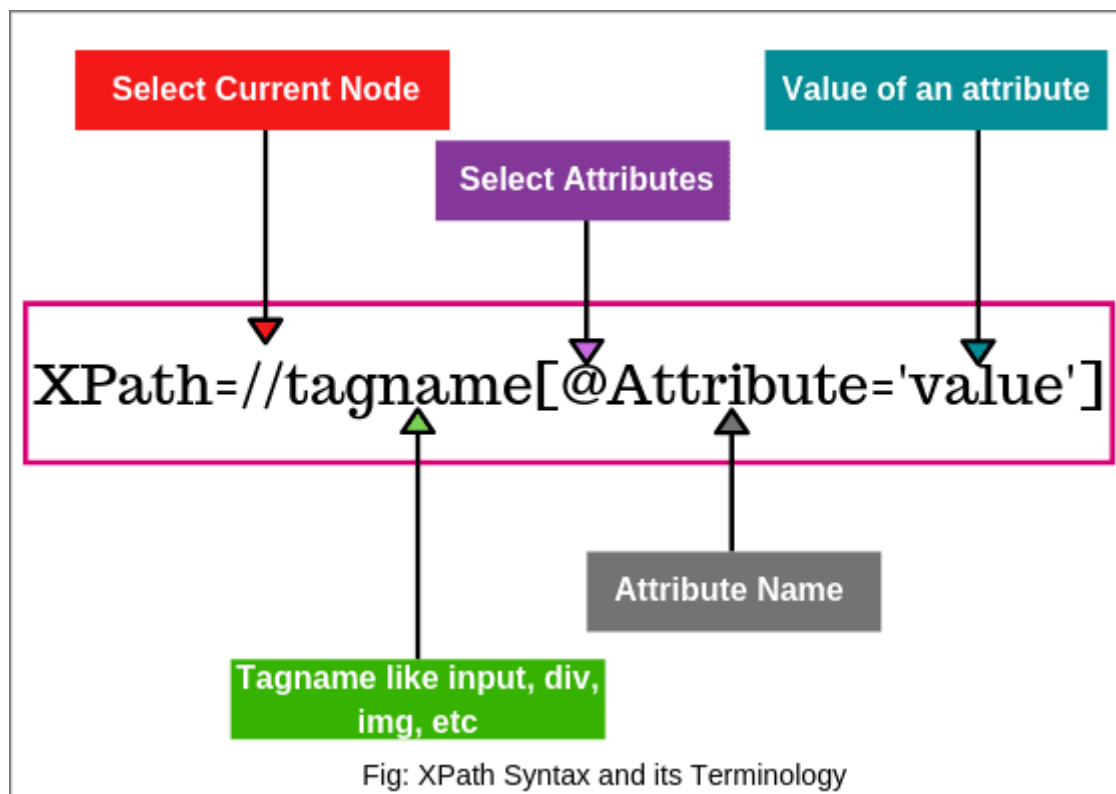
Parallelism leads to overlapping of central processing units and input-output tasks in one process with the central processing unit and input-output tasks of another process. Whereas in concurrency the speed is increased by overlapping the input-output activities of one process with CPU process of another process.

[Fuente](#)

XPath

Date: 2022-03-21

XPath stands for XML Path Language. It uses a non-XML syntax to provide a flexible way of addressing (pointing to) different parts of an XML document. It can also be used to test addressed nodes within a document to determine whether they match a pattern or not.



[Fuente](#)

[XPath CheatSheet XPath Tutorial](#)

Twitter API Elevated Access Level Questions

Date: 2022-04-01

- In English, please describe how you plan to use Twitter data and/or APIs. The more detailed the response, the easier it is to review and approve.
- Are you planning to analyze Twitter data? Please describe how you will analyze Twitter data including any analysis of Tweets or Twitter users.
- Will your App use Tweet, Retweet, Like, Follow, or Direct Message functionality? Please describe your planned use of these features.
- Do you plan to display Tweets or aggregate data about Twitter content outside Twitter? Please describe how and where Tweets and/or data about Twitter content will be displayed outside of Twitter.
- Will your product, service, or analysis make Twitter content or derived information available to a government entity? Please list all government entities you intend to provide Twitter content or derived information to under this use case.

pandas.DataFrame.explode

Date: 2022-04-07

Transforma todos los elementos de una lista (or *list-like*) en una fila y replica los índices.

```
df['SplitTitle'] = df['Title'].apply(lambda x : x.split(' '))

df = df.explode('SplitTitle')

df.loc[1]
```

[Documentation](#)

Regular Expressions (RegEx)

Date: 2022-04-08

Define a search pattern that can be used to search for things in a string.

[freeCodeCamp Video](#)

Natural Language Processing (NLP)

Date: 2022-04-08

Se refiere a la rama de la IA que se ocupa de darle a las computadoras la habilidad de entender texto y palabras escritas de la misma manera que lo pueden hacer los seres humanos.

[Fuente - IB;](#)

Title

Date: 2022-03-14

```
codigo
```

[Markdown CheatSheet](#)

DUDAS A DESARROLLAR

pruebas unitarias? milestone? AirFlow? Azzure? AWS? VPN? Pipeline? PEP8? global python html.unescape
python concurrent vs multiprocessing (libraries) multiprocessing vs multithreading
