



# Almacenamiento y captura de datos

Claudio Aracena

GobLab - Universidad Adolfo Ibáñez  
Chatbot Chile



# Contenidos

- Captura de datos desde archivos
- Base de datos
- Captura y almacenamiento de datos en BD
- Captura de datos de la Web (Web scraping)
- **Captura de datos de API (ej: Twitter)**
- Captura y almacenamiento en arquitecturas Big data

Códigos y clase en:

<https://github.com/caracena/almacenamiento-captura-datos>



# Clase de hoy

## Captura de datos de API

- Application Programming Interface (API)
- Conexión API vía requests
- Conexión API vía librerías

## Preparación última clase

- Descargar VirtualBox
- Descargar Cloudera 5.13 VM





# Application Programming Interface (API)

Una API es una forma estandarizada de compartir data en la web. Muchos sitios web comparten data utilizando APIs.

Algunos ejemplos de APIs son:

- Twitter API
- Google Maps API
- Github API
- Biblioteca del Congreso Nacional API



## Conexión API vía requests

Una de las formas de conectarse a una API es mediante peticiones HTTP. Dependiendo lo que indique la documentación de la API se deberán realizar requerimientos GET, POST, PUT o DELETE.

## Conexión API vía librerías

Otra forma de realizar peticiones a una API es utilizar alguna librería construida para esos fines. Generalmente, existen librerías de conexión para aplicaciones populares como Twitter, Google, Facebook, entre otras.



# Preparación última clase

## Descargar VirtualBox

- <https://www.virtualbox.org/wiki/Downloads>

## Descargar Cloudera 5.13 VM

- [https://www.cloudera.com/downloads/quickstart\\_vms/5-13.html](https://www.cloudera.com/downloads/quickstart_vms/5-13.html)