

APLICACIÓN BIG DATA SOBRE EL EFECTO DEL CORONAVIRUS EN EL MUNDO



EU Open Data Portal

- Mario Pérez Sánchez-Montañez
- David Camuñas Sánchez
- Francesco Zingariello

ÍNDICE

- ¿Qué es el BigData?
- Fuente de datos
- Creación de la base de datos
- Estructura de la base de datos
- Demostración

¿QUÉ ES BIG DATA?

- Tipos de datos:
 - Datos estructurados
 - Datos semiestructurados
 - Datos no estructurados



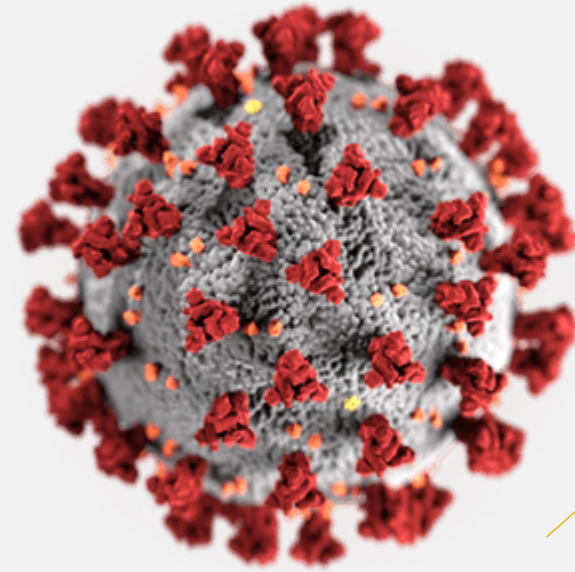
Fuente: <http://lluna4tc.blogspot.com/2017/10/12-big-data.html>

FUENTE DE DATOS

El conjunto de datos contiene los últimos datos públicos disponibles sobre COVID-19, incluida una actualización diaria de la situación, la curva epidemiológica y la distribución geográfica global (UE / EEE y el Reino Unido, en todo el mundo).



EU **Open Data** Portal



SCRIPT PARA ACTUALIZAR LA BASE DE DATOS



```
import pandas as pd
import wget
import os
from pymongo import MongoClient

url = 'https://www.ecdc.europa.eu/sites/default/files' \
      '/documents/COVID-19-geographic-disbtributionworldwide.xlsx'

def excel_to_json():
    filename = wget.download(url)
    data = pd.read_excel(filename)
    os.remove(filename)
    return data.to_dict('records')

def upload_data(uri):
    client = MongoClient(uri)
    database = client['covid']
    collection = database['Worldwide']
    collection.drop()
    collection.insert_many(excel_to_json())

upload_data(os.sys.argv[1])
```

CREACIÓN DE LA BASE DE DATOS

- Base de datos documental con *MongoDB*
- Utilización de *Azure Cosmos DB*
- Automatización del proceso de actualización datos



ESTRUCTURA DE LA BASE DE DATOS



Fields	Global Probability	Type
[covid.Worldwide]	100.0%	Collection
▶ _id	100.0%	ObjectId
▶ cases	100.0%	Int32
▶ continentExp	100.0%	String
▶ countriesAndTerritories	100.0%	String
▶ countryterritoryCode	100.0%	String
▶ dateRep	100.0%	Date
▶ day	100.0%	Int32
▶ deaths	100.0%	Int32
▶ geold	100.0%	String
▶ month	100.0%	Int32
▶ popData2018	100.0%	Double
▶ year	100.0%	Int32

