



Universidad Simón Bolívar
Departamento de Ciencias de la Computación
CS3391 – Taller de Bases de Datos

Proyecto 1:

Los últimos 3 años estuvieron marcados por la pandemia más importante que nos ha afectado en los últimos 100 años. Millones de personas murieron y varias organizaciones se dedicaron a recopilar datos que ayudaron a miles de investigadores a analizar lo que realmente estaba pasando.

Una de estas organizaciones es **Our World In Data** <https://ourworldindata.org/coronavirus> Para este proyecto utilizaremos el dataset **owid-covid-data.csv** disponible en <https://github.com/owid/covid-19-data/blob/master/public/data> (aprox. 44MB). En ese mismo enlace conseguirán la descripción de cada uno de los campos del archivo.

Este proyecto consiste en automatizar el proceso de bajar el dataset e importar estos datos a una base de datos PostgreSQL. El modelo final al que se importarán los datos debe estar normalizado, ya que los datos en la fuente original están repetidos por cada día. La recomendación es copiar los datos del CSV a una tabla con la misma estructura, y desde esta copiar los datos a cada tabla final

Adicionalmente será necesario crear múltiples queries que contesten las siguientes preguntas:

- En los datos entre muchas otras existen dos columnas: **new_cases** y **new_cases_smoothed** (suavizados en una ventana de 7 días). En un SELECT, crear una columna nueva calculada en base a **new_cases** que deberá ser igual a **new_cases_smoothed**. ¿Es siempre igual?
- Para cada país, ¿En qué fecha llegó a la mitad del total de casos hasta hoy?
- Para cada país, ¿Cuál es la tasa de muerte promedio antes y después de alcanzar el 66.666% de vacunación? (mostrar columnas separadas para el antes y el después)
- Basado en la pregunta anterior, cual es el porcentaje de vacunación por país que iguala lo más posible la tasa de muerte promedio antes y después de ese porcentaje
- Por cada continente, mostrar la tasa de crecimiento total de COVID per capita (ordenados de mayor a menor)

El/los archivos a entregar deberán ser enviados por email a jhovanny.villegas@gmail.com a más tardar el día 9/Marzo@23:59. Sin embargo, se recomienda hacer el proyecto antes del próximo lunes como una manera de reforzar lo aprendido.

Deben contener:

- Diagrama de la BD
- Script (bash) que:
 - Baja los datos de la fuente original (wget ó equivalente). NO INCLUIR el archivo **owid-covid-data.csv** en el zip a enviar.
 - Comandos para importar estos datos en PostgreSQL (psql)
- Script que crea el *schema* de tablas en la base de datos
- Queries (SELECT) respuesta a las preguntas