# Proyecto - Big Data

Estudiante: Ernesto Rivera

Instrucciones para correr este entregable:

Una vez descomprimido el archivo proyecto\_erivera.zip y abrir una consola (terminal) en la carpeta descomprimida:

1. Para crear la imagen
   1. **docker build -t bigdatafull .**
2. Correr el contenedor para la base de datos:
   1. **docker run --name bigdata-db -e POSTGRES\_PASSWORD=testPassword -p 5433:5432 -d postgres**
3. Correr el contenedor con Jupyter y copier URL
   1. **docker run -p 8888:8888 -i -t --name proyecto bigdatafull**
   2. Copiar el URL con el token que es como <http://127.0.0.1:8888/?token=9763aaa75aea6f0bd38b80ed49dd8ab1bd6f5662994ea1cf>
4. Acceder en el browser (Chrome, Firefox) y pegar el URL copiado en el paso anterior.
5. Abrir el document de Jupyter Proyecto.
6. Si está en Windows o Mac, se puede comentar la segunda línea de código bajo el título “Para cargar los datos en diferentes sistemas operativos”. Omitir este paso si está en Linux.
7. Para conectarse a la base de datos y realizar consultas:
   1. Conectarse al contenedor, en una nueva terminal
      1. **docker exec -it proyecto bash**
   2. Dentro del container, puede usar **psql**
      1. **psql -U postgres -h 172.17.0.1 -p 5433**
      2. password es testPassword
   3. Con psql, se pueden hacer consultas como. El esquema de las tablas se puede ver en el cuaderno de jupyter en <http://127.0.0.1:8888/notebooks/Proyecto.ipynb#Esquema-de-las-tablas-resultantes>
      1. **select column\_name, data\_type, numeric\_scale from information\_schema.columns where table\_name = 'infosafi';**
      2. **select \* from infosafi limit 10;**
      3. **select \* from cartera limit 10;**
      4. **select \* from utilidades limit 10;**
8. Para correr las pruebas **pytest**:
   1. Conectarse al contenedor, en una nueva terminal
      1. **docker exec -it proyecto bash**
      2. **pytest**
9. Para salir del contenedor de Jupyter, hay que oprimir dos veces en el teclado “Ctrl+C”, esto detiene el contenedor.
10. Para detener el contenedor de base de datos
    1. **docker stop bigdata-db**
11. Para borrar ambos contenedores
    1. **docker container rm proyecto**
    2. **docker container rm bigdata-db**