

Instituto Tecnológico de Costa Rica
Programa Ciencia de los Datos
Big Data

Archivos necesarios para ejecutar el código:

- actividad
- build_image
- ciclistas
- conftest
- Dockerfile
- functions
- main
- run_image
- rutas
- test_unit_tests

Instrucciones para el código principal:

1. Descargar y descomprimir el zip en un directorio local.
2. El código debe correrse en un sistema operativo de Linux, o bien con WLS 2 para Windows o máquina virtual de Linux. Debe tener el programa Docker instalado. Este código fue ejecutado mediante el programa Windows Terminal usando Ubuntu 20.04
3. Abra el command prompt de su sistema y busque la dirección en la que guardó la carpeta. Debe nombrarla "tarea_1".
4. Luego, debe construir la imagen. Utilice el comando ***docker <build --tag tarea_1 .>*** En caso de que no funcione pruebe con ***<bash build_image.sh>*** o bien con ***<sh>*** o ***<./>***.
5. Ahora debe correr la imagen con el comando ***<docker run -p 8888:8888 -i -t tarea_1 /bin/bash>*** o en su defecto ***<run_image.sh>***.
6. Dentro del contenedor ejecute el comando ***<spark-submit main.py ciclistas.csv rutas.csv actividad.csv>***.

Instrucciones para las pruebas unitarias:

1. Dentro del contenedor ejecute ***<pytest -s>***. Si desea omitir el contenido de las pruebas utilice ***<pytest>***.