**Programa Ciencia de Datos – FUNDATEC**

**Módulo 4 – Curso Big Data**

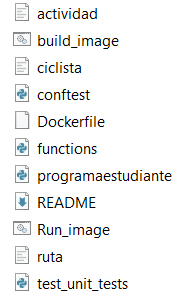
**Tarea 1**

Sebastián Porras y Luis Vargas

Setiembre, 2022

**Instrucciones:** a continuación, se presentan indicaciones para crear un contenedor, ejecutar el programa principal y realizar las respectivas pruebas unitarias.

**Contendido de archivos en la carpeta comprimida:**



* Archivos de datos .csv: actividad, ciclista, ruta
* Archivos para crear y configurar el contenedor: build\_image, Dockerfile, Run\_image
* Archivos .py: conftest, functions, programaestudiante , test\_unit\_tests

**Instrucciones para el código principal:**

Descargar y descomprimir el zip en un directorio local. Nombre esta carpeta ‘Tarea1ModuloIV’.

Abra el command prompt del sistema, navegue los directorios y establezca a la carpeta ‘Tarea1ModuloIV’ como el directorio actual.

Construya la imagen utilizando el comando de docker que contiene el archivo ‘build\_image.sh’, <docker build --tag tarea1 .> Alternativamente, también puede usar <bash build\_image.sh>

Luego, ejecutar la imagen con el comando presente en el archivo ‘Run\_image.sh’, para crear el contenedor: <docker run -p 8888:8888 -i -t tarea1 /bin/bash>

Una vez dentro del contenedor, ejecutar el comando <spark-submit programaestudiante.py ciclista.csv ruta.csv actividad.csv>

**Instrucciones para las pruebas unitarias**

Finalmente, puede usar el comando <pytest –s> para ejecutar las pruebas unitarias definidas en el archivo <test\_unit\_tests> y visualizar su contenido.