

Programa Ciencia de Datos – FUNDATEC

Módulo 4 – Curso Big Data

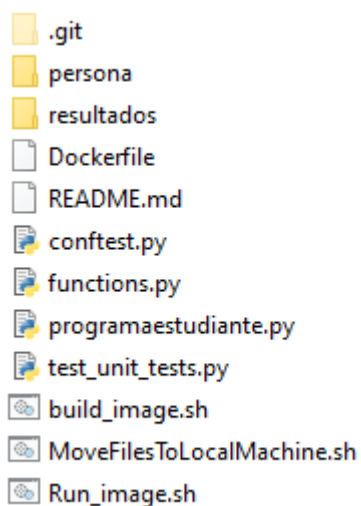
Tarea 2

Sebastián Porras

Setiembre, 2022

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

Contenido de archivos en la carpeta comprimida:



- Carpeta de datos persona
- Archivos para crear y configurar el contenedor: build_image, Dockerfile, Run_image
- Archivos .py: conftest, functions, programaestudiante, test_unit_tests
- Archivos .csv: Dentro de la carpeta resultados

Instrucciones para el código principal:

Descargar y descomprimir el zip en un directorio local. Nombre esta carpeta 'Tarea2ModuloIV'.

Abra el command prompt del sistema, navegue los directorios y establezca a la carpeta 'Tarea2ModuloIV' como el directorio actual.

Construya la imagen utilizando el comando de docker que contiene el archivo 'build_image.sh',
<docker build --tag tarea2 .> 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 tarea2 /bin/bash>

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

Instrucciones para las pruebas unitarias

Finalmente, puede usar el comando `<pytest -s -vv>` para ejecutar las pruebas unitarias definidas en el archivo `<test_unit_tests>` y visualizar su contenido.

En caso de que se quiera extraer los archivos generados entonces se debe, con el container corriendo, abrir otra terminal en la dirección donde se haya guardado la tarea y se debe copiar y pegar el contenido del archivo `MoveFilesToLocalMachine.sh`. Después se generará una carpeta llamada `resultados`.