

Instrucciones de ejecución de Tarea 1

Jose Piedra Venegas

1. Construir imagen de Docker (tarea1_jpiedra)

```
sh build_image.sh
```

2. Correr imagen de Docker (tarea1_jpiedra)

```
sh run_image.sh
```

3. Dentro del shell de la imagen (tarea1_jpiedra) ejecutar spark-submit

```
sh run_programaestudiante.sh
```

4. Las pruebas se ejecutan con

```
pytest
```

y son las siguientes:

- test_multicolumn_join: prueba de join de columnas de DataFrames
- test_aggregation: prueba de agregaciones parciales
- test_top: prueba de promedios ponderados
- test_null: prueba de promedios ponderados con datos nulos (créditos)
- test_single: prueba de promedios ponderados con una sola fila de estudiante
- test_missing: prueba de promedios ponderados con datos faltantes (2 estudiantes)