



Instituto Tecnológico de Costa Rica

Big Data.

Prof. Dr. Juan Manuel Esquivel

William Fallas Coto.

Tarea #1

Conjunto de datos de pruebas.

Los dataset utilizados para la tarea son los siguientes:

- Estudiante.csv
- Curso.csv
- Estudiante.csv

Los mismos fueron generados con una herramienta en línea:

www.mockaroo.com

Comandos importantes

Construir el contenedor

```
docker build --tag bigdata_wfc .
```

Correr la imagen

```
docker run -i -t bigdata_wfc /bin/bash
```

Correr la tarea

```
spark-submit tarea.py "Estudiante.csv","Curso.csv","Nota.csv"
```

Salida esperada

```
20/07/20 20:02:22 INFO CodeGenerator: Code generated
+-----+-----+-----+
| carrera | nombre | Promedio |
+-----+-----+-----+
| Education | Christophorus Lemm | 9.3 |
| Education | Remington Carverhill | 9.1 |
| Sales | Agathe Fibbens | 9.9 |
| Sales | Corabel Grason | 9.8 |
| Engineering | Jermaine Meachem | 9.4 |
| Engineering | Kristien Jewell | 9.3 |
| philosophy | Hernando Woodhouse | 9.9 |
| philosophy | Riordan Amps | 9.8 |
| philosophy | Zechariah Cubbino | 9.8 |
| Accounting | Mona Isakovitch | 9.2 |
| Accounting | Jens Epilet | 9.1 |
| Legal | Peadar Gosdin | 9.7 |
| Legal | Livvy Berndtssen | 9.0 |
| Legal | Haley Hoggins | 9.0 |
| Fisica | Bellina Matfin | 9.0 |
| Fisica | Ethelbert McGoogan | 8.6 |
| Marketing | Roddy Phil | 10.0 |
| Marketing | Jodi Pawlyn | 9.63 |
| Law | Lucilia Carlisso | 9.9 |
| Law | Bartholemy Lownes | 9.9 |
| Mathematics | Franky McNaughton | 9.9 |
| Mathematics | Averyl Lowing | 9.8 |
| Human Resources | Catie Marzele | 9.8 |
| Human Resources | Rog Lillian | 9.1 |
+-----+-----+-----+
```

Archivos contenidos en la carpeta tarea.

funciones.py: Contiene las funciones generales, necesarias en la solución de la tarea.

tarea.py: Programa principal, únicamente hace los llamados a las funciones contenidas en el archivo de funciones.

test_*.py: Contiene los casos de prueba, cada función definida en el archivo funciones.py tiene un archivo test_ con sus respectivas pruebas.