

Instituto Tecnológico de Costa Rica

Programa de Ciencia de los Datos - Módulo Big Data

Tarea #3

- Esteban Sáenz Villalobos (esaenz7@gmail.com)
- Entrega: 29 de agosto 2021, 23:00.
- Observaciones: Ejecutar el programa siguiendo las instrucciones detalladas a continuación.

Instrucciones

1. Para cargar el contenedor con todos los recursos necesarios, ejecute los archivos:

- a) **clean_docker.sh**. Este script borrará contenedores e imágenes antiguos correspondientes a este proyecto. **Atención: el comando realiza una acción de "prune" para la limpieza.**
- b) **build_image.sh**. Construye una imagen a partir del archivo DockerFile.
- c) **run_image.sh**. Este script creará 2 contenedores y una red local en docker de la siguiente forma:

```
* Red: bigdatanet, IP: 10.7.84.0/24.  
* Host principal (sesión bash donde se ejecutarán los  
comandos): bigdata_tarea3_esv_1, IP: 10.7.84.101.  
* Host secundario (base de datos): bigdata_tarea3_esv_2  
(postgres), IP: 10.7.84.102.
```

d) Estos parámetros corresponden a la instancia de postgres dentro del ambiente de docker:

```
* Host: 10.7.84.102  
* Puerto: 5432  
* Usuario: postgres  
* Clave: testPassword
```

2. Programa principal:

a) Ejecute el archivo:

```
#run_main.sh
```

Este comando ejecutará un query psql para crear las tablas necesarias en la base de datos postgres. Luego cargará el servidor de Jupyter para acceder al directorio en donde se encuentra el notebook con el trabajo propiamente.

El notebook "BIGDATA_07_2021_Tarea3_ESV.ipynb" almacena los recursos, el código, la documentación y los resultados solicitados. Se ejecuta de forma completa y secuencial. Adicional se incluyen copias en formatos HTML y PDF.

b) Luego de cerrar la sesión del servidor de Jupyter se puede ejecutar el siguiente comando desde la misma sesión BASH del contenedor principal.

```
#run_read.sh
```

Este comando ejecutará un query psql para realizar una lectura de los datos contenidos en las tablas recién creadas y utilizadas por el código del notebook.

c) Una vez haya terminado, puede ejecutar el comando **clean_docker.sh** en el host de docker para limpiar los recursos creados para este trabajo.

- El repositorio completo de la tarea se encuentra también en el siguiente enlace [github/esaenz7](https://github.com/esaenz7).
-