

Tarea evaluable

CE_5075 7.1

Big data aplicado



APARTADO 1

Siguiendo las instrucciones que hemos visto en el apartado 6 de los apuntes, crea un clúster sobre Docker con 1 nodo maestro, 3 nodos worker y un nodo para JupyterLab. Puedes utilizar las mismas imágenes que hemos usado en los apuntes. El worker 3 debe exponer el puerto 8083 para su interfaz web.

Aprovechamos las imágenes ofrecidas en los apuntes y montamos el clúster de Docker.

```
david@srv-ubuntu ~/cluster (0.033s)
ll
total 28
drwxrwxr-x  2 david david 4096 Mar 30 15:55 ./
drwxr-x--- 17 david david 4096 Mar 30 14:37 ../
-rw-rw-r--  1 david david  626 Mar 29 19:27 build.sh
-rw-rw-r--  1 david david 1007 Mar 30 15:55 docker-compose.yml
-rw-rw-r--  1 david david 1539 Mar 29 19:22 jupyterlab.Dockerfile
-rw-rw-r--  1 david david 1953 Mar 29 19:20 spark-master.Dockerfile
-rw-rw-r--  1 david david 1897 Mar 29 19:45 spark-worker.Dockerfile
```

```
david@srv-ubuntu ~/cluster (0.064s)
sudo docker images
REPOSITORY          TAG                 IMAGE ID            CREATED             SIZE
jupyterlab           latest              ce089b04437d       About a minute ago 2.11GB
spark-worker         latest              8c9b7844b848       4 minutes ago      1.38GB
spark-master         latest              bb3130112246       8 minutes ago      1.38GB
```

Además modificamos el docker-compose.yml para añadir un worker más, será el "spark-worker-3" y le asignamos el puerto especificado por el enunciado.

```
version: "3.8"

volumes:
  hdfs-simulat:
    name: "hadoop-distributed-file-system"
    driver: local

services:
  spark-master:
    image: spark-master
    container_name: spark-master
    ports:
      - 7077:7077
      - 8080:8080
    volumes:
      - hdfs-simulat:/opt/workspace

  spark-worker-1:
    image: spark-worker
    container_name: spark-worker-1
    ports:
      - 8081:8081
    volumes:
      - hdfs-simulat:/opt/workspace
    depends_on:
      - spark-master
```

```
  spark-worker-2:
    image: spark-worker
    container_name: spark-worker-2
    ports:
      - 8082:8081
    volumes:
      - hdfs-simulat:/opt/workspace
    depends_on:
      - spark-master

  spark-worker-3:
    image: spark-worker
    container_name: spark-worker-3
    ports:
      - 8083:8081
    volumes:
      - hdfs-simulat:/opt/workspace
    depends_on:
      - spark-master

  jupyterlab:
    image: jupyterlab
```

Y después de levantar los contenedores con “`sudo docker-compose up`” comprobamos en otra terminal que están activos.

```
sudo docker container ls
[sudo] password for david:
CONTAINER ID   IMAGE          COMMAND                  CREATED        STATUS        PORTS                                                                 NAMES
c98fc6d0c7a0   spark-worker   "/bin/sh -c 'bin/spa..." 16 seconds ago Up 15 seconds 0.0.0.0:8083->8081/tcp, [::]:8083->8081/tcp spark
d4289b084fc67   spark-worker   "/bin/sh -c 'bin/spa..." 2 hours ago   Up 15 seconds 0.0.0.0:8082->8081/tcp, [::]:8082->8081/tcp spark
b48ce7123653   spark-worker   "/bin/sh -c 'bin/spa..." 2 hours ago   Up 15 seconds 0.0.0.0:8081->8081/tcp, [::]:8081->8081/tcp spark
2538282cdd82   spark-master   "/bin/sh -c 'bin/spa..." 2 hours ago   Up 15 seconds 0.0.0.0:7077->7077/tcp, [::]:7077->7077/tcp, 0.0.0.0:8080->8080/tcp, [::]:8080->8080/tcp spark-master
f6316c194ea6   jupyterlab     "/bin/sh -c 'jupyter..." 2 hours ago   Up 15 seconds 0.0.0.0:8888->8888/tcp, [::]:8888->8888/tcp jupyterlab
```

Si probamos a acceder a la interfaz web del worker 3 vemos que funciona.

