

INSTALACIÓN DE HADOOP EN DOCKER

Instalación Hadoop en Docker

Paso 1: Clonar el repositorio de Docker-Hadoop desde GitHub

1. Abre la terminal cmd.
2. Clona el repositorio de Docker-Hadoop ejecutando el siguiente comando:

```
C:\Users\ESP\Desktop\SBD\Practica Hadoop Docker\docker-hadoop>git clone https://github.com/big-data-europe/docker-hadoop
```

Paso 2: Iniciar los contenedores necesarios usando Docker Compose

1. Cambia al directorio del repositorio clonado:
2. Inicia los contenedores necesarios con el siguiente comando:
3. Este comando ejecutará todos los servicios definidos en el archivo `docker-compose.yml`. La primera vez que se ejecute, descargará las imágenes necesarias de Docker.

```
C:\Users\ESP\Desktop\SBD\Practica Hadoop Docker\docker-hadoop>docker-compose up -d
[+] Running 9/9
  Network docker-hadoop_default          Created           0.2s
  Volume "docker-hadoop_hadoop_namenode" Created           0.0s
  Volume "docker-hadoop_hadoop_datanode" Created           0.0s
  Volume "docker-hadoop_hadoop_historyserver" Created           0.0s
  Container datanode                     Started           5.8s
  Container namenode                     Started           6.1s
  Container resourcemanager              Started           5.2s
  Container historyserver                 Started           5.7s
  Container nodemanager                  Started           6.6s
```

Paso 3: Acceder al contenedor Docker

1. Verifica que los cinco contenedores estén corriendo con:

```
C:\Users\ESP\Desktop\SBD\Practica Hadoop Docker\docker-hadoop>docker ps
CONTAINER ID   IMAGE                                COMMAND                  CREATED        STATUS
9123b82259c7   bde2020/hadoop-nodemanager:2.0.0-hadoop3.2.1-java8  "/entrypoint.sh /run..." About a minute ago Up About a minute (healthy)
92b79a698494   bde2020/hadoop-datanode:2.0.0-hadoop3.2.1-java8     "/entrypoint.sh /run..." About a minute ago Up About a minute (healthy)
c18825d67663   bde2020/hadoop-resourcemanager:2.0.0-hadoop3.2.1-java8 "/entrypoint.sh /run..." About a minute ago Up About a minute (healthy)
2ca35ceb0de8   bde2020/hadoop-namenode:2.0.0-hadoop3.2.1-java8     "/entrypoint.sh /run..." About a minute ago Up About a minute (healthy)
9f53eb7f8060   bde2020/hadoop-historyserver:2.0.0-hadoop3.2.1-java8 "/entrypoint.sh /run..." About a minute ago Up About a minute (healthy)
8d4452f29e2d   apache/nifi-registry:latest          nifi-registry           2 weeks ago   Up About a minute
24d6ce861c19   apache/nifi:latest                   nifi                     2 weeks ago   Up About a minute
```

2. Para entrar en el contenedor docker, ejecuta:

```
C:\Users\ESP\Desktop\SBD\Practica Hadoop Docker>docker exec -it namenode bash
```

Listado de comandos de Hadoop

1. Listar archivos en un directorio:

```
root@2ca35ceb0de8:/# hdfs dfs -ls /user/  
Found 1 items  
drwxr-xr-x  - root supergroup          0 2024-10-24 17:45 /user/root  
root@2ca35ceb0de8:/#
```

2. Crear un directorio:

```
root@2ca35ceb0de8:/# hdfs dfs -mkdir -p /user/root
```

3. Copiar archivos del sistema de archivos local a HDFS:

```
C:\Users\ESP>docker cp C:\Users\ESP\Desktop\SBD\ejemplo.txt namenode:/tmp  
Successfully copied 2.05kB to namenode:/tmp
```

```
root@2ca35ceb0de8:/# hdfs dfs -copyFromLocal /tmp/ejemplo.txt /user/root  
2024-10-24 18:12:06,965 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false  
root@2ca35ceb0de8:/# hdfs dfs -ls /user/root  
Found 1 items  
-rw-r--r--  3 root supergroup          27 2024-10-24 18:12 /user/root/ejemplo.txt
```

4. Ver el contenido de un archivo:

```
root@2ca35ceb0de8:/tmp# hdfs dfs -cat ejemplo.txt  
2024-10-27 18:44:02,040 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false  
, remoteHostTrusted = false  
Esto es un ejemplo de texto
```

5. Copiar un fichero de una ruta a otra dentro de HDFS

```
root@2ca35ceb0de8:/tmp# hdfs dfs -cp ejemplo.txt /out  
2024-10-27 18:47:12,898 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false  
, remoteHostTrusted = false  
2024-10-27 18:47:13,231 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false  
, remoteHostTrusted = false
```

6. Mover archivos en HDFS:

```
root@2ca35ceb0de8:/tmp/out# hdfs dfs -mv ejemplo.txt /in
```

7. Eliminar archivos o directorios en HDFS:

```
root@2ca35ceb0de8:/tmp# hdfs dfs -rm /out  
Deleted /out
```

8. Acceder a la ayuda para ver más comandos:

```
root@2ca35ceb0de8:/tmp/out# hdfs dfs -help  
Usage: hadoop fs [generic options]  
[-appendToFile <localsrc> ... <dst>]  
[-cat [-ignoreCrc] <src> ...]  
[-checksum <src> ...]  
[-chgrp [-R] GROUP PATH...]  
[-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]  
[-chown [-R] [OWNER][:[GROUP]] PATH...]  
[-copyFromLocal [-f] [-p] [-l] [-d] [-t <thread count>] <localsrc>  
[-copyToLocal [-f] [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]  
[-count [-q] [-h] [-v] [-t <storage type>]] [-u] [-x] [-e] <path>  
[-cp [-f] [-p | -p[topax]] [-d] <src> ... <dst>]  
[-createSnapshot <snapshotDir> [<snapshotName>]]  
[-deleteSnapshot <snapshotDir> <snapshotName>]  
[-df [-h] [<path> ...]]  
[-du [-s] [-h] [-v] [-x] <path> ...]  
[-expunge [-immediate]]  
[-find <path> ... <expression> ...]  
[-get [-f] [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]  
[-getfacl [-R] <path>]
```