



UNIVERSIDAD
DE
CÓRDOBA

**Instituto de Estudios de Posgrado
Universidad de Córdoba**

MÁSTER UNIVERSITARIO EN INTELIGENCIA COMPUTACIONAL E
INTERNET DE LAS COSAS

**INSTALACIÓN DE HADOOP Y
CONFIGURACIÓN DE HDFS**

Introducción al Big Data Analysis

Autora:

Alba Márquez-Rodríguez

Profesora:

Aurora Ramírez Quesada

Córdoba, Diciembre 2023

Índice

1. Introducción	2
2. Ejercicios	3
2.1. Ejercicio de comprobación de la instalación	3
2.1.1. Comprobar el estado del sistema HDFS	3
2.1.2. Crear un directorio de prueba para el ejemplo	4
2.1.3. Copiar un fichero en el directorio de usuario	4
2.1.4. Ejecución de un ejemplo MapReduce	4
2.1.5. Visualización del resultado	5
2.1.6. Terminar la Ejecución	7

1. Introducción

Estas prácticas consisten en completar el proceso de instalación y configuración según las secciones del documento de la práctica. Al final del documento se han propuesto ejercicios para comprobar que la instalación es correcta.

En esta memoria se indica en cada paso el comando de Hadoop utilizado y muestra la salida obtenida mediante una captura de pantalla.

2. Ejercicios

Se ha completado la instalación y comprobaciones recomendadas en el documento de la práctica. A continuación se explica y adjuntan las capturas de los ejercicios propuestos.

2.1. Ejercicio de comprobación de la instalación

En este apartado se han realizado una serie de pasos para comprobar la correcta instalación de Hadoop. En cada paso, se ha debido ejecutar el comando adecuado y realizar una captura de pantalla del resultado obtenido por consola.

2.1.1. Comprobar el estado del sistema HDFS

Para comprobar el estado del sistema debemos ejecutar el comando `texttt./hdfs fsck /` mientras nos encontramos en el directorio `/hadoop-3.3.6/bin`. El resultado obtenido es el siguiente:

```
osboxes@osboxes:~/hadoop-3.3.6/bin$ ./hdfs fsck /
Connecting to namenode via http://localhost:9870/fsck?ugi=osboxes&path=%2F
FSCK started by osboxes (auth:SIMPLE) from /127.0.0.1 for path / at Thu Dec 14 10:15:27 EST 2023

Status: HEALTHY
Number of data-nodes: 1
Number of racks: 1
Total dirs: 1
Total symlinks: 0

Replicated Blocks:
Total size: 0 B
Total files: 0
Total blocks (validated): 0
Minimally replicated blocks: 0
Over-replicated blocks: 0
Under-replicated blocks: 0
Mis-replicated blocks: 0
Default replication factor: 1
Average block replication: 0.0
Missing blocks: 0
Corrupt blocks: 0
Missing replicas: 0
Blocks queued for replication: 0

Erasure Coded Block Groups:
Total size: 0 B
Total files: 0
Total block groups (validated): 0
Minimally erasure-coded block groups: 0
Over-erasure-coded block groups: 0
Under-erasure-coded block groups: 0
Unsatisfactory placement block groups: 0
Average block group size: 0.0
Missing block groups: 0
Corrupt block groups: 0
Missing internal blocks: 0
Blocks queued for replication: 0
FSCK ended at Thu Dec 14 10:15:27 EST 2023 in 14 milliseconds
```

Con esto hemos comprobado que el estado del sistema de ficheros creado es correcto.

2.1.2. Crear un directorio de prueba para el ejemplo

Para probar la ejecución de un proceso MapReduce que escriba sobre el sistema de ficheros recién creado, se debe añadir un directorio denominado *prueba*.

Estando en el directorio `/hadoop-3.3.6` se ha ejecutado el comando `bin/hdfs dfs -mkdir /user/osboxes/prueba`. Esta vez no se ha obtenido nada por la terminal.

2.1.3. Copiar un fichero en el directorio de usuario

Se deben escribir tres comandos para realizar las siguientes operaciones:

1. Copiar un fichero local al sistema HDFS para usarlo como entrada del proceso MapReduce. El fichero local será el de licencia de uso de Hadoop, disponible en el directorio de instalación de Hadoop (ubicación: `/home/¡tu_usuario!/hadoop-3-3-6`)
2. Mostrar el contenido del directorio actual.

Estando en el directorio `/hadoop-3.3.6` se ha ejecutado el comando `bin/hdfs dfs -put /home/osboxes/hadoop-3.3.6/LICENSE.txt /user/osboxes/prueba`. Sin obtener nada por la terminal.

Luego se ha ejecutado el comando `bin/hdfs dfs -ls` pero al haber guardado el fichero en la carpeta *prueba* no se mostraba el fichero copiado. Por eso se ha ejecutado el comando `bin/hdfs dfs -ls prueba`. Lo obtenido es lo siguiente:

```
osboxes@osboxes:~/hadoop-3.3.6$ bin/hdfs dfs -ls
Found 1 items
drwxr-xr-x  - osboxes supergroup          0 2023-12-14 10:38 prueba
osboxes@osboxes:~/hadoop-3.3.6$ bin/hdfs dfs -ls prueba
Found 1 items
-rw-r--r--   1 osboxes supergroup      15217 2023-12-14 10:38 prueba/LICENSE.txt
```

2.1.4. Ejecución de un ejemplo MapReduce

Estando en el directorio `/hadoop-3.3.6` se ha ejecutado el comando `bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.6.jar wordcount hdfs:///user/osboxes/prueba/LICENSE.txt hdfs:///user/osboxes/prueba/output`. La prueba de que se ha ejecutado correctamente es la siguiente:

```

FILE: Number of bytes written=1881287
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=30434
HDFS: Number of bytes written=9894
HDFS: Number of read operations=15
HDFS: Number of large read operations=0
HDFS: Number of write operations=4
HDFS: Number of bytes read erasure-coded=0
Map-Reduce Framework
  Map input records=270
  Map output records=1672
  Map output bytes=20756
  Map output materialized bytes=12507
  Input split bytes=118
  Combine input records=1672
  Combine output records=657
  Reduce input groups=657
  Reduce shuffle bytes=12507
  Reduce input records=657
  Reduce output records=657
  Spilled Records=1314
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=30
  Total committed heap usage (bytes)=220200960
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=15217
File Output Format Counters
  Bytes Written=9894

```

```
osboxes@osboxes:~/hadoop-3.3.6$
```

2.1.5. Visualización del resultado

Se debe mostrar por consola el resultado accediendo al directorio de salida. Ejecutando el comando `bin/hdfs dfs -text prueba/output/part-r-00000` se obtiene el siguiente resultado:

trade	1	
trademark,		1
trademarks,		1
transfer		1
transformation		1
translation		1
types.	1	
under	10	
union	1	
unless	1	
uriparser2		1
use	5	
use,	4	
using	1	
various	1	
verbal,	1	
version	1	
warranties		1
warranty		1
warranty,		1
was	1	
where	1	
wherever		1
whether	4	
which	2	
whole,	2	
whom	1	
with	11	
within	8	
without	3	
work	5	
work,	2	
work.	1	
works	1	

2.1.6. Terminar la Ejecución

Se debe utilizar el script adecuado para deshabilitar los nodos del sistema de ficheros HDFS. Este fichero es el fichero `stop-dfs.sh` que se ejecutando con el comando `sbin/stop-dfs.sh`. Por terminal se muestra lo siguiente:

```
osboxes@osboxes:~/hadoop-3.3.6$ sbin/stop-dfs.sh
Stopping namenodes on [localhost]
Stopping datanodes
Stopping secondary namenodes [osboxes]
```


Referencias