

Análisis aerolínea

Práctica obligatoria

Antonio Cabrera

Trabajo para el doble grado de
Ingeniería del Software y Matemática Computacional



Asignatura de procesamiento de datos

U-tad

España

Mayo 2024

Contents

1	Preparación del entorno de la máquina virtual	3
1.1	Preparación de Cassandra	3
2	Desplegar HDFS	6
3	Desplegar PostgreSQL	8

List of Figures

1.1	Actualización del sistema operativo	3
1.2	Instalación de Python 2	4
1.3	Descarga de Cassandra	4
1.4	Descompresión de Cassandra	4
1.5	Eliminación del archivo .tar.gz	5
2.1	Despliegue de HDFS	6
2.2	Comprobación de HDFS	6
2.3	Comando de HDFS	7

Chapter 1

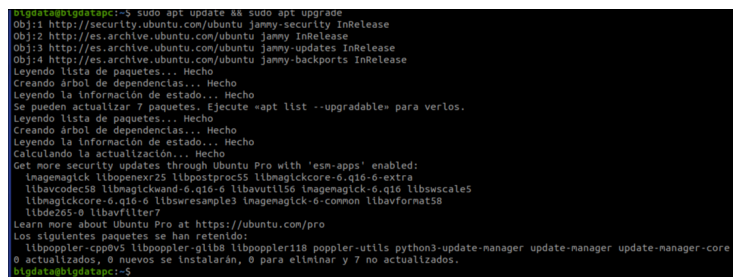
Preparación del entorno de la máquina virtual

Partiremos de la máquina virtual proporcionada por el profesor, la cual tiene instalado el sistema operativo Ubuntu 22.04.3 LTS.

1.1 Preparación de Cassandra

Primero instalaremos Cassandra, para ello primero actualizaremos el sistema operativo. El comando `sudo apt update` actualiza la lista de paquetes disponibles y sus versiones, mientras que el comando `sudo apt upgrade` instala las actualizaciones disponibles.

```
1 sudo apt update && sudo apt upgrade
```

A terminal window showing the output of the command 'sudo apt update && sudo apt upgrade'. The output includes updates for security, jammy-security, jammy-backports, and various system packages like libmagick, libavcodec, and libavformat. It also lists packages that are held back and provides a link to learn more about Ubuntu Pro.

```
bigdata@bigdatapc:~$ sudo apt update && sudo apt upgrade
Obj:1 http://security.ubuntu.com/ubuntu jammy-security InRelease
Obj:2 http://es.archive.ubuntu.com/ubuntu jammy InRelease
Obj:3 http://es.archive.ubuntu.com/ubuntu jammy-updates InRelease
Obj:4 http://es.archive.ubuntu.com/ubuntu jammy-backports InRelease
Leyendo lista de paquetes... Hecho
Creando árbol de dependencias... Hecho
Leyendo la información de estado... Hecho
Se pueden actualizar 7 paquetes. Ejecute «apt list --upgradable» para verlos.
Leyendo lista de paquetes... Hecho
Creando árbol de dependencias... Hecho
Leyendo la información de estado... Hecho
Calculando la actualización... Hecho
Get more security updates through ubuntu Pro with 'esm-apps' enabled:
inagemagick libopenexr25 libpostproc55 libmagickcore-6.q16-6-extra
libavcodec58 libmagickwand-6.q16-6 libavutil56 inagemagick-6.q16 libswscales
libmagickcore-6.q16-6 libswresample3 inagemagick-6-common libavformat58
libdecor-0 libavfilter7
Learn more about Ubuntu Pro at https://ubuntu.com/pro
Los siguientes paquetes se han retenido:
  libpoppler-cpp0v5 libpoppler-glib8 libpoppler-isa poppler-utils python3-update-manager update-manager update-manager-core
0 actualizados, 0 nuevos se instalarán, 0 para eliminar y 7 no actualizados.
bigdata@bigdatapc:~$
```

Figure 1.1: Actualización del sistema operativo

Ahora instalaremos Python 2, ya que Cassandra requiere esta versión de Python. Para ello, ejecutamos el siguiente comando:

```
1 sudo apt install python2
```

```
bigdata@bigdatapc:~$ sudo apt install python2
Leyendo lista de paquetes... Hecho
Creando árbol de dependencias... Hecho
Leyendo la información de estado... Hecho
python2 ya está en su versión más reciente (2.7.18-3).
0 actualizados, 0 nuevos se instalarán, 0 para eliminar y 7 no actualizados.
bigdata@bigdatapc:~$
```

Figure 1.2: Instalación de Python 2

Después, descargaremos el archivo .tar.gz de Cassandra desde la página oficial de Apache. Para ello, ejecutamos el siguiente comando:

```
1 wget https://dlcdn.apache.org/cassandra/3.11.16/apache-cassandra-3.11.16-bin.tar.gz
```

```
bigdata@bigdatapc:~$ wget https://dlcdn.apache.org/cassandra/3.11.16/apache-cassandra-3.11.16-bin.tar.gz
--2024-05-10 20:03:05-- https://dlcdn.apache.org/cassandra/3.11.16/apache-cassandra-3.11.16-bin.tar.gz
Resolviendo dlcdn.apache.org (dlcdn.apache.org)... 151.101.2.132, 2a01:4ee2::1644
Conectando con dlcdn.apache.org (dlcdn.apache.org)[151.101.2.132]:443... conectado.
Petición HTTP enviada, esperando respuesta... 200 OK
Longitud: 31111361 (30M) [application/x-gzip]
Guardando como: 'apache-cassandra-3.11.16-bin.tar.gz.1'
apache-cassandra-3.11.16-bin.tar.g 100%[=====] 29,67M 9,33MB/s en 3,2s
2024-05-10 20:03:08 (9,33 MB/s) - 'apache-cassandra-3.11.16-bin.tar.gz.1' guardado [31111361/31111361]
bigdata@bigdatapc:~$
```

Figure 1.3: Descarga de Cassandra

Descomprimos el archivo .tar.gz con el siguiente comando:

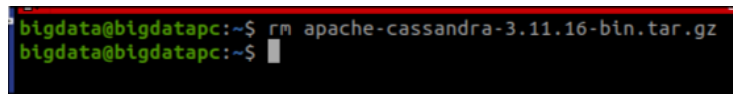
```
1 tar -xvzf apache-cassandra-3.11.16-bin.tar.gz
```

```
bigdata@bigdatapc:~$ tar -xvzf apache-cassandra-3.11.16-bin.tar.gz
apache-cassandra-3.11.16/bin/
apache-cassandra-3.11.16/conf/
apache-cassandra-3.11.16/conf/triggers/
apache-cassandra-3.11.16/doc/
apache-cassandra-3.11.16/doc/cql3/
apache-cassandra-3.11.16/interface/
apache-cassandra-3.11.16/lib/
apache-cassandra-3.11.16/lib/sigar-bin/
apache-cassandra-3.11.16/pylib/
apache-cassandra-3.11.16/pylib/cqlshlib/
apache-cassandra-3.11.16/pylib/cqlshlib/test/
apache-cassandra-3.11.16/pylib/cqlshlib/test/config/
apache-cassandra-3.11.16/tools/
apache-cassandra-3.11.16/tools/bin/
apache-cassandra-3.11.16/tools/lib/
apache-cassandra-3.11.16/CASSANDRA-14092.txt
apache-cassandra-3.11.16/CHANGES.txt
apache-cassandra-3.11.16/LICENSE.txt
apache-cassandra-3.11.16/NEWS.txt
```

Figure 1.4: Descompresión de Cassandra

Por último, eliminamos el archivo .tar.gz con el siguiente comando:

```
1 rm apache-cassandra-3.11.16-bin.tar.gz
```

A terminal window with a black background and green text. The prompt is 'bigdata@bigdatapc:~\$'. The command 'rm apache-cassandra-3.11.16-bin.tar.gz' has been entered and executed. The prompt is now 'bigdata@bigdatapc:~\$' with a cursor at the end.

```
bigdata@bigdatapc:~$ rm apache-cassandra-3.11.16-bin.tar.gz
bigdata@bigdatapc:~$
```

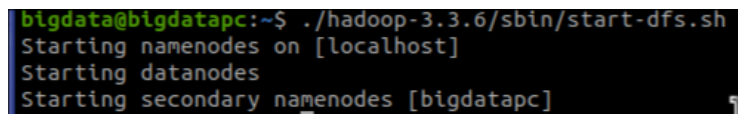
Figure 1.5: Eliminación del archivo .tar.gz

Chapter 2

Desplegar HDFS

HDFS ya está instalado por defecto en la máquina virtual en la carpeta `/hadoop-3.3.6`. Para desplegar HDFS ejecutamos el siguiente comando:

```
1 ~/hadoop-3.3.6/sbin/start-dfs.sh
```

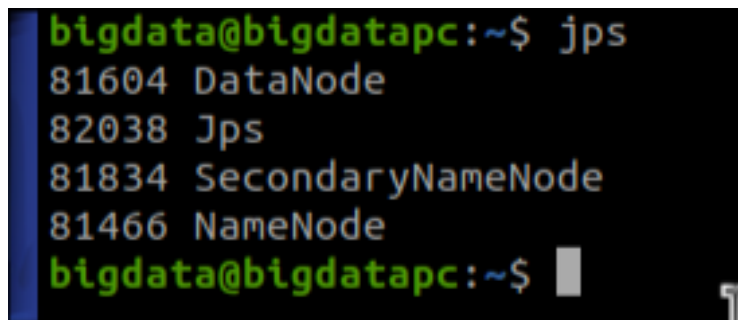


```
bigdata@bigdatapc:~$ ./hadoop-3.3.6/sbin/start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [bigdatapc]
```

Figure 2.1: Despliegue de HDFS

Ahora para comprobar que HDFS se ha desplegado correctamente, ejecutamos el siguiente comando:

```
1 jps
```



```
bigdata@bigdatapc:~$ jps
81604 DataNode
82038 Jps
81834 SecondaryNameNode
81466 NameNode
bigdata@bigdatapc:~$
```

Figure 2.2: Comprobación de HDFS

Ahora ya podemos ejecutar comandos de HDFS, como por ejemplo el siguiente comando que muestra los archivos en el sistema de archivos HDFS:

```
1 ~/hadoop-3.3.6/bin/hdfs dfs -ls /
```

```
bigdata@bigdatapc:~$ ~/hadoop-3.3.6/bin/hdfs dfs -ls /  
Found 1 items  
drwx-wx-wx - bigdata_superuser 0 2024-01-29 01:29 /tmp
```

Figure 2.3: Comando de HDFS

Chapter 3

Desplegar PostgreSQL

Postgres también está instalado por defecto en la máquina virtual, además se arranca por defecto al iniciar la sesión en la máquina virtual. Para comprobar que Postgres se ha desplegado correctamente, ejecutamos el siguiente comando:

```
1 sudo systemctl status postgresql
```