

Funcionamiento de un programa MapReduce

Para comprobar que la instalación realizada funciona correctamente, se ha ejecutado un programa MapReduce sobre el nodo configurado.

Todos los pasos descritos a continuación se han realizado una vez que la máquina virtual ha sido creada y aprovisionada con **vagrant up** y para el usuario “vagrant”.

Dentro de la carpeta del proyecto, abrimos un terminal y lanzamos el comando **vagrant ssh** para entrar dentro de la máquina virtual.

Una vez levantada la máquina virtual, dentro del directorio del proyecto se crea una carpeta llamada “sync”. Esa carpeta está sincronizada con la carpeta remota de la máquina virtual llamada “vagrant”. Listamos el contenido de esa carpeta con el comando **ls /vagrant**:

```
cdh5-repository_1.0_all.deb
```

La carpeta únicamente muestra un paquete descargado y utilizado durante el aprovisionamiento de la máquina. Volvemos a la carpeta “sync” y guardamos en ella el archivo “donquijote.txt” y la carpeta comprimida “WordCountSimple” disponibles en el espacio virtual de la asignatura. Descomprimos la carpeta, escribimos **exit** en el terminal para salir de la máquina virtual y escribimos **vagrant reload** para que la carpeta sync vuelva a sincronizarse con la carpeta remota /vagrant (es posible que este paso no sea necesario y que la máquina virtual detecte el nuevo contenido de forma automática, pero es un paso que recomienda hacer la documentación oficial de Vagrant). Entramos de nuevo en la máquina virtual con **vagrant ssh** y listamos con **ls /vagrant** el contenido de la carpeta sincronizada. Ahora se puede ver que dentro de la máquina virtual existen los nuevos archivos que hemos añadido:

```
cdh5-repository_1.0_all.deb donquijote.txt WordCountSimple
```

Creamos un directorio de trabajo dentro del HDFS y copiamos en él el archivo “donquijote.txt”:

```
hadoop fs -mkdir trabajo
```

```
hadoop fs -put /vagrant/donquijote.txt trabajo
```

Para comprobar que el archivo se ha copiado de forma correcta, mostramos el contenido de la carpeta de trabajo con el comando `hadoop fs -ls trabajo`:

```
Found 1 items
-rw-r--r-- 1 vagrant supergroup 2198903 2020-06-24 09:08 trabajo/donquijote.txt
```

Entramos dentro de la carpeta con el comando `cd /vagrant/WordCountSimple` y compilamos el programa mediante el comando `make`, el cual nos producirá un archivo ".jar":

```
bin Makefile src WordCountSimple.jar
```

Una vez hecho esto, salimos del directorio con el comando `cd` y ejecutamos el trabajo MapReduce:

```
hadoop jar /vagrant/WordCountSimple/WordCountSimple.jar trabajo
salida
```

Se obtiene una salida parcial parecida a la siguiente:

```
20/06/24 10:23:25 WARN mapred.JobClient: Use GenericOptionsParser for parsing the
20/06/24 10:23:26 INFO input.FileInputFormat: Total input paths to process : 1
20/06/24 10:23:26 INFO mapred.JobClient: Running job: job_202006240903_0001
20/06/24 10:23:27 INFO mapred.JobClient: map 0% reduce 0%
20/06/24 10:23:37 INFO mapred.JobClient: map 100% reduce 0%
20/06/24 10:23:42 INFO mapred.JobClient: map 100% reduce 100%
20/06/24 10:23:45 INFO mapred.JobClient: Job complete: job_202006240903_0001
```

Una vez ejecutado el trabajo, se habrá creado una carpeta llamada "salida". Para ver el contenido de esa carpeta, lo listamos con el comando `hadoop fs -ls salida`:

```
Found 3 items
-rw-r--r-- 1 vagrant supergroup 0 2020-06-24 10:23 salida/_SUCCESS
drwxr-xr-x - vagrant supergroup 0 2020-06-24 10:23 salida/_logs
-rw-r--r-- 1 vagrant supergroup 448894 2020-06-24 10:23 salida/part-r-00000
```

El resultado de esta ejecución se encuentra en el archivo "part-r-00000". Para poder ver este resultado, lo recuperamos del HDFS y lo copiamos a la carpeta /vagrant con el siguiente comando:

```
hadoop fs -get salida/part-r-00000 /vagrant
```

Con lo anterior, ya tendríamos el resultado recuperado en nuestra carpeta sincronizada “sync” de la máquina anfitriona y podríamos consultar el resultado abriéndolo con algún editor de texto:

```
1  !Mal 1
2  "Al 1
3  "Cuando 2
4  "Cuidados 1
5  "De 2
6  "Defects," 1
7  "Desnudo 1
8  "Dijo 1
9  "Dime 1
10 "Don 1
```

Una vez realizados todos los pasos anteriores, se puede comprobar que la instalación se ha realizado de forma correcta y que tanto Hadoop como MapReduce presentan un funcionamiento adecuado.