

1_Descubriendo HDFS

1. Lo primero que tenemos que hacer es habilitar la transferencia de ficheros entre vuestra máquina virtual y vuestro host. En las settings del VMWare asociadas a la VMCloudera que tenéis, habilitad la opción shared folders y añadid una carpeta en vuestros host Windows desde donde copiaremos los archivos necesarios para trabajar.

2. Una vez creada, arrancad la MV. La carpeta en Windows compartida estará en la ruta “/mnt/hgfs”

3. Abre un terminal en la MV y cámbiale el idioma a español

4. Escribe el siguiente comando

a. Hadoop fs

5. Escribe

a. Hadoop fs -ls /

6. Crea un directorio local de trabajo hadoop

a. Por ejemplo /home/cloudera/ejercicios

mkdir ejercicios

rmdir ejercicios (si está vacío)

rm -r borrar en cascada

b. Copia el archivo “shakespeare.tar.gz” en ese directorio

cp origen destino

c. Descomprime el archivo cd ..

i. “tar zxvf shakespeare.tar.gz”

d. Copia la carpeta que acabas de descomprimir en HDFS en la ruta

/user/cloudera/shakespeare. Tu home a partir de ahora será /user/cloudera

i. “hadoop fs -put shakespeare /user/cloudera/shakespeare”

7. Lista el contenido de tu home en HDFS para comprobar que se ha copiado la carpeta shakespeare.

i. “hadoop fs -ls /user/cloudera”

8. Observa el contenido de la carpeta shakespeare en hdfs

i. “hadoop fs -ls /user/cloudera/shakespeare”

9. Borra la subcarpeta “glossary” de la carpeta shakespeare en HDFS

a. “hadoop fs -rm /user/cloudera/shakespeare/glossary”

10. Comprueba que se ha borrado

11. Lista las primeras 50 últimas líneas de la subcarpeta “histories”. Puedes usar los comando “cat” y “tail” (head -> primeras)

a. “hadoop fs -cat /user/cloudera/shakespeare/histories | tail -n 50”

12. Copia al Sistema de ficheros local de tu MV el fichero “poems” en la ruta

/home/cloudera/ejercicios/shakespeare/shakepoems.txt

a. “hadoop fs -get /user/cloudera/shakespeare/poems

/home/cloudera/ejercicios/shakespeare/shakepoems.txt”

13. Muestra las últimas líneas de shakepoems.txt copiado en tu local por pantalla

tail 5 shakepoems.txt

14. Si has terminado muy rápido, juega un poco con los comandos disponibles en la Shell de hadoop fs. Para ello introduce “hadoop fs” y observa las posibilidades.