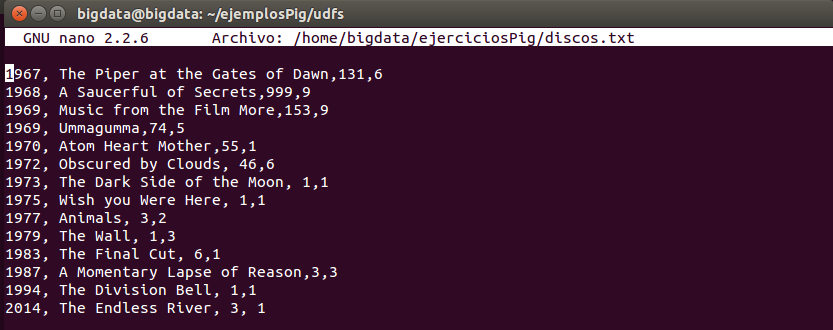
1. Crear un fichero llamado discos.txt



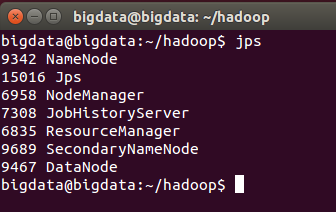
1. Arrancar HDFS, Yarn y el job history

$ cd $HADOOP\_HOME

$ ./sbin/start-dfs.sh

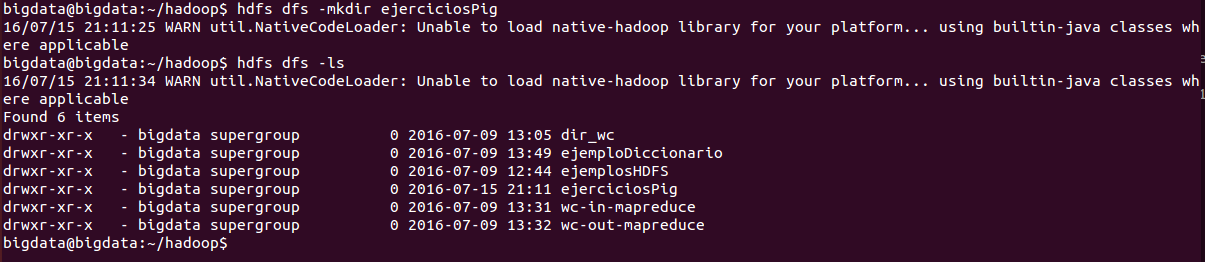
$ ./sbin/start-yarn.sh

$ sbin/mr-jobhistory-daemon.sh start historyserver

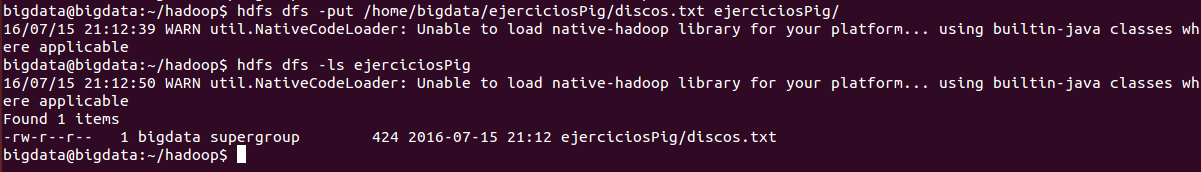


1. Subir el fichero a HDFS dentro de la carpeta /ejerciciosPig/discografia.txt

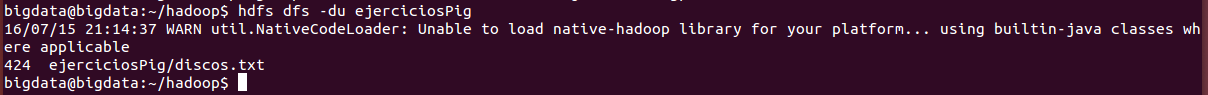
Creamos la carpeta en HDFS



Subimos el fichero

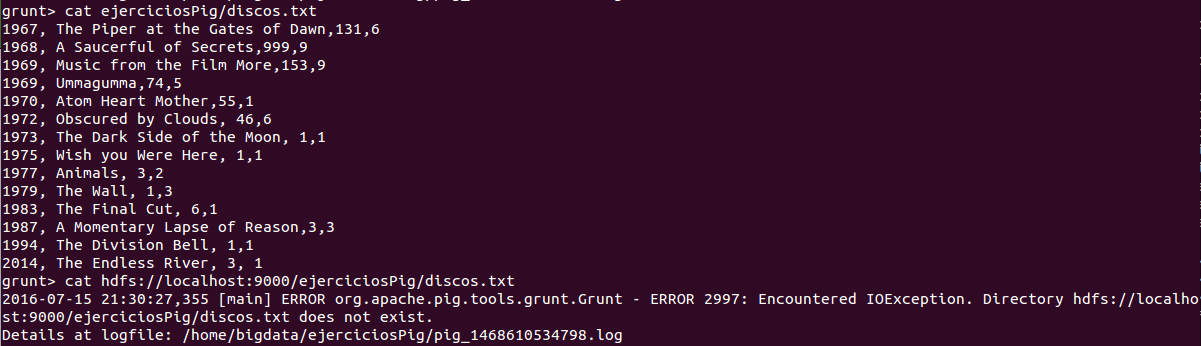


1. Ejecutar la instrucción ls sobre Hadoop para indicar el tamaño del fichero



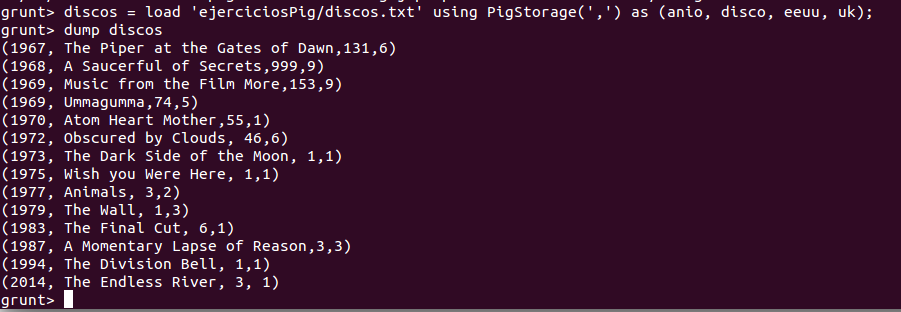
1. Arrancar pig en modo distribuido (si se desea eliminar trazas de log) y ejecutar el siguiente comando: cat hdfs://localhost:9000/ejerciciosPig/discografia.txt para confirmar que los primeros puntos han funcionado correctamente y el fichero está subido a HDFS.



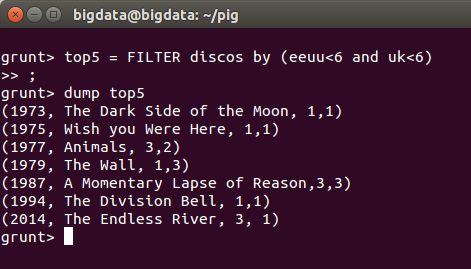


Utilizando el comando localhost:9000 me dice que el fichero no existe. Pero puedo acceder a él con un cat normal.

1. Cargar el fichero de hdfs en una variable llamada discos.

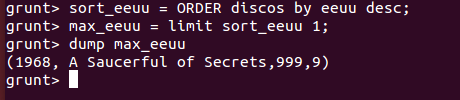


1. Calcular los discos que estuvieron a la vez en el top 5 de EEUU y de UK (indicar también el resultado).



1. Obtener la máxima y mínima posición que ocuparon los discos de Pink Floyd en EEUU y en UK (indicar también el resultado)  empleando los comandos de LATIN PIG.

Máxima posición EEUU:



1. Explica con tus propias palabras lo que se desea obtener con los siguientes comandos e indica el resultado obtenido.

a = foreach discos generate anio;

b = distinct a;

dump b;

1. (opcional) Empleando UDFs extrae información útil de la discografía.