



La finalidad de esta práctica es manejar el sistema de archivos de Hadoop (HDFS) utilizando los comandos **shell** que nos oferta.

Vete haciendo capturas de pantalla de todos los pasos que vayas dando, acompañándolas de comentarios descriptivos de los mismos.

## CONTENIDO

### APARTADO A

## Prácticas con HDFS

1. Muestra la ayuda del comando para manejar el sistema de archivos de HDFS.
2. Muestra el contenido de la raíz de HDFS.
3. Visualiza dicha carpeta raíz desde el gestor de archivos del navegador. ¿En qué puerto se localiza?
4. Crea en HDFS un nuevo directorio llamado **datos** dentro de la carpeta **/user/maria\_dev**.
5. Comprobar que existe.
6. Mostrarlo desde el navegador.
7. Crea un fichero llamado **practicas.txt** en tu directorio **home** de tu usuario en Linux con alguna frase dentro.
8. Copiarlo en HDFS, en concreto al directorio **datos** anterior.
9. Comprueba su existencia desde la utilidad del navegador.
10. Haz clic sobre el fichero ¿Cuál es el tamaño del fichero en HDFS? ¿Cuánto ocupa realmente en HDFS? ¿Cuántas veces está replicado?
11. Visualizar su contenido en HDFS.
12. HDFS es un sistema de archivos distribuido que está dentro de una carpeta local de nuestro Linux. En su archivo de configuración está su ubicación. ¿Podrías localizarla? Explora su contenido desde Linux ¿Qué carpetas tiene dentro? ¿Dónde están los datos?
13. Si damos clic sobre el fichero en el entorno del navegador aparece el **Block Id** del fichero que nos indica el nombre del subdirectorio dentro del sistema de archivos local del punto anterior ¿Podrías mostrar su contenido?
14. Vamos a crear otro ejemplo con un fichero grande. Investiga como crear automáticamente desde Linux con un comando un archivo de 1GB en la carpeta **home** de tu usuario en Linux.
15. Copia el archivo anterior al directorio **/datos** de nuestro HDFS.
16. Comprueba en la página web que ha creado múltiples bloques ¿Cuántos ha creado? ¿De qué tamaño son? Como solamente tenemos un nodo aparecen todos los bloques en el mismo, pero en un clúster real cada bloque estaría en un nodo distinto.
17. ¿Puedes localizar en el sistema de archivos local dichos bloques?
18. Vamos a crear otro directorio llamado **practicas** dentro de la carpeta **/user/maria\_dev**.



19. Copiamos **prueba.txt** desde datos a **practicas**.
20. Comprobamos el contenido de **practicas**.
21. Comprobamos el contenido de **prueba.txt** con un comando de HDFS.
22. Borramos el fichero **prueba.txt**.
23. Borra el directorio **practicas**.