Práctica: Uso de Comandos HDFS para Gestión de Archivos en Hadoop

Objetivo:

Aprender y practicar los comandos básicos del Sistema de Archivos Distribuido de Hadoop (HDFS) para gestionar archivos y directorios.

Requisitos:

- Un entorno Hadoop configurado y operativo.
- Acceso a HDFS a través de la línea de comandos.
- Conocimientos básicos de la línea de comandos de Linux.

Instrucciones:

Realizar pruebas con los siguientes comandos

• Ver versión de hadoop:

hadoop versión

• Verificar el estado del HDFS:

hdfs dfsadmin -report

• Listar directorios y archivos:

hdfs dfs -ls /

Crear un nuevo directorio:

hdfs dfs -mkdir /Nombre usuario

• Transferir una carpeta de tu computadora al servidor:

scp -P 10022 -r C:\Users\Usuario\Downloads\datos_RUOA
Usuario@132.248.172.201:/home/Usuario

- Subir un archivo desde el sistema local a HDFS:
 hdfs dfs -put /Usuario/archivo.txt /Usuario/datos/
- Descargar un archivo desde HDFS al sistema local: hdfs dfs -get /Usuario/datos/archivo.txt /Usuario/
- Mover o renombrar un archivo dentro de HDFS:

hdfs dfs -mv /user/estudiante/datos/archivo.txt /user/estudiante/datos/archivo_renombrado.txt

• Eliminar un archivo en HDFS:

hdfs dfs -rm /ruta/datos/archivo renombrado.txt

- Eliminar un directorio en HDFS:
 hdfs dfs -rm -r /ruta/datos
- Ver el contenido del archivo: hdfs dfs -cat /ruta/archivo.txt
- Mostrar las primeras líneas de un archivo: hdfs dfs -head /ruta/archivo.txt
- Mostrar las últimas líneas de un archivo: hdfs dfs -tail /user/estudiante/archivo.txt
- Copiar un archivo dentro de HDFS:
 hdfs dfs -cp /ruta/origen /ruta/destino
- Crea un solo archivo .csv a partir de varios:

hdfs dfs -appendToFile 2000.csv 2001.csv 2002.csv 2003.csv 2004.csv 2005.csv 2006.csv 2007.csv 2008.csv /user/hadoop/flightsUSA.csv

Tarea: Descarga cada archivo .csv de la RUOA de 2015 hasta finales de 2021: Por minuto

https://www.ruoa.unam.mx/csv_data/more/minuto.php

Sube los archivos a hdfs con los comandos asignados en una carpeta llamada datos_RUOA_minutos que este en tu directorio, checar que se encuentren en el directorio correcto, asi mismo crear un solo archivo csv que contenga todos los archivos descargados, tomar capturas de pantalla de todo el proceso y crear un archivo donde expliques cada paso que realizaste.