Report Template

Name : SWETHA K
Roll Number : 235229143
Class : II M.Sc Data Science
Subject : Big Data Analytics Lab
Date : 16-07-2024

Exercise-1

- Create a directory: /user/yourname/data

hdfs dfs -mkdir swetha

- Upload a file from local: yourfile.txt to /user/yourname/data

hdfs dfs put Hello.txt swetha

- List all files in /user/yourname/data

hdfs dfs ls swetha

- Create a subdirectory: /user/yourname/data/processed

hdfs dfs -mkdir swetha/swethasub

- Move yourfile.txt to /user/yourname/data/processed

hdfs dfs -mv swetha/swefile.txt

- Copy yourfile.txt to /user/yourname/data/backup.txt

swetha/backup.txt

- Change permissions of /user/yourname/data/processed/yourfile.txt to read-only for others.

hdfs dfs chmod o=r sub/swefile.txt

- Delete the backup file: /user/yourname/data/backup.txt

hdfs dfs -rm swetha/backup.txt

- View the content of yourfile.txt in HDFS.

hdfs dfs -cat swetha/swethasub/swefile.txt

- Delete the /user/yourname/data/processed directory.

hdfs dfs -rm -r swetha/swethasub

Exercise-II:

- Create a directory structure: /user/yourname/projects/data and /user/yourname/projects/processed.

hdfs dfs -mkdir -p swetha/projects/data
hdfs dfs -mkdir -p swetha/projects/processed

- Upload multiple files (e.g., file1.txt, file2.txt, file3.txt) from the local filesystem to /user/yourname/projects/data.

touch swe1.txt
touch swe2.txt
touch swe3.txt
hdfs dfs put swe1.txt swe2.txt swe3.txt /projects/data

- List all files in /user/yourname/projects/data and note their sizes and permissions.

hdfs dfs -ls swetha/projects/data

- Move all .txt files from /user/yourname/projects/data to /user/yourname/projects/data/archive.

hdfs dfs -mv swetha/projects/data/*.txt swetha/projects/data/archive

- Create a subdirectory within /user/yourname/projects/data: /user/yourname/projects/data/archive.

hdfs dfs -mkdir swetha/projects/data/archive

- Create a Hadoop Archive (HAR) from the /user/yourname/projects/data/archive directory.

hadoop archive -archiveName data.har -p swetha/projects/data swetha/projects/data/archive

- Change the replication factor of the files in /user/yourname/projects/data/archive to 3.

hdfs dfs -setrep -R 3 swetha/projects/data/archive

- Verify the replication factor of the files in the archive directory.

hdfs dfs -stat %r swetha/projects/data/archive/*

- Use wildcards to delete all .txt files from /user/yourname/projects/data.

hdfs dfs -rm swetha/projects/data/*.txt

- Create a file named summary.txt in /user/yourname/projects/processed, summarizing the contents of the .txt files in the archive directory (e.g., number of records, any other relevant metrics).

echo "Summary of .txt files" > summary.txt
hdfs dfs -put summary.txt swetha/projects/processed

- Copy summary.txt back to /user/yourname/projects/data for backup.

hdfs dfs -cp swetha/projects/processed/summary.txt swetha/projects/data

- Change the permissions of summary.txt to allow only the owner to read and write, while others have no access.

hdfs dfs -chmod 600 swetha/projects/processed/summary.txt

- Implement a recursive listing of all files in /user/yourname/projects to ensure all files are organized.

hdfs dfs -ls -R swetha/projects

- Delete the /user/yourname/projects/data/archive directory and all its contents.

hdfs dfs -rm -r swetha/projects/data/archive

- Create a backup of the entire /user/yourname/projects directory to a new location in HDFS.

hdfs dfs -cp -R swetha/projects swetha/projects_backup