Hadoop File Management Tasks -

This document provides question and answer format explanations for performing file management tasks in Hadoop using the HDFS shell commands. These tasks are intended for graduate students working in the Cloudera environment.

Problem Statement 1:

You are working on a Big Data project and need to organize raw data files in HDFS. Your task is to create appropriate directories and upload required data files into the HDFS file system.

Q1. How do you add files and directories in HDFS?

Answer:

To add files and directories in Hadoop HDFS:

1. Create a directory:

hdfs dfs -mkdir /user/student/data

2. Upload a file to HDFS:

hdfs dfs -put sample.txt /user/student/data/

3. Upload multiple files:

hdfs dfs -put \*.csv /user/student/data/

Q2. How do you retrieve files from HDFS?

Problem Statement 2:

You need to verify the contents of your uploaded files and fetch them for further processing. Perform actions to view and download files from HDFS to your local system.

Answer:

To retrieve files from Hadoop HDFS:

4. 1. List files in a directory:

hdfs dfs -ls /user/student/data

5. 2. View contents of a file:

hdfs dfs -cat /user/student/data/sample.txt

6. 3. Download a file to the local system:

hdfs dfs -get /user/student/data/sample.txt /home/student/

Q3. How do you delete files and directories in HDFS?

Answer:

Problem Statement 3:

After completing your processing tasks, you are required to clean up the HDFS by deleting unnecessary files and directories. Demonstrate how to perform these deletions.

To delete files and directories from Hadoop HDFS:

1. Delete a file:

hdfs dfs -rm /user/student/data/sample.txt

2. Delete a directory and its contents:

hdfs dfs -rm -r /user/student/data

3. Force delete (if needed):

hdfs dfs -rm -r -f /user/student/data

Practice Questions and Answers

Q1. Write the command to create a directory named '/project/logs' in HDFS.

Answer: hdfs dfs -mkdir /project/logs

Q2. How can you upload a file named 'log1.txt' from your desktop to the '/project/logs' directory in HDFS?

Answer: hdfs dfs -put ~/Desktop/log1.txt /project/logs/

Q3. Write the command to read the content of a file '/project/logs/log1.txt' in HDFS.

Answer: hdfs dfs -cat /project/logs/log1.txt

Q4. How do you delete a file named 'temp.csv' located in '/project/data' in HDFS?

Answer: hdfs dfs -rm /project/data/temp.csv

Viva Questions

1. What is the purpose of the Hadoop Distributed File System (HDFS)?

2. How does HDFS handle large data files?

3. What is the difference between -put and -copyFromLocal commands in HDFS?

4. Can you retrieve deleted files from HDFS? Why or why not?

5. What happens when you try to create a directory that already exists in HDFS?

6. How can you check the space usage of a directory in HDFS?

7. What is the role of the NameNode in HDFS file operations?