

Exercise: Working with Text Files in HDFS

Exercise: Working with Text Files in HDFS

Objective:

In this exercise, you will practice basic operations with text files in Hadoop Distributed File System (HDFS). You will perform tasks like uploading, downloading, managing directories, and performing simple manipulations on text files in HDFS.

Note: This exercise assumes you have a Hadoop cluster set up with HDFS configured. If you don't have a cluster, you can use tools like Hadoop's standalone mode or platforms like Hortonworks Sandbox.

Tasks:

1. Create Directories:

- a. Create a directory called "test_files" in HDFS.
- b. Inside the "test_files" directory, create subdirectories for different categories (e.g., "poems", "essays", "stories").

2. Upload Text Files to HDFS:

- a. Prepare a few sample text files on your local machine (e.g., `poem1.txt`, `essay1.txt`, `story1.txt`) with some content.
- b. Upload the sample text files to the appropriate subdirectories in HDFS under "test_files".

3. List and View Files:

- a. List the contents of the "test_files" directory in HDFS to verify that your uploaded files are present.
- b. Use the command to view the contents of one of the uploaded text files without downloading it.

4. Merge Text Files:

- a. Create a new directory called "merged" inside the "test_files" directory.

5. Download Files from HDFS:

Download the merged text file from the "merged" directory in HDFS to your local machine.

6. Copy, Move, and Delete Files:

- a. Create a duplicate copy of one of the uploaded text files within the same subdirectory.
- b. Move the duplicate copy to a different category's subdirectory.
- c. Delete the original copy of the text file from HDFS.

7. Check Disk Usage:

Check the disk usage of the HDFS directories to see the space occupied by your uploaded and merged files.

8. Cleanup:

After completing the exercise, remember to clean up by removing any temporary files or directories created during the exercise.