**Kush Parihar**

**2203023**

**LY-ISA1**

**Assignment No. 4**

Aim: File Management in Hadoop

Objective: The objective of this assignment is to practice basic file operations in the Hadoop Distributed File System (HDFS).

THEORY: File management in Hadoop is facilitated primarily through the Hadoop Distributed File System (HDFS), which is designed to store and manage large files across a distributed cluster of machines. HDFS divides files into smaller blocks and replicates them across nodes to ensure fault tolerance and efficient data access. HDFS is a distributed file system that is designed to handle large datasets by distributing them across numerous nodes in a cluster. It is resilient to node failure, which ensures data reliability. HDFS stores metadata on a dedicated server, known as the Name Node, while actual data is stored on other servers called Data Nodes.

Hadoop's HDFS provides fault tolerance, scalability, and data locality, making it well-suited for handling large-scale data storage and processing tasks in a distributed environment. While HDFS is a foundational component, other distributed storage solutions have also emerged, offering additional features and capabilities beyond HDFS.

## Prerequisites

Hadoop 3.3.6 should be installed.

A test file should be at /Users/komal/Downloads/dev/input/BDA.txt .

## Assignment Steps

### Navigate to the Hadoop Directory

Open your terminal and navigate to the Hadoop directory using the following command:

cd /Users/nikhil/Downloads/dev/hadoop-3.3.6

### Start HDFS

Start the Hadoop Distributed File System (HDFS) by running:

sbin/start-dfs.sh

### Create Directory in HDFS

Create a new directory named

in HDFS using the following command:

**input**

hdfs dfs -mkdir /input

### Upload File to HDFS

Upload the HDFS:

**test.txt**

file from your local system to the

directory in

**input**

hdfs dfs -put /Users/nikhil/Downloads/dev/input/test.txt /input

### List Directory Contents

List the contents of the uploaded:

directory to verify that the file has been

**input**

hdfs dfs -ls /input

### Display File Contents

Display the contents of the uploaded

file:

**test.txt**

hdfs dfs -cat /input/test.txt

### Copy File Within HDFS

Make a copy of directory:

**test.txt**

and name it

within the same

**test\_copy.txt**

hdfs dfs -cp /input/test.txt /input/test\_copy.txt

### Rename File in HDFS

Rename the copied file from

**test\_copy.txt**

to **test\_moved.txt** :

hdfs dfs -mv /input/test\_copy.txt /input/test\_moved.txt

### Delete File from HDFS

Delete the

**test\_moved.txt**

file from HDFS:

hdfs dfs -rm /input/test\_moved.txt

### Stop HDFS

Finally, stop the HDFS service:

sbin/stop-dfs.sh

# Conclusion

In this assignment, we walked through the essential file management operations in Hadoop's Distributed File System (HDFS). We started by navigating to the Hadoop installation directory and initiating the HDFS service. We then performed various operations like creating a directory, uploading a file, listing directory contents,

displaying file content, copying, renaming, and deleting files within HDFS. Finally, we stopped the HDFS service.