**Big Data Frameworks CSE3120**

Lab – 3 Hadoop introduction and HDFS commands

**Name:** Naveen Nidadavolu

**Roll No:** 22MIA1049

**Aim:** To understand and demonstrate the usage of various HDFS commands for file and directory operations in the Hadoop Distributed File System (HDFS).

**Algorithm/Procedure**

1. Start the Hadoop environment and ensure HDFS services are running.
2. Execute each HDFS command as described, to perform operations like directory creation, file uploads, and retrieval.
3. Verify results using appropriate HDFS commands (ls, cat, etc.).

**Preliminary Commands**

**1.start-all.sh**

* **Command:** start-all.sh
* **Usage:** Starts all essential Hadoop daemons, including the HDFS (NameNode and DataNode) and YARN (ResourceManager and NodeManager) services. It ensures that the distributed environment is fully operational.

A computer screen shot of white text

Description automatically generated

**2.jps**

* **Command:** jps
* **Usage:** Displays the list of Java processes running on the system, such as NameNode, DataNode, ResourceManager, and NodeManager. Useful for verifying whether Hadoop daemons are running.A screenshot of a computer

  Description automatically generated

**HDFS Commands and Usage Descriptions**

**1. version**

* **Command:**

hdfs version

* **Usage:** Displays the Hadoop version currently installed and running. It helps verify the installation and compatibility of Hadoop.

A computer screen shot of a computer code

Description automatically generated

**2. mkdir**

* **Command:**

hdfs dfs -mkdir /directory\_name

* **Usage:** Creates a new directory in HDFS. This is useful for organizing files in the distributed file system.



**3. touchz**

* **Command:**

hdfs dfs -touchz /path\_to\_file

* **Usage:** Creates an empty file at the specified path in HDFS. If the file already exists, the command does nothing. Useful for creating placeholder files in HDFS.

A screenshot of a computer

Description automatically generated

**4. ls**

* **Command:**

hdfs dfs -ls /path

* **Usage:** Lists the files and directories at the specified path in HDFS. It provides details like file permissions, owner, group, size, and modification date.

A screenshot of a computer

Description automatically generated

**5. put**

* **Command:**

hdfs dfs -put localfile /path\_in\_hdfs

* **Usage:** Uploads a file from the local filesystem to HDFS. It is used to populate HDFS with data for processing

A computer screen with white text

Description automatically generated

**6. copyFromLocal**

* **Command:**

hdfs dfs -copyFromLocal /local/path /hdfs/path

* **Usage:** Copies a file or directory from the local file system to HDFS. It works similarly to put but supports recursive copying for directories.

A screenshot of a computer

Description automatically generated

**7. get**

* **Command:**

hdfs dfs -get /hdfs/path /local/path

* **Usage:** Downloads files or directories from HDFS to the local filesystem. It is useful for retrieving processed data or backups from HDFS.

A computer screen shot of a program

Description automatically generated

**8. copyToLocal**

* **Command:**

hdfs dfs -copyToLocal /hdfs/path /local/path

* **Usage:** Copies files or directories from HDFS to the local file system, similar to get. It is often used for backups or local analysis.

A computer screen shot of a program

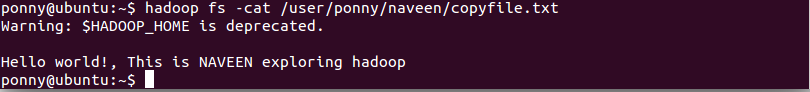
Description automatically generated

**9. cat**

* **Command:**

hdfs dfs -cat /file\_path\_in\_hdfs

* **Usage:** Displays the content of a file stored in HDFS. It is useful for verifying file content without downloading it.



**10. mv**

* **Command:**

hdfs dfs -mv /source\_path /destination\_path

* **Usage:** Moves or renames files or directories within HDFS. This is helpful for organizing or restructuring data in HDFS.

A screenshot of a computer program

Description automatically generated

**11. cp**

* **Command:**

hdfs dfs -cp /source\_path /destination\_path

* **Usage:** Copies files or directories within HDFS. It is useful for creating duplicates of data for backup or other purposes.

A screenshot of a computer program

Description automatically generated