**Ex no: 4 - HDFS – Advanced Commands**

|  |  |
| --- | --- |
| **LIST FILES** | |
| **hdfs dfs -ls /** | List all the files/directories for the given hdfs destination path |
| **hdfs dfs -ls -d /hadoop** | Directories are listed as plain files. In this case, this command will list the details of hadoop folder |
| **hdfs dfs -ls -h /data** | Format file sizes in a human-readable fashion (eg 64.0m instead of 67108864) |
| **hdfs dfs -ls -R /hadoop** | Recursively list all files in hadoop directory and all subdirectories in hadoop directory |
| **hdfs dfs -ls /hadoop/dat\*** | List all the files matching the pattern. In this case, it will list all the files inside hadoop directory which starts with 'dat' |

|  |  |
| --- | --- |
| **READ/WRITE FILES** | |
| **hdfs dfs -text /hadoop/derby.log** | HDFS Command that takes a source file and outputs the file in text format on the terminal. The allowed formats are zip and TextRecordInputStream |
| **hdfs dfs -cat /hadoop/test** | This command will display the content of the HDFS file test on your stdout |
| **hdfs dfs -appendToFile /home/ubuntu/test1/hadoop/text2** | |

|  |  |
| --- | --- |
| **UPLOAD/DOWNLOAD FILES** | |
| **hdfs dfs -put /home/ubuntu/sample /hadoop** | Copies the file from local file system to HDFS |
| **hdfs dfs -put -f /home/ubuntu/sample /hadoop** | Copies the file from local file system to HDFS, and in case the local already exists in the given destination path, using -f option with put command will overwrite it |
| **hdfs dfs -put -l /home/ubuntu/sample /hadoop** | Copies the file from local file system to HDFS. Allow DataNode to lazily persist the file to disk. Forces replication factor of 1 |
| **hdfs dfs -put -p /home/ubuntu/sample /hadoop** | Copies the file from local file system to HDFS. Passing -p preserves access and modification times, ownership and the mode |
| **hdfs dfs -get /newfile /home/ubuntu/** | Copies the file from HDFS to local file system |
| **hdfs dfs -get -p /newfile /home/ubuntu/** | Copies the file from HDFS to local file system. Passing -p preserves access and modification times, ownership and the mode |
| **hdfs dfs -get /hadoop/\*.txt /home/ubuntu/** | Copies all the files matching the pattern from local file system to HDFS |
| **hdfs dfs -copyFromLocal /home/ubuntu/sample /hadoop** | Works similarly to the put command, except that the source is restricted to a local file reference |
| **hdfs dfs -copyToLocal /newfile /home/ubuntu/** | Works similarly to the put command, except that the destination is restricted to a local file reference |
| **hdfs dfs -moveFromLocal /home/ubuntu/sample /hadoop** | Works similarly to the put command, except that the source is deleted after it's copied |

|  |  |
| --- | --- |
| **FILE MANAGEMENT** | |
| **hdfs dfs -cp /hadoop/file1 /hadoop1** | Copies file from source to destination on HDFS. In this case, copying file1 from hadoop directory to hadoop1 directory |
| **hdfs dfs -cp -p /hadoop/file1 /hadoop1** | Copies file from source to destination on HDFS. Passing -p preserves access and modification times, ownership and the mode |

|  |  |
| --- | --- |
| **hdfs dfs -cp -f /hadoop/file1 /hadoop1** | Copies file from source to destination onHDFS. Passing -f overwrites the destination when if it already exists |
| **hdfs dfs -mv /hadoop/file1 /hadoop1** | Move files that match the specified file pattern <src> to a destination <dst>.  When moving multiple files, the destination must be a directory |
| **hdfs dfs -rm /hadoop/file1** | Deletes the file (sends it to the trash) |
| **hdfs dfs -rm -r /hadoop**  **hdfs dfs -rm -R /hadoop**  **hdfs dfs -rmr /hadoop** | Deletes the directory and any content under it recursively |
| **hdfs dfs -rm -skipTrash /hadoop** | The -skipTrash option will bypass trash, if enabled, and delete the specified file(s) immediately |
| **hdfs dfs -rm -f /hadoop** | If the file does not exist, do not display a diagnostic message or modify the exit status to reflect an error |
| **hdfs dfs -rmdir /hadoop1** | Delete a directory |
| **hdfs dfs -mkdir /hadoop2** | Create a directory in specified HDFS location |
| **hdfs dfs -mkdir -f /hadoop2** | Create a directory in specified HDFS location. This command does not fail even if the directory already exists |
| **hdfs dfs -touchz /hadoop3** | Creates a file of zero length at <path> with current time as the timestamp of that <path> |

|  |  |
| --- | --- |
| **OWNERSHIP AND VALIDATION** | |
| **hdfs dfs -checksum /hadoop/file1** | Dump checksum information for files that match the file pattern <src> to stdout |
| **hdfs dfs -chmod 755 /hadoop/file1** | Changes permissions of the file |
| **hdfs dfs -chmod -R 755 /hadoop** | Changes permissions of the files recursively |
| **hdfs dfs -chown ubuntu:ubuntu /hadoop** | Changes owner of the file. 1st ubuntu in the command is owner and 2nd one is group |
| **hdfs dfs -chown -R ubuntu:ubuntu /hadoop** | Changes owner of the files recursively |
| **hdfs dfs -chgrp ubuntu /hadoop** | Changes group association of the file |
| **hdfs dfs -chgrp -R ubuntu /hadoop** | Changes group association of the files recursively |

|  |  |
| --- | --- |
| **FILESYSTEM** | |
| **hdfs dfs -df /hadoop** | Shows the capacity, free and used space of the filesystem |
| **hdfs dfs -df -h /hadoop** | Shows the capacity, free and used space of the filesystem. -h parameter Formats the sizes of files in a human-readable fashion |
| **hdfs dfs -du /hadoop/file** | Show the amount of space, in bytes, used by the files that match the specified file pattern |
| **hdfs dfs -du -s /hadoop/file** | Rather than showing the size of each individual file that matches the pattern, shows the total (summary) size |
| **hdfs dfs -du -h /hadoop/file** | Show the amount of space, in bytes, used by the files that match the  specified file pattern. Formats the sizes of files in a human-readable fashion |

|  |  |
| --- | --- |
| **ADMINISTRATION** | |
| **hdfs balancer -threshold 30** | Runs a cluster balancing utility. Percentage of disk capacity. This overwrites the default threshold |
| **hadoop version** | To check the version of Hadoop |
| **hdfs fsck /** | It checks the health of the Hadoop file system |
| **hdfs dfsadmin -safemode leave** | The command to turn off the safemode of NameNode |
| **hdfs dfsadmin -refreshNodes** | Re-read the hosts and exclude files to update the set of Datanodes that are allowed to connect to the Namenode and those that should be decommissioned or recommissioned |
| **hdfs namenode -format** | Formats the NameNode |