1. Hadoop Namenode Commands

|  |  |
| --- | --- |
| **Command** | **Description** |
| hadoop namenode -format | Format HDFS filesystem from Namenode |
| hadoop namenode -upgrade | Upgrade the NameNode |
| start-dfs.sh | Start HDFS Daemons |
| stop-dfs.sh | Stop HDFS Daemons |
| start-mapred.sh | Start MapReduce Daemons |
| stop-mapred.sh | Stop MapReduce Daemons |
| hadoop namenode -recover -force | Recover namenode metadata after a cluster failure (may lose data) |

2. Hadoop fsck Commands

|  |  |
| --- | --- |
| **Command** | **Description** |
| hadoop fsck / | Filesystem check on HDFS |
| hadoop fsck / -files | Display files during check |
| hadoop fsck / -files -blocks | Display files and blocks during check |
| hadoop fsck / -files -blocks -locations | Display files, blocks and its location during check |
| hadoop fsck / -files -blocks -locations -racks | Display network topology for data-node locations |
| hadoop fsck -delete | Delete corrupted files |
| hadoop fsck -move | Move corrupted files to /lost+found directory |

3. Hadoop Job Commands

|  |  |
| --- | --- |
| **Command** | **Description** |
| hadoop job -submit <job-file> | Submit the job |
| hadoop job -status <job-id> | Print job status completion percentage |
| hadoop job -list all | List all jobs |
| hadoop job -list-active-trackers | List all available TaskTrackers |
| hadoop job -set-priority <job-id> <priority> | Set priority for a job. Valid priorities: VERY\_HIGH, HIGH, NORMAL, LOW, VERY\_LOW |
| hadoop job -kill-task <task-id> | Kill a task |
| hadoop job -history | Display job history including job details, failed and killed jobs |

4. Hadoop dfsadmin Commands

|  |  |
| --- | --- |
| **Command** | **Description** |
| hadoop dfsadmin -report | Report filesystem info and statistics |
| hadoop dfsadmin -metasave file.txt | Save namenode’s primary data structures to file.txt |
| hadoop dfsadmin -setQuota 10 /quotatest | Set Hadoop directory quota to only 10 files |
| hadoop dfsadmin -clrQuota /quotatest | Clear Hadoop directory quota |
| hadoop dfsadmin -refreshNodes | Read hosts and exclude files to update datanodes that are allowed to connect to namenode. Mostly used to commission or decommsion nodes |
| hadoop fs -count -q /mydir | Check quota space on directory /mydir |
| hadoop dfsadmin -setSpaceQuota /mydir 100M | Set quota to 100M on hdfs directory named /mydir |
| hadoop dfsadmin -clrSpaceQuota /mydir | Clear quota on a HDFS directory |
| hadooop dfsadmin -saveNameSpace | Backup Metadata (fsimage & edits). Put cluster in safe mode before this command. |

5. Hadoop Safe Mode (Maintenance Mode) Commands

The following dfsadmin commands helps the cluster to enter or leave safe mode, which is also called as maintenance mode. In this mode, Namenode does not accept any changes to the name space, it does not replicate or delete blocks.

|  |  |
| --- | --- |
| **Command** | **Description** |
| hadoop dfsadmin -safemode enter | Enter safe mode |
| hadoop dfsadmin -safemode leave | Leave safe mode |
| hadoop dfsadmin -safemode get | Get the status of mode |
| hadoop dfsadmin -safemode wait | Wait until HDFS finishes data block replication |

6. Hadoop Configuration Files

|  |  |
| --- | --- |
| **File** | **Description** |
| hadoop-env.sh | Sets ENV variables for Hadoop |
| core-site.xml | Parameters for entire Hadoop cluster |
| hdfs-site.xml | Parameters for HDFS and its clients |
| mapred-site.xml | Parameters for MapReduce and its clients |
| masters | Host machines for secondary Namenode |
| slaves | List of slave hosts |

7. Hadoop mradmin Commands

|  |  |
| --- | --- |
| **Command** | **Description** |
| hadoop mradmin -safemode get | Check Job tracker status |
| hadoop mradmin -refreshQueues | Reload mapreduce configuration |
| hadoop mradmin -refreshNodes | Reload active TaskTrackers |
| hadoop mradmin -refreshServiceAcl | Force Jobtracker to reload service ACL |
| hadoop mradmin -refreshUserToGroupsMappings | Force jobtracker to reload user group mappings |

8. Hadoop Balancer Commands

|  |  |
| --- | --- |
| **Command** | **Description** |
| start-balancer.sh | Balance the cluster |
| hadoop dfsadmin -setBalancerBandwidth <bandwidthinbytes> | Adjust bandwidth used by the balancer |
| hadoop balancer -threshold 20 | Limit balancing to only 20% resources in the cluster |

9. Hadoop Filesystem Commands

|  |  |
| --- | --- |
| **Command** | **Description** |
| hadoop fs -mkdir mydir | Create a directory (mydir) in HDFS |
| hadoop fs -ls | List files and directories in HDFS |
| hadoop fs -cat myfile | View a file content |
| hadoop fs -du | Check disk space usage in HDFS |
| hadoop fs -expunge | Empty trash on HDFS |
| hadoop fs -chgrp hadoop file1 | Change group membership of a file |
| hadoop fs -chown huser file1 | Change file ownership |
| hadoop fs -rm file1 | Delete a file in HDFS |
| hadoop fs -touchz file2 | Create an empty file |
| hadoop fs -stat file1 | Check the status of a file |
| hadoop fs -test -e file1 | Check if file exists on HDFS |
| hadoop fs -test -z file1 | Check if file is empty on HDFS |
| hadoop fs -test -d file1 | Check if file1 is a directory on HDFS |

10. Additional Hadoop Filesystem Commands

|  |  |
| --- | --- |
| **Command** | **Description** |
| hadoop fs -copyFromLocal <source> <destination> | Copy from local fileystem to HDFS |
| hadoop fs -copyFromLocal file1 data | e.g: Copies file1 from local FS to data dir in HDFS |
| hadoop fs -copyToLocal <source> <destination> | copy from hdfs to local filesystem |
| hadoop fs -copyToLocal data/file1 /var/tmp | e.g: Copies file1 from HDFS data directory to /var/tmp on local FS |
| hadoop fs -put <source> <destination> | Copy from remote location to HDFS |
| hadoop fs -get <source> <destination> | Copy from HDFS to remote directory |
| hadoop distcp hdfs://192.168.0.8:8020/input hdfs://192.168.0.8:8020/output | Copy data from one cluster to another using the cluster URL |
| hadoop fs -mv file:///data/datafile /user/hduser/data | Move data file from the local directory to HDFS |
| hadoop fs -setrep -w 3 file1 | Set the replication factor for file1 to 3 |
| hadoop fs -getmerge mydir bigfile | Merge files in mydir directory and download it as one big file |