Apache Hadoop Top 20 Commands

Apache Hadoop Commands

In this tutorial, we will see the list of Hadoop commands which are used to perform multiple tasks like submit jobs, check the status of jobs, kill jobs, check the list of jobs, command to format namenode, command to rollback namenode in its previous state and so on.

Commands	Description
$hadoop job -submit	This command is used to submit the job.
$hadoop job -status	This command is used to print the map and reduce the end percentage and all job counters.
$hadoop job -counter	This command will print group-name counter-name and counter value.
$hadoop job -kill job-id	This command will kill the job.
$hadoop job -events job-id from-event-# #-of-events.	This command is used to imprint the event details received by the job tracker for the delivered scope.
$hadoop job -history [all]jobOutputDir	This command is used to print job details such as the job is completed, failed, or killed. The option (all) will give more details such as successful tasks and the attempts composed for each task.
$hadoop job -list [all]	This command will display jobs that are yet to complete.
$hadoop job -kill-task task-id	This command will kill the task.
$hadoop job -fail-task task-id	This command will fail the task.

$hadoop job -set-priority job-id priority	This command will change the priority of the job. The values allowed for priority are very high, high, normal, low, and very low.
$hadoop dfsadmin -report	This command is used to describe the filesystem.
$hadoop dfsadmin -refreshNodes	This command will 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.
$hadoop dfsadmin -finalizeUpgrade	This command will finalize the upgrade of HDFS.
$hadoop dfsadmin -upgradeProgress status / details / force	This command will request the current distributed upgrade status, a detailed status, or force the upgrade to proceed.
$hadoop dfsadmin -metasave filename	This command saves the Namenode's fundamental data structures to the filename in the directory named by hadoop.log.dir property.
$hadoop dfsadmin -restoreFailedStorage true / false / check	This command is used to turn on/off planned effort to recover broken storage replicas.
$hadoop namenode -format	This command will format the namenode.
$hadoop namenode -rollback	This command is used to revert the namenode to the earlier version. This command is used when the Hadoop cluster is stopped.
$hadoop namenode -finalize	This command will remove the previous state of the files system. The recent upgrade will become permanent. The rollback option will not be available anymore. Once the finalization is completed post that, the namenode is shutdown.
$hadoop namenode -importCheckpoint	This command will load an image from a checkpoint directory and save it into the current one. Checkpoint dir is read from property fs. checkpoint. dir.