

# Commands Guide

## Table of contents

1 Overview.....	2
1.1 Generic Options.....	2
2 User Commands .....	3
2.1 archive .....	3
2.2 distcp .....	3
2.3 fs .....	3
2.4 fsck .....	3
2.5 jar .....	4
2.6 job .....	4
2.7 pipes .....	5
2.8 queue .....	6
2.9 version .....	6
2.10 CLASSNAME .....	6
2.11 classpath .....	7
3 Administration Commands .....	7
3.1 balancer .....	7
3.2 daemonlog .....	7
3.3 datanode.....	8
3.4 dfsadmin .....	8
3.5 mradmin.....	9
3.6 jobtracker .....	10
3.7 namenode .....	10
3.8 secondarynamenode .....	10
3.9 tasktracker .....	11

## 1. Overview

All hadoop commands are invoked by the bin/hadoop script. Running the hadoop script without any arguments prints the description for all commands.

```
Usage: hadoop [--config confdir] [COMMAND] [GENERIC_OPTIONS]
[COMMAND_OPTIONS]
```

Hadoop has an option parsing framework that employs parsing generic options as well as running classes.

COMMAND_OPTION	Description
--config confdir	Overwrites the default Configuration directory. Default is \${HADOOP_HOME}/conf.
GENERIC_OPTIONS	The common set of options supported by multiple commands.
COMMAND COMMAND_OPTIONS	Various commands with their options are described in the following sections. The commands have been grouped into <a href="#">User Commands</a> and <a href="#">Administration Commands</a> .

### 1.1. Generic Options

The following options are supported by [dfsadmin](#), [fs](#), [fsck](#) and [job](#). Applications should implement [Tool](#) to support [GenericOptions](#).

GENERIC_OPTION	Description
-conf <configuration file>	Specify an application configuration file.
-D <property=value>	Use value for given property.
-fs <local namenode:port>	Specify a namenode.
-jt <local jobtracker:port>	Specify a job tracker. Applies only to <a href="#">job</a> .
-files <comma separated list of files>	Specify comma separated files to be copied to the map reduce cluster. Applies only to <a href="#">job</a> .
-libjars <comma seperated list of jars>	Specify comma separated jar files to include in the classpath. Applies only to <a href="#">job</a> .
-archives <comma separated list of archives>	Specify comma separated archives to be unarchived on the compute machines. Applies only to <a href="#">job</a> .

## 2. User Commands

Commands useful for users of a hadoop cluster.

### 2.1. archive

Creates a hadoop archive. More information can be found at [Hadoop Archives](#).

Usage: `hadoop archive -archiveName NAME <src>* <dest>`

COMMAND_OPTION	Description
<code>-archiveName NAME</code>	Name of the archive to be created.
<code>src</code>	Filesystem pathnames which work as usual with regular expressions.
<code>dest</code>	Destination directory which would contain the archive.

### 2.2. distcp

Copy file or directories recursively. More information can be found at [Hadoop DistCp Guide](#).

Usage: `hadoop distcp <srcurl> <desturl>`

COMMAND_OPTION	Description
<code>srcurl</code>	Source Url
<code>desturl</code>	Destination Url

### 2.3. fs

Usage: `hadoop fs [GENERIC OPTIONS] [COMMAND_OPTIONS]`

Runs a generic filesystem user client.

The various COMMAND\_OPTIONS can be found at [File System Shell Guide](#).

### 2.4. fsck

Runs a HDFS filesystem checking utility. See [Fsck](#) for more info.

Usage: `hadoop fsck [GENERIC OPTIONS] <path> [-move | -delete | -openforwrite] [-files [-blocks [-locations | -racks]]]`

COMMAND_OPTION	Description
<path>	Start checking from this path.
-move	Move corrupted files to /lost+found
-delete	Delete corrupted files.
-openforwrite	Print out files opened for write.
-files	Print out files being checked.
-blocks	Print out block report.
-locations	Print out locations for every block.
-racks	Print out network topology for data-node locations.

## 2.5. jar

Runs a jar file. Users can bundle their Map Reduce code in a jar file and execute it using this command.

Usage: `hadoop jar <jar> [mainClass] args...`

The streaming jobs are run via this command. Examples can be referred from [Streaming examples](#)

Word count example is also run using jar command. It can be referred from [Wordcount example](#)

## 2.6. job

Command to interact with Map Reduce Jobs.

Usage: `hadoop job [GENERIC OPTIONS] [-submit <job-file>] | [-status <job-id>] | [-counter <job-id> <group-name> <counter-name>] | [-kill <job-id>] | [-events <job-id> <from-event-#> <#-of-events>] | [-history [all] <jobOutputDir>] | [-list [all]] | [-kill-task <task-id>] | [-fail-task <task-id>] | [-set-priority <job-id> <priority>]`

COMMAND_OPTION	Description
-submit <job-file>	Submits the job.
-status <job-id>	Prints the map and reduce completion

	percentage and all job counters.
<code>-counter &lt;job-id&gt; &lt;group-name&gt; &lt;counter-name&gt;</code>	Prints the counter value.
<code>-kill &lt;job-id&gt;</code>	Kills the job.
<code>-events &lt;job-id&gt; &lt;from-event-#&gt; &lt;#-of-events&gt;</code>	Prints the events' details received by jobtracker for the given range.
<code>-history [all] &lt;jobOutputDir&gt;</code>	<code>-history &lt;jobOutputDir&gt;</code> prints job details, failed and killed tip details. More details about the job such as successful tasks and task attempts made for each task can be viewed by specifying the [all] option.
<code>-list [all]</code>	<code>-list all</code> displays all jobs. <code>-list</code> displays only jobs which are yet to complete.
<code>-kill-task &lt;task-id&gt;</code>	Kills the task. Killed tasks are NOT counted against failed attempts.
<code>-fail-task &lt;task-id&gt;</code>	Fails the task. Failed tasks are counted against failed attempts.
<code>-set-priority &lt;job-id&gt; &lt;priority&gt;</code>	Changes the priority of the job. Allowed priority values are VERY_HIGH, HIGH, NORMAL, LOW, VERY_LOW

## 2.7. pipes

Runs a pipes job.

Usage: `hadoop pipes [-conf <path>] [-jobconf <key=value>, <key=value>, ...] [-input <path>] [-output <path>] [-jar <jar file>] [-inputformat <class>] [-map <class>] [-partitioner <class>] [-reduce <class>] [-writer <class>] [-program <executable>] [-reduces <num>]`

COMMAND_OPTION	Description
<code>-conf &lt;path&gt;</code>	Configuration for job
<code>-jobconf &lt;key=value&gt;, &lt;key=value&gt;, ...</code>	Add/override configuration for job
<code>-input &lt;path&gt;</code>	Input directory
<code>-output &lt;path&gt;</code>	Output directory

-jar <jar file>	Jar filename
-inputformat <class>	InputFormat class
-map <class>	Java Map class
-partitioner <class>	Java Partitioner
-reduce <class>	Java Reduce class
-writer <class>	Java RecordWriter
-program <executable>	Executable URI
-reduces <num>	Number of reduces

## 2.8. queue

command to interact and view Job Queue information

Usage : `hadoop queue [-list] | [-info <job-queue-name> [-showJobs]] | [-showacls]`

COMMAND_OPTION	Description
-list	Gets list of Job Queues configured in the system. Along with scheduling information associated with the job queues.
-info <job-queue-name> [-showJobs]	Displays the job queue information and associated scheduling information of particular job queue. If -showJobs options is present a list of jobs submitted to the particular job queue is displayed.
-showacls	Displays the queue name and associated queue operations allowed for the current user. The list consists of only those queues to which the user has access.

## 2.9. version

Prints the version.

Usage: `hadoop version`

## 2.10. CLASSNAME

hadoop script can be used to invoke any class.

Usage: `hadoop CLASSNAME`

Runs the class named CLASSNAME.

## 2.11. classpath

Prints the class path needed to get the Hadoop jar and the required libraries.

Usage: `hadoop classpath`

## 3. Administration Commands

Commands useful for administrators of a hadoop cluster.

### 3.1. balancer

Runs a cluster balancing utility. An administrator can simply press Ctrl-C to stop the rebalancing process. See [Rebalancer](#) for more details.

Usage: `hadoop balancer [-threshold <threshold>]`

COMMAND_OPTION	Description
<code>-threshold &lt;threshold&gt;</code>	Percentage of disk capacity. This overwrites the default threshold.

### 3.2. daemonlog

Get/Set the log level for each daemon.

Usage: `hadoop daemonlog -getlevel <host:port> <name>`

Usage: `hadoop daemonlog -setlevel <host:port> <name> <level>`

COMMAND_OPTION	Description
<code>-getlevel &lt;host:port&gt; &lt;name&gt;</code>	Prints the log level of the daemon running at <host:port>. This command internally connects to <code>http://&lt;host:port&gt;/logLevel?log=&lt;name&gt;</code>
<code>-setlevel &lt;host:port&gt; &lt;name&gt; &lt;level&gt;</code>	Sets the log level of the daemon running at <host:port>. This command internally connects to <code>http://&lt;host:port&gt;/logLevel?log=&lt;name&gt;</code>

### 3.3. datanode

Runs a HDFS datanode.

Usage: `hadoop datanode [-rollback]`

COMMAND_OPTION	Description
<code>-rollback</code>	Rollsback the datanode to the previous version. This should be used after stopping the datanode and distributing the old hadoop version.

### 3.4. dfsadmin

Runs a HDFS dfsadmin client.

Usage: `hadoop dfsadmin [GENERIC OPTIONS] [-report] [-safemode enter | leave | get | wait] [-refreshNodes] [-finalizeUpgrade] [-upgradeProgress status | details | force] [-metasave filename] [-setQuota <quota> <dirname>...<dirname>] [-clrQuota <dirname>...<dirname>] [-help [cmd]]`

COMMAND_OPTION	Description
<code>-report</code>	Reports basic filesystem information and statistics.
<code>-safemode enter   leave   get   wait</code>	Safe mode maintenance command. Safe mode is a Namenode state in which it 1. does not accept changes to the name space (read-only) 2. does not replicate or delete blocks. Safe mode is entered automatically at Namenode startup, and leaves safe mode automatically when the configured minimum percentage of blocks satisfies the minimum replication condition. Safe mode can also be entered manually, but then it can only be turned off manually as well.
<code>-refreshNodes</code>	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.
<code>-finalizeUpgrade</code>	Finalize upgrade of HDFS. Datanodes delete their previous version working directories,



	followed by Namenode doing the same. This completes the upgrade process.
<code>-upgradeProgress status   details   force</code>	Request current distributed upgrade status, a detailed status or force the upgrade to proceed.
<code>-metasave filename</code>	Save Namenode's primary data structures to <filename> in the directory specified by <code>hadoop.log.dir</code> property. <filename> will contain one line for each of the following <ol style="list-style-type: none"> <li>1. Datanodes heart beating with Namenode</li> <li>2. Blocks waiting to be replicated</li> <li>3. Blocks currently being replicated</li> <li>4. Blocks waiting to be deleted</li> </ol>
<code>-setQuota &lt;quota&gt; &lt;dirname&gt;...&lt;dirname&gt;</code>	Set the quota <quota> for each directory <dirname>. The directory quota is a long integer that puts a hard limit on the number of names in the directory tree. Best effort for the directory, with faults reported if <ol style="list-style-type: none"> <li>1. N is not a positive integer, or</li> <li>2. user is not an administrator, or</li> <li>3. the directory does not exist or is a file, or</li> <li>4. the directory would immediately exceed the new quota.</li> </ol>
<code>-clrQuota &lt;dirname&gt;...&lt;dirname&gt;</code>	Clear the quota for each directory <dirname>. Best effort for the directory. with fault reported if <ol style="list-style-type: none"> <li>1. the directory does not exist or is a file, or</li> <li>2. user is not an administrator.</li> </ol> It does not fault if the directory has no quota.
<code>-help [cmd]</code>	Displays help for the given command or all commands if none is specified.

### 3.5. mradmin

Runs MR admin client

Usage: `hadoop mradmin [ GENERIC OPTIONS ] [-refreshQueueAcls]`

COMMAND_OPTION	Description
<code>-refreshQueueAcls</code>	Refresh the queue acls used by hadoop, to check access during submissions and administration of the job by the user. The properties present in <code>mapred-queue-acls.xml</code> is reloaded by the

	queue manager.
--	----------------

### 3.6. jobtracker

Runs the MapReduce job Tracker node.

Usage: `hadoop jobtracker [-dumpConfiguration]`

COMMAND_OPTION	Description
<code>-dumpConfiguration</code>	Dumps the configuration used by the JobTracker alongwith queue configuration in JSON format into Standard output used by the jobtracker and exits.

### 3.7. namenode

Runs the namenode. More info about the upgrade, rollback and finalize is at [Upgrade](#)  
[Rollback](#)

Usage: `hadoop namenode [-format] | [-upgrade] | [-rollback] | [-finalize] | [-importCheckpoint]`

COMMAND_OPTION	Description
<code>-format</code>	Formats the namenode. It starts the namenode, formats it and then shut it down.
<code>-upgrade</code>	Namenode should be started with upgrade option after the distribution of new hadoop version.
<code>-rollback</code>	Rollsback the namenode to the previous version. This should be used after stopping the cluster and distributing the old hadoop version.
<code>-finalize</code>	Finalize will remove the previous state of the files system. Recent upgrade will become permanent. Rollback option will not be available anymore. After finalization it shuts the namenode down.
<code>-importCheckpoint</code>	Loads image from a checkpoint directory and save it into the current one. Checkpoint dir is read from property <code>fs.checkpoint.dir</code>

### 3.8. secondarynamenode

Runs the HDFS secondary namenode. See [Secondary Namenode](#) for more info.

Usage: `hadoop secondarynamenode [-checkpoint [force]] | [-geteditsize]`

COMMAND_OPTION	Description
<code>-checkpoint [force]</code>	Checkpoints the Secondary namenode if EditLog size $\geq$ <code>fs.checkpoint.size</code> . If <code>-force</code> is used, checkpoint irrespective of EditLog size.
<code>-geteditsize</code>	Prints the EditLog size.

### 3.9. tasktracker

Runs a MapReduce task Tracker node.

Usage: `hadoop tasktracker`