hadoop job

The `hadoop job` command enables you to manage MapReduce jobs.

Syntax

```
hadoop job [Generic Options]
        [-submit <job-file>]
 [-status <job-id>]
 [-counter <job-id> <group-name> <counter-name>]
 [-kill <job-id>]
 [-unblacklist <job-id> <hostname>]
 [-unblacklist-tracker <hostname>]
 [-set-priority <job-id> <priority>]
 [-events <job-id> <from-event-#> <#-of-events>]
 [-history <jobOutputDir>]
 [-list [all]]
 [-list-active-trackers]
 [-list-blacklisted-trackers]
 [-list-attempt-ids <job-id> <task-type> <task-state>]
 [-kill-task <task-id>]
 [-fail-task <task-id>]
        [-blacklist-tasktracker <hostname>]
        [-showlabels]
```

Parameters

Command Options

The following command options are supported for `hadoop job`:

Parameter	Description
`-submit <job-file>`	Submits the job.
`-status <job-id>`	Prints the map and reduce completion percentage and all job counters.
`-counter <job-id> <group-name> <counter-name>`	Prints the counter value.
`-kill <job-id>`	Kills the job.
`-unblacklist <job-id> <hostname>`	Removes a tasktracker job from the jobtracker's blacklist.
`-unblacklist-tracker <hostname>`	Admin only. Removes the TaskTracker at `<hostname>` from the JobTracker's global blacklist.
`-set-priority <job-id> <priority>`	Changes the priority of the job. Valid priority values are `VERY_HIGH`, `HIGH, NORMAL`, `LOW`, and `VERY_LOW`. The job scheduler uses this property to determine the order in which jobs are run.
`-events <job-id> <from-event-#> <#-of-events>`	Prints the events' details received by jobtracker for the given range.
`-history <jobOutputDir>`	Prints job details, failed and killed tip details.
`-list [all]`	The `-list all` option displays all jobs. The `-list` command without the `all` option displays only jobs which are yet to complete.
`-list-active-trackers`	Prints all active tasktrackers.
`-list-blackisted-trackers`	Prints the TaskTracker nodes that JobTracker blacklisted with the reason for blacklisting.
`-list-attempt-ids <job-id><task-type>`	Lists the IDs of task attempts.
`-kill-task <task-id>`	Kills the task. Killed tasks are *not* counted against failed attempts.

`-fail-task <task-id>`	Fails the task. Failed tasks are counted against failed attempts.
`-blacklist-tasktracker <hostname>`	Pauses all current tasktracker jobs and prevent additional jobs from being scheduled on the tasktracker.
`-showlabels`	Dumps label information of all active nodes.

Generic Options

The following generic options are supported for the `hadoop job` command: `-conf <configuration file>`, `-D <property=value>`, `-fs <local|file system URI>`, `-jt <local|jobtracker:port>`, `-files <file1,file2,file3,...>`, `-libjars <libjar1,libjar2,libjar3,...>`, and `-archives <archive1,archive2,archive3,...>`. For more information on generic options, see Generic Options.

Examples

Submitting Jobs

The `hadoop job -submit` command enables you to submit a job to the specified jobtracker.

```
$ hadoop job -jt darwin:50020 -submit job.xml
```

Stopping Jobs Gracefully

Use the `hadoop kill` command to stop a running or queued job.

```
$ hadoop job -kill <job-id>
```

Viewing Job History Logs

Run the `hadoop job -history` command to view the history logs summary in specified directory.

```
$ hadoop job -history output-dir
```

This command will print job details, failed and killed tip details.

Additional details about the job such as successful tasks and task attempts made for each task can be viewed by adding the `-all` option:

```
$ hadoop job -history all output-dir
```

Blacklisting Tasktrackers

The `hadoop job` command when run as root or using `sudo` can be used to manually blacklist tasktrackers:

```
hadoop job -blacklist-tasktracker <hostname>
```

Manually blacklisting a tasktracker pauses any running jobs and prevents additional jobs from being scheduled.
For a detailed discussion see TaskTracker Blacklisting.