

Hadoop file system commands

A table of all HDFS operations is reproduced below. The following conventions are used for parameters:

- italics denote variables to be filled out by the user
- “path” means any file or directory name
- “path...” means one or more file or directory names
- “file” means any filename.
- “src” and “dest” are path names in a directed operation.
- “localSrc” and “localDest” are paths as above, but on the local file system. All other file and path names refer to objects inside HDFS
- parameters in [brackets] are optional

Command: <code>hadoop fs ...</code>	Description
<code>-ls path</code>	Lists the contents of the directory specified by <i>path</i> , showing the names, permissions, owner, size and modification date for each entry.
<code>-ls -R path</code>	Behaves like <code>-ls</code> , but recursively displays entries in all subdirectories of <i>path</i> .
<code>-du path</code>	Shows disk usage, in bytes, for all files which match <i>path</i> ; filenames are reported with the full HDFS protocol prefix.
<code>-du -s path</code>	Like <code>-du</code> , but prints a summary of disk usage of all files/directories in the path.
<code>-mv src dest</code>	Moves the file or directory indicated by <i>src</i> to <i>dest</i> , within HDFS.
<code>-cp src dest</code>	Copies the file or directory identified by <i>src</i> to <i>dest</i> , within HDFS.
<code>-rm path</code>	Removes the file or empty directory identified by <i>path</i> .

<code>-rm -r <i>path</i></code>	Removes the file or directory identified by <i>path</i> . Recursively deletes any child entries (i.e., files or subdirectories of <i>path</i>).
<code>-put <i>localSrc dest</i></code>	Copies the file or directory from the local file system identified by <i>localSrc</i> to <i>dest</i> within the DFS.
<code>-copyFromLocal <i>localSrc dest</i></code>	Identical to <code>-put</code>
<code>-moveFromLocal <i>localSrc dest</i></code>	Copies the file or directory from the local file system identified by <i>localSrc</i> to <i>dest</i> within HDFS, then deletes the local copy on success.
<code>-get [-crc] <i>src localDest</i></code>	Copies the file or directory in HDFS identified by <i>src</i> to the local file system path identified by <i>localDest</i> .
<code>-getmerge <i>src localDest</i> [addnl]</code>	Retrieves all files that match the path <i>src</i> in HDFS, and copies them to a single, merged file in the local file system identified by <i>localDest</i> .
<code>-cat <i>filename</i></code>	Displays the contents of <i>filename</i> on stdout.
<code>-copyToLocal [-crc] <i>src localDest</i></code>	Identical to <code>-get</code>
<code>-moveToLocal [-crc] <i>src localDest</i></code>	Works like <code>-get</code> , but deletes the HDFS copy on success.
<code>-mkdir <i>path</i></code>	Creates a directory named <i>path</i> in HDFS. Creates any parent directories in <i>path</i> that are missing (e.g., like <code>mkdir -p</code> in Linux).
<code>-setrep [-R] [-w] <i>rep path</i></code>	Sets the target replication factor for files identified by <i>path</i> to <i>rep</i> . (The actual replication factor will move toward the target over time)
<code>-touchz <i>path</i></code>	Creates a file at <i>path</i> containing the current time as a timestamp. Fails if a file already exists at <i>path</i> , unless the file is already size 0.
<code>-test [-ezd] <i>path</i></code>	Returns 1 if <i>path</i> exists; has zero length; or is a directory, or 0 otherwise.
<code>-stat [format] <i>path</i></code>	Prints information about <i>path</i> . <i>format</i> is a string which accepts file size in blocks (%b), filename (%n), block size (%o), replication (%r), and modification date (%y, %Y).

-tail [-f] <i>file</i>	Shows the last 1KB of file on stdout.
-chmod [-R] <i>mode path...</i>	Changes the file permissions associated with one or more objects identified by <i>path...</i> . Performs changes recursively with -R. <i>mode</i> is a 3-digit octal mode, or {augo}+/-{rwxX}. Assumes a if no scope is specified and does not apply a umask.
-chown [-R] [<i>owner</i>] [: [<i>group</i>]] <i>path...</i>	Sets the owning user and/or group for files or directories identified by <i>path...</i> . Sets owner recursively if -R is specified.
-chgrp [-R] <i>group path...</i>	Sets the owning group for files or directories identified by <i>path...</i> . Sets group recursively if -R is specified.
-help <i>cmd</i>	Returns usage information for any of the commands listed above. You must omit the leading '-' character in <i>cmd</i> .

This material is taken from the “Hadoop Tutorial from Yahoo!” by Yahoo! Inc., available from <http://developer.yahoo.com/hadoop/tutorial/>.