

HDFS Cheat Sheet by Matthew Rathbone

Commonly used commands for HDFS. This is not an exhaustive list, but the most useful ones on a day to day basis.

Listing Files

This is how you inspect HDFS to see what it contains. Use these commands to find files and their associated metadata.

<code>hadoop fs -ls ./example</code>	List files in a directory
<code>hadoop fs -ls ./example/matthew*</code>	List files matching a pattern
<code>hadoop fs -ls -h ./example/</code>	List files with human-friendly file sizes (eg 1.3M vs 1331325)
<code>hadoop fs -ls -R ./example/</code>	Recursively list files in this directory AND all child directories.
<code>hadoop fs -ls -d ./example/</code>	List files, but with directories shown as files. So in this case it will show info about the example folder itself.

Uploading/Downloading Files

Moving files between HDFS and the local filesystem and back.

<code>hadoop fs -put ./localfile.txt ./example/</code>	Upload a file from your local machine to a specific directory on HDFS.
<code>hadoop fs -put -f ./localfile.txt ./example/</code>	Upload a file and overwrite any existing file on HDFS.
<code>hadoop fs -put -l ./localfile.txt ./example/</code>	Upload a file and set a replication factor of 1 (you probably shouldn't ever really use this).
<code>hadoop fs -get ./example/remotefile.txt ./</code>	Download a file from HDFS to your local machine.
<code>hadoop fs -get -p ./example/remotefile.txt ./</code>	Download a file from HDFS to your local machine, preserving metadata (eg modified time).
<code>hadoop fs -get ./example/*.txt ./</code>	Download a set of files that match a pattern to your local machine.

Reading & Writing Files

Reading file contents without downloading the file itself.

<code>hadoop fs -text ./example/file.txt</code>	Print the contents of a file to the terminal, decompressing if necessary.
<code>hadoop fs -cat ./example/*.txt</code>	Print to the terminal the contents of all files that match the provided pattern. Note - this will NOT decompress like 'text' will.
<code>hadoop fs [-cat,-text] -ignoreCrc ./example/*.txt</code>	As above, but disable the verification checksum.
<code>hadoop fs -appendToFile ./localfile.txt ./example/remotefile.txt</code>	Append the contents of a local file to a file on HDFS. <u>This is only supported in Hadoop versions 2.1.1+</u>

File Management

Organize your files with these commands.

<code>hadoop fs -mv ./example/f1.txt ./example/f2.txt</code>	Move a file to a different file/directory (omit the filename to name it the same)
<code>hadoop fs -cp ./example/f1.txt ./example/f2.txt</code>	Copy a file to a different file/directory (omit the filename to name it the same)
<code>hadoop fs -rm ./example/f1.txt</code>	Delete a file (sends it to the trash)
<code>hadoop fs -rm -skipTrash ./example/f1.txt</code>	Actually delete the file. No trash.
<code>hadoop fs -rm -r ./example/directory</code>	Recursively delete a directory and its contents
<code>hadoop fs -touchz ./example/somefile</code>	Create a zero-length file (great for creating _SUCCESS files).
<code>hadoop fs -touchz ./example/somefile</code>	Create a zero-length file (great for creating _SUCCESS files).

HDFS Administration

Check on important stuff that is less about your files and more about HDFS.

<code>hadoop fs -df -h ./example</code>	Show capacity and used space of the filesystem. Will show partition space remaining if you have partitions.
<code>hadoop fs -du -h ./example/*.txt</code>	show the amount of space used by matching files
<code>hadoop fs -expunge</code>	Empty the Trash (useful if you -rm without -skipTrash)
<code>hadoop fs -chown owner:group ./example</code>	Change ownership of a file (use -R for the directory)
<code>hadoop fs -chmod 0700 ./example/file.txt</code>	Change the mode of the file (eg to 0700)
<code>hadoop fs -checksum ./example/*.txt</code>	Fetch checksum information for the matching files (requires a datanode roundtrip, slow and intensive).