

Hadoop Cheat Sheet

hiregion.com

User Commands	
archive	Create a hadoop archive
d stcp	Make a distributed copy
fs	fs command – see file system
fsck	HDFS file system check
jar	Run a jar file
job	Interact with MapReduce jobs
p pes	Run a pipe job
queue	Interact with job queue
version	Print hadoop version
CLASSNAME	Run java class

Admin Commands	
balancer	Balance data in cluster
daemonlog	Get/set daemon log level
datanode	Rollback datanode (-rollback)
dfsadmin	Administer hdfs
jobtracker	Run a MapReduce job Tracker
namenode	Run a namenode
secondarynamenode	Runs HDFS secondary namenode
tasktracker	Run a MapReduce task Tracker

Hadoop streaming	
\$HADOOP_HOME/bin/hadoop jar \	
\$HADOOP_HOME/hadoop-streaming.jar \	
-input myInputDir \	
-output myOutputDir \	
-mapper myPythonScript.py \	
-reducer /bin/wc \	
-file myPythonScript.py	

fs – file system	
cat	Concatenate – source to stdout
chgrp	Change group association
chmod	Change permission on files
chown	Change ownership of files
copyFromLocal	Copy from local system to hdfs
copyToLocal	Copy to local system
count	Get Count of dirs, files, bytes
cp	Copy files from URI to UR
du	Get file size
dus	Get aggregate size
expurge	Delete from trash
get	Similar to copyToLocal
getmerge	Merge files
ls	List files
lsr	Recursive ls
mkcir	Create a directory
moveFromLocal	Move file from local to hdfs
moveToLocal	Move file from hdfs to local
mv	Move from URI to URI
put	Similar to copyFromLocal
rm	Remove file, dir
rmr	Recursive remove
setrep	Change replication factor
stat	Get stat info
tail	Get tail content of a file
test	Tests a file
text	Get text formatted output
touchz	Create a zero length file