Hadoop commands

List the information in hdfs, or for certain folder
hadoop fs –ls
hadoop fs –ls myinput

put "local" document on to hdfs
hadoop fs –put ~/Hortonworks.1 /user/guest/Hortonworks

see the tail of certain file
hadoop fs –tail purchases.txt

Rename a file
hadoop fs –mv purchases.txt newname.txt

Remove a file
hadoop fs –rm newname.txt

Make a folder
hadoop fs –mkdir myinput

Put some file into some folder
hadoop fs –put purchases.txt myinput

Use cat to see a short part of information
hadoop fs –cat joboutput/part-00000 | less

Run map reduce
hs mapper.py reducer.py myinput joboutput

Copy a short part of a file into a testfile
head -50 ../data/purchases.txt > testfile

Cat while doing map and reduce not using hs (need to sort first)
cat testfile | ./mapper.py | sort | ./reducer.py

Remove a file
rm -rf /var/log/httpd/access
it will remove the whole directories and files within /var/log/httpd/access

rm -f /var/log/httpd/access.log
it will be forced to remove file /var/log/httpd/access.log

For sandbox:
Get some data online
wget http://en.wikipedia.org/wiki/Hortonworks

Copy the data over to hdfs on sandbox
hadoop fs –put ~/Hortonworks.1 /user/guest/Hortonworks

Open pyspark
pyspark

Solve write permission error to hdfs
[root@sandbox ~]# su hdfs
[hdfs@sandbox root]$ hadoop fs -chmod -R 777 /user/guest
[hdfs@sandbox root]$ exit
[rooßt@sandbox ~]# hadoop fs -put ~/Hortonworks /user/guest/Hortonworks
[root@sandbox ~]# hadoop fs -ls /user/guest/