

Last login: Sun Nov 15 09:01:22 on ttys000
(base) lejahu@a-10-27-12-70 ~ % ssh -Y lh2514@dumbo.es.its.nyu.edu

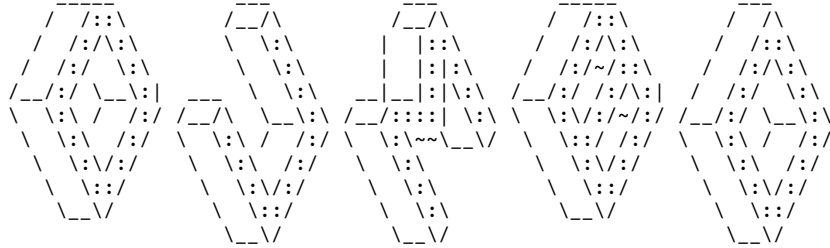
Password:

Last login: Sun Nov 15 08:34:25 2020 from a-10-27-12-70.dynapool.vpn.nyu.edu
Rocks 6.2 (SideWinder)
Profile built 16:54 24-Aug-2015

Kickstarted 16:58 24-Aug-2015

~~~~~  
THIS NYU-OPERATED FACILITY MAY BE ACCESSED ONLY BY AUTHORIZED USERS.  
UNAUTHORIZED PERSONS ... TYPE 'exit' NOW  
See <https://wikis.nyu.edu/display/NYUHPC/NYU+HPC+Access+Policy> for details  
of the access policy. Questions regarding the access policy should be  
directed (by e-mail) to [security@nyu.edu](mailto:security@nyu.edu) or (by phone) to 212-998-3333  
~~~~~

WELCOME TO



<https://wikis.nyu.edu/display/NYUHPC/Clusters+-+Dumbo>

```
[lh2514@login-2-1 ~]$ ls
Airbnb.class          NYC_Mapper.py
Airbnb.jar            NYC_output1
Airbnb.java           NYC_Reducer.py
AirbnbMapper.class    NYPD_Complaint_Data_Historic.csv
AirbnbMapper.java     output_11
AirbnbReducer.class   output_hw3.txt
AirbnbReducer.java    part-r-00000
Cleaned.csv           project
CleanMapper.py        SelectFields.jar
CleanReducer.py       SelectFieldsMapper.class
com                  SelectFieldsMapper.java
data.csv              SelectFieldsReducer.class
hw1                   SelectFieldsReducer.java
input.csv             SelectFieldsRunner.class
input_hw3.txt         SelectFieldsRunner.java
IntSumReducer.class   smallWeather1.txt
IntSumReducer.java    WordCount.class
LineCountMapper.class WordCount.jar
LineCountMapper.java  WordCount.java
LinesCount.jar        WordCountMapper.class
LinesCountRunner.class WordCountMapper.class.save
LinesCountRunner.java WordCountMapper.java
Main.class            WordCountReducer.class
Main.java             WordCountReducer.java
myDir                 Word_output
myNewDir
```

```
[lh2514@login-2-1 ~]$ hdfs dfs -put Cleaned.csv /user/lh2514/class1
```

```
[lh2514@login-2-1 ~]$ hdfs dfs -ls /user/lh2514/class1
```

Found 20 items

Permissions	Owner	Group	Size	Mod Time	Path
drwxr-xr-x+	lh2514	users	0	2020-10-12 01:27	/user/lh2514/class1/Airbnb_output
drwxr-xr-x+	lh2514	users	0	2020-10-12 01:35	/user/lh2514/class1/Airbnb_output1
-rw-r--r--+	3 lh2514	users	52955815	2020-11-15 09:16	/user/lh2514/class1/Cleaned.csv
-rw-r--r--+	3 lh2514	users	1324	2020-11-12 00:53	/user/lh2514/class1/NYC_Mapper.py
-rw-r--r--+	3 lh2514	users	1037	2020-11-12 00:52	/user/lh2514/class1/NYC_Reducer.py
drwxr-xr-x+	lh2514	users	0	2020-11-15 08:43	/user/lh2514/class1/NYC_output1
-rw-r--r--+	3 lh2514	users	265727269	2020-11-12 00:47	/user/lh2514/class1/NYPD_Complaint_Data_Historic.csv
-rw-r--r--+	3 lh2514	users	165	2020-11-14 14:29	/user/lh2514/class1/SelectFieldsMapper.java
-rw-r--r--+	3 lh2514	users	168	2020-11-14 14:28	/user/lh2514/class1/SelectFieldsReducer.java
-rw-r--r--+	3 lh2514	users	1851	2020-11-14 14:28	/user/lh2514/class1/SelectFieldsRunner.java
drwxr-xr-x+	lh2514	users	0	2020-09-22 01:03	/user/lh2514/class1/Word_output
drwxr-xr-x+	lh2514	users	0	2020-09-23 14:21	/user/lh2514/class1/Word_output1
-rw-r--r--+	3 lh2514	users	7077974	2020-10-12 01:26	/user/lh2514/class1/data.csv

```

-rw-r--r--+ 3 lh2514 users 264678694 2020-11-05 01:14 /user/lh2514/class1/input.csv
-rw-rwxr--+ 3 lh2514 users 314 2020-09-12 16:15 /user/lh2514/class1/input_hw2.txt
-rw-r--r--+ 3 lh2514 users 289 2020-09-19 18:44 /user/lh2514/class1/input_hw3.txt
drwxr-xr-x+ - lh2514 users 0 2020-11-14 14:51 /user/lh2514/class1/output_NYC
drwxr-xr-x+ - lh2514 users 0 2020-09-14 13:58 /user/lh2514/class1/output_hw2
drwxr-xr-x+ - lh2514 users 0 2020-11-12 01:45 /user/lh2514/class1/output_nyc
drwxr-xr-x+ - lh2514 users 0 2020-11-12 01:08 /user/lh2514/class1/python_code
[lh2514@login-2-1 ~]$ hadoop jar LineCount.jar LineCountRunner /user/lh2514/class1/Cleaned.csv /user/lh2514/class1/NYC_
output2
Not a valid JAR: /home/lh2514/LineCount.jar
[lh2514@login-2-1 ~]$ javac -classpath "$(yarn classpath)" -d . LineCountMapper.java
[lh2514@login-2-1 ~]$ javac -classpath "$(yarn classpath)" -d . IntSumReducer.java
[lh2514@login-2-1 ~]$ avac -classpath "$(yarn classpath)":. -d . LineCountRunner.java
-bash: avac: command not found
[lh2514@login-2-1 ~]$ javac -classpath "$(yarn classpath)":. -d . LineCountRunner.java
javac: file not found: LineCountRunner.java
Usage: javac <options> <source files>
use -help for a list of possible options
[lh2514@login-2-1 ~]$ javac -classpath "$(yarn classpath)":. -d . LinesCountRunner.java
[lh2514@login-2-1 ~]$ hadoop jar LinesCount.jar LinesCountRunner /user/lh2514/class1/Cleaned.csv /user/lh2514/class1/NY
C_output2
20/11/15 09:20:37 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the T
ool interface and execute your application with ToolRunner to remedy this.
20/11/15 09:20:38 INFO input.FileInputFormat: Total input paths to process : 1
20/11/15 09:20:38 INFO mapreduce.JobSubmitter: number of splits:1
20/11/15 09:20:38 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1604346392376_2183
20/11/15 09:20:39 INFO impl.YarnClientImpl: Submitted application application_1604346392376_2183
20/11/15 09:20:39 INFO mapreduce.Job: The url to track the job: http://babar.es.its.nyu.edu:8088/proxy/application_1604
346392376_2183/
20/11/15 09:20:39 INFO mapreduce.Job: Running job: job_1604346392376_2183
20/11/15 09:20:43 INFO mapreduce.Job: Job job_1604346392376_2183 running in uber mode : false
20/11/15 09:20:43 INFO mapreduce.Job: map 0% reduce 0%
20/11/15 09:20:48 INFO mapreduce.Job: map 100% reduce 0%
20/11/15 09:20:54 INFO mapreduce.Job: map 100% reduce 97%
20/11/15 09:20:55 INFO mapreduce.Job: map 100% reduce 100%
20/11/15 09:21:00 INFO mapreduce.Job: Job job_1604346392376_2183 completed successfully
20/11/15 09:21:00 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=897900
    FILE: Number of bytes written=6883146
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=52955923
    HDFS: Number of bytes written=20
    HDFS: Number of read operations=99
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=64
  Job Counters
    Launched map tasks=1
    Launched reduce tasks=32
    Data-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=11944
    Total time spent by all reduces in occupied slots (ms)=425928
    Total time spent by all map tasks (ms)=2986
    Total time spent by all reduce tasks (ms)=70988
    Total vcore-milliseconds taken by all map tasks=2986
    Total vcore-milliseconds taken by all reduce tasks=70988
    Total megabyte-milliseconds taken by all map tasks=12230656
    Total megabyte-milliseconds taken by all reduce tasks=436150272
  Map-Reduce Framework
    Map input records=1048576
    Map output records=1048576
    Map output bytes=16777216
    Map output materialized bytes=897772
    Input split bytes=108
    Combine input records=0
    Combine output records=0
    Reduce input groups=1
    Reduce shuffle bytes=897772
    Reduce input records=1048576
    Reduce output records=1
    Spilled Records=2097152
    Shuffled Maps =32
    Failed Shuffles=0
    Merged Map outputs=32
    GC time elapsed (ms)=1682
    CPU time spent (ms)=36150

```

Physical memory (bytes) snapshot=12624015360
Virtual memory (bytes) snapshot=123441782784
Total committed heap usage (bytes)=39564345344

Shuffle Errors

BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0

File Input Format Counters
Bytes Read=52955815

File Output Format Counters
Bytes Written=20

[lh2514@login-2-1 ~]\$ hdfs dfs -ls /user/lh2514/class1/NYC_output2

Found 33 items

-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/_SUCCESS
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00000
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00001
-rw-r--r--	3	lh2514	users	20	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00002
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00003
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00004
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00005
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00006
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00007
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00008
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00009
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00010
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00011
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00012
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00013
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00014
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00015
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00016
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00017
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00018
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00019
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00020
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00021
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00022
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00023
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00024
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00025
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00026
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00027
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00028
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00029
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00030
-rw-r--r--	3	lh2514	users	0	2020-11-15	09:20	/user/lh2514/class1/NYC_output2/part-r-00031

[lh2514@login-2-1 ~]\$ hdfs dfs -cat /user/lh2514/class1/NYC_output2/part-r-00002

Total Lines 1048576

[lh2514@login-2-1 ~]\$