Goel, Ajay: Nu Id (001897443)

INFO 7250
Engineering Big-Data Systems
Summer Full 2019
Assignment 3

PART 2

db.nyse.createIndex({stock_symbol:1})

```
> show collections
nyse
> db.nyse.createIndex({stock_symbol:1})
{
        "createdCollectionAutomatically" : false,
        "numIndexesBefore" : 1,
        "numIndexesAfter" : 2,
        "ok" : 1
}
```

db.nyse.getIndexes()

```
> db.nyse.getIndexes()
[
        {
                "v" : 2,
                "key" : {
                        "_id" : 1
                },
                "name" : "_id_",
                "ns" : "bigdata_Ass3.nyse"
        },
        {
                "v" : 2,
                "key" : {
                        "stock_symbol" : 1
                },
                "name" : "stock_symbol_1",
                "ns" : "bigdata_Ass3.nyse"
        }
]
```

PART 3

Created NYSE database first in collection named as nyse3

```
[>
[>
[> db.nyse3.count()
 9211031
 >
```

Db.nyse3.createIndex({stock_symbol:1})

```
{
        "_id" : ObjectId("5cf964cf22ddcc18c7f083bf"),
        "exchange" : "NYSE",
        "stock_symbol" : "AEA",
        "date" : "2010-01-07",
        "stock_price_open" : 5.47,
        "stock_price_high" : 5.65,
        "stock_price_low" : 5.4,
        "stock_price_close" : 5.62,
        "stock_volume" : 228900,
        "stock_price_adj_close" : 5.62
}
Type "it" for more
> db.nyse3.createIndex({stock_symbol:1})
{
        "createdCollectionAutomatically" : false,
        "numIndexesBefore" : 1,
        "numIndexesAfter" : 2,
        "ok" : 1
}
>
```

PART 4

All hadoop commands are invoked by the bin/hadoop script. Running the hadoop script without any arguments prints the description for all commands.

Usage: hadoop [--config confdir] [--loglevel loglevel] [COMMAND] [GENERIC_OPTIONS] [COMMAND_OPTIONS]

Execute each hadoop command once, and place the screenshots into a word file. If a command cannot be executed for any reason (such as, a distributed environment is needed), you may write the definition of the command, and skip execution.

http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-common/FileSystemShell.html

a. Ls
hadoop fs -ls /
list all the files present in the directory

```
Ajays-MacBook-Pro:neu ajaygoel$ hadoop fs -ls /
2019-06-06 16:58:46,287 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... us
ing builtin-java classes where applicable
[Found 2 items
drwxr-xr-x   - ajaygoel supergroup          0 2019-05-31 16:56 /test
drwxr-xr-x   - ajaygoel supergroup          0 2019-05-31 16:42 /test2
```

b. Cat
Copies source paths to stdout

hadoop fs -cat /test/mybook.txt

```
Ajays-MacBook-Pro:neu ajaygoel$ hadoop fs -cat /test/mybook.txt
2019-06-06 17:04:03,797 WARN util.NativeCodeLoader: Unable to load nativ
The Project Gutenberg EBook of Change Signals, by Ralph Henry Barbour

This eBook is for the use of anyone anywhere in the United States and mo
other parts of the world at no cost and with almost no restrictions
whatsoever.  You may copy it, give it away or re-use it under the terms
the Project Gutenberg License included with this eBook or online at
www.gutenberg.org.  If you are not located in the United States, you'll
to check the laws of the country where you are located before using this

Title: Change Signals
       A Story of the New Football

Author: Ralph Henry Barbour

Illustrator: Charles M. Relyea

Release Date: May 28, 2019 [EBook #59623]

Language: English

Character set encoding: UTF-8

*** START OF THIS PROJECT GUTENBERG EBOOK CHANGE SIGNALS ***
```

c. appendToFile

hadoop fs -appendToFile
/Users/ajaygoel/Documents/neu/BigData/Assignments/Assignment3/test.txt
/test/mybook.txt

```
This web site includes information about Project Gutenberg-tm,
including how to make donations to the Project Gutenberg Liter
Archive Foundation, how to help produce our new eBooks, and ho
subscribe to our email newsletter to hear about new eBooks.

Hi! My name is Ajay GoelAjays-MacBook-Pro:neu ajaygoel$ █
```

d. mkdir
   Hadoop fs -mkdir -p /ajay

```
2019-06-06 17:33:19,307 WARN util.NativeCodeLoader: Unable to load nati
Found 3 items
drwxr-xr-x   - ajaygoel supergroup          0 2019-06-06 17:33 /ajay
drwxr-xr-x   - ajaygoel supergroup          0 2019-06-06 17:21 /test
drwxr-xr-x   - ajaygoel supergroup          0 2019-05-31 16:42 /test2
Ajays-MacBook-Pro:neu ajaygoel$ █
```
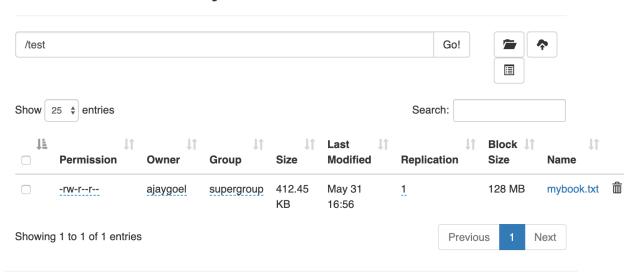
e. copyFromLocal
   copies local file from local to Hadoop directory

   hadoop fs -copyfromLocal /usr/Documents/neu/test.txt /test/mybook.txt

Browse Directory

| /test | | | | | | Go! | |

Show 25 ⟡ entries   Search: [   ]

	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
☐	-rw-r--r--	ajaygoel	supergroup	412.45 KB	May 31 16:56	1	128 MB	mybook.txt 🗑

Showing 1 to 1 of 1 entries   Previous **1** Next

f. head

hadoop fs -head /test/mybook.txt

```
 drwxr-xr-x    - ajaygoel supergroup         0 2019-05-31 16:42 /test2
[Ajays-MacBook-Pro:neu ajaygoel$ hadoop fs -head /test/mybook.txt
 2019-06-06 17:47:21,420 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
 The Project Gutenberg EBook of Change Signals, by Ralph Henry Barbour

 This eBook is for the use of anyone anywhere in the United States and most
 other parts of the world at no cost and with almost no restrictions
 whatsoever.  You may copy it, give it away or re-use it under the terms of
 the Project Gutenberg License included with this eBook or online at
 www.gutenberg.org.  If you are not located in the United States, you'll have
 to check the laws of the country where you are located before using this ebook.
```

g. tail

hadoop fs -tail /test/mybook.txt

```
[Each IlAjays-MacBook-Pro:neu ajaygoel$ hadoop fs -tail /test/mybook.txt
 2019-06-06 17:49:00,933 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
 gutenberg.org/donate

 Section 5. General Information About Project Gutenberg-tm electronic works.

 Professor Michael S. Hart was the originator of the Project
 Gutenberg-tm concept of a library of electronic works that could be
 freely shared with anyone. For forty years, he produced and
 distributed Project Gutenberg-tm eBooks with only a loose network of
 volunteer support.

 Project Gutenberg-tm eBooks are often created from several printed
 editions, all of which are confirmed as not protected by copyright in
 the U.S. unless a copyright notice is included. Thus, we do not
 necessarily keep eBooks in compliance with any particular paper
 edition.

 Most people start at our Web site which has the main PG search
 facility: www.gutenberg.org

 This Web site includes information about Project Gutenberg-tm
```

h. df
Displays free space.
Hadoop fd -df -h

```
[Ajays-MacBook-Pro:neu ajaygoel$ hadoop fs -df -h
 2019-06-06 18:00:25,324 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
 Filesystem               Size   Used   Available  Use%
 hdfs://localhost:9000  465.6 G  440 K    361.7 G    0%
```

i. checksum

Returns the checksum information of a file.
Hadoop fs -checksum /test/mybook.txt

```
.ncck3um.   /tc3t . ro u uirectory
Ajays-MacBook-Pro:neu ajaygoel$ hadoop fs -checksum /test/mybook.txt
2019-06-06 18:07:21,219 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... usin
/test/mybook.txt        MD5-of-0MD5-of-512CRC32C        00000200000000000000000f45996b61f5190b528265200c6bbb413
Ajays-MacBook-Pro:neu ajaygoel$
```

j. chmod

Change the permissions of files. With -R, make the change recursively through the directory structure. The user must be the owner of the file, or else a super-user

hadoop fs -chmod 544 /test/mybook.txt

```
2019-06-06 18:14:10,084 WARN util.NativeCodeLoader: Unable to load native-hadoop library
Found 2 items
-rw-r--r--   1 ajaygoel supergroup         48 2019-06-06 18:05 /test/myBook.txt
-r-xr--r--   1 ajaygoel supergroup     422374 2019-06-06 18:05 /test/mybook.txt
Ajays-MacBook-Pro:neu ajaygoel$
```

k. chown
Change the owner of files. The user must be a super-user.
Hadoop fs -chown 777 /test/myBook.txt

```
Found 2 items
-rw-r--r--   1 777      supergroup         48 2019-06-06 18:05 /test/myBook.txt
-r-xr--r--   1 ajaygoel supergroup     422374 2019-06-06 18:05 /test/mybook.txt
iaus MacBook Pro:neu ajaygoel$ hadoop fs -chown 444 /test/myBook txt
```

l. copyToLocal

hadoop fs -copyToLocal /test/mybook.txt /Users/ajaygoel/Documents

```
-rw-r--r--    1 ajaygoel  staff    8581682 Feb 28 12:35 goel_ajay_001897443
drwxr-xr-x@  11 ajaygoel  staff        352 Jun  3 17:07 highres
-rw-r--r--    1 ajaygoel  staff     422374 Jun  6 18:25 mybook.txt
drwxr-xr-x  1011 ajaygoel  staff      32352 Feb 27 14:58 pics_xs 27022019
```

m. count
return count of the number of files present in the directory
Hadoop fs -count /test

```
Usage: hadoop fs [generic options] -count [-q] [-h] [-v] [-t [<storage ty
Ajays-MacBook-Pro:Documents ajaygoel$ hadoop fs -count /test
2019-06-06 18:30:12,696 WARN util.NativeCodeLoader: Unable to load native
           1            2            422422 /test
Ajays-MacBook-Pro:Documents ajaygoel$
```

n. cp
hadoop fs -cp /test/sample.txt /ajay

```
Ajays-MacBook-Pro:Documents ajaygoel$ hadoop fs -cp /test/sample.txt /ajay
2019-06-06 18:38:53,540 WARN util.NativeCodeLoader: Unable to load native-hadoop li
Ajays-MacBook-Pro:Documents ajaygoel$ hadoop fs -ls /ajay
2019-06-06 18:39:07,454 WARN util.NativeCodeLoader: Unable to load native-hadoop li
Found 1 items
-rw-r--r--   1 ajaygoel supergroup       6120 2019-06-06 18:38 /ajay/sample.txt
Ajays-MacBook-Pro:Documents ajaygoel$
```

o. cw

Usage: hadoop fs -du [-s] [-h] [-x] URI [URI ...]

Displays sizes of files and directories contained in the given directory or the length of a file in case its just a file.

```
jays-MacBook-Pro:Documents ajaygoel$
jays-MacBook-Pro:Documents ajaygoel$ hadoop fs -du -h /test
019-06-06 19:03:27,867 WARN util.NativeCodeLoader: Unable to
8       48        /test/myBook.txt
12.5 K  412.5 K   /test/mybook.txt
.0 K     6.0 K    /test/sample.txt
jays-MacBook-Pro:Documents ajaygoel$
```

p. dus

Displays a summary of file lengths

```
dus: DEPRECATED: Please use 'du -s' instead.
dus: `.': No such file or directory
[Ajays-MacBook-Pro:Documents ajaygoel$ hadoop fs -du -s /ajay
 2019-06-06 19:06:00,530 WARN util.NativeCodeLoader: Unable to I
 6120  6120  /ajay
 Ajays-MacBook-Pro:Documents ajaygoel$
```

q. expunge

Usage: hadoop fs -expunge

Permanently delete files in checkpoints older than the retention threshold from trash directory, and create new checkpoint.

When checkpoint is created, recently deleted files in trash are moved under the checkpoint. Files in checkpoints older than fs.trash.interval will be permanently deleted on the next invocation of -expunge command.

r. Find

Finds all files that match the specified expression and applies selected actions to them. If no *path* is specified then defaults to the current working directory. If no expression is specified then defaults to -print.

hadoop fs -find /test -print

```
[Ajays-MacBook-Pro:Documents ajaygoel$ hadoop fs -find /test -print
 2019-06-06 19:10:15,059 WARN util.NativeCodeLoader: Unable to load na
 /test
 /test/myBook.txt
 /test/mybook.txt
 /test/sample.txt
 Ajays-MacBook-Pro:Documents ajaygoel$
```

s. Get

Copy files to the local file system. Files that fail the CRC check may be copied with the -ignorecrc option. Files and CRCs may be copied using the -crc option

hadoop fs -get /test/mybook.txt
/Users/ajaygoel/Documents/neu/BigData/Assignments/Assignment3

```
/test/sample.txt
[Ajays-MacBook-Pro:Documents ajaygoel$ hadoop fs -get /test/mybook.txt /Users/ajaygoel/Documents/neu/BigData/Assignments/Assignment3
2019-06-06 19:12:29,745 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[Ajays-MacBook-Pro:Documents ajaygoel$ cd /Users/ajaygoel/Documents/neu/BigData/Assignments/Assignment3
[Ajays-MacBook-Pro:Assignment3 ajaygoel$ ls
Assignment3.docx          Questions_Assignment3.docx     nyse.sh               sample.txt          ~$signment3.docx
NYSE                      mybook.txt                     nyse_zip.zip          test.txt
Ajays-MacBook-Pro:Assignment3 ajaygoel$ 
```

t. Getfattr

Displays the extended attribute names and values (if any) for a file or directory

```
Ajays-MacBook-Pro:Assignment3 ajaygoel$ hadoop fs -getfattr -d /test/mybook.txt
2019-06-06 19:15:26,458 WARN util.NativeCodeLoader: Unable to load native-hadoop
# file: /test/mybook.txt
Ajays-MacBook-Pro:Assignment3 ajaygoel$ 
```

u. Getmerge
hadoop fs -moveFromLocal
/Users/ajaygoel/Documents/Neu/BigData/Assignments/Assignment3/sample4.txt
/gutenberg/sample4.txt

hadoop fs -getmerge -nl /gutenberg/sample3.txt /gutenberg/sample4.txt
/Users/ajaygoel/Documents/Neu/BigData/Assignments/Assignment3/test2

```
 79  SMALL LETTER Y          F9  SMALL LETTER U WITH GRAVE
 7A  SMALL LETTER Z          FA  SMALL LETTER U WITH ACUTE
 7B  LEFT CURLY BRACKET      FB  SMALL LETTER U WITH CIRCUMFLEX
 7C  VERTICAL LINE           FC  SMALL LETTER U WITH DIAERESIS
 7D  RIGHT CURLY BRACKET     FD  SMALL LETTER Y WITH ACUTE
 7E  TILDE                   FE  SMALL LETTER THORN (Icelandic)
                             FF  SMALL LETTER Y WITH DIAERESIS

 Hi! My name is AJay Goel
```

v. Help
Usage: `hadoop fs -help`

Return usage output

```
[Ajays-MacBook-Pro:3.1.2 ajaygoel$
[Ajays-MacBook-Pro:3.1.2 ajaygoel$ hadoop fs -help
 2019-06-10 17:52:34,981 WARN util.NativeCodeLoader: Unable to load native
 Usage: hadoop fs [generic options]
        [-appendToFile <localsrc> ... <dst>]
        [-cat [-ignoreCrc] <src> ...]
        [-checksum <src> ...]
        [-chgrp [-R] GROUP PATH...]
        [-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
        [-chown [-R] [OWNER][:[GROUP]] PATH...]
        [-copyFromLocal [-f] [-p] [-l] [-d] [-t <thread count>] <localsrc
        [-copyToLocal [-f] [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]
        [-count [-q] [-h] [-v] [-t [<storage type>]] [-u] [-x] [-e] <path
```

w. Lsr
   hadoop fs -lsr /gutenberg

```
lsr: '.': No such file or directory
[Ajays-MacBook-Pro:3.1.2 ajaygoel$ hadoop fs -lsr /gutenberg
2019-06-10 17:54:53,228 WARN util.NativeCodeLoader: Unable to load native-hadoop librar
lsr: DEPRECATED: Please use 'ls -R' instead.
-rw-r--r--   1 ajaygoel supergroup       6120 2019-06-08 16:41 /gutenberg/sample3.txt
-rw-r--r--   1 ajaygoel supergroup         24 2019-06-10 17:28 /gutenberg/sample4.txt
Ajays-MacBook-Pro:3.1.2 ajaygoel$ 
```

x. moveToLocal
   hadoop fs -moveToLocal /Gutenberg/sample3.txt
   /Users/ajaygoel/Documents/Neu/BigData/Assignments/Assignment3/Local.txt

y. put
   Same as moveFromLocal

z. rm
   removes file from the directory
   hadoop fs -rm /Assignment3/NYSE/NYSE_daily_prices_0.csv

```
Ajays-MacBook-Pro:3.1.2 ajaygoel$ hadoop fs -rm /Assignment3/NYSE/NYSE_daily_prices_0.csv
2019-06-11 17:30:44,042 WARN util.NativeCodeLoader: Unable to load native-hadoop library for yc
Deleted /Assignment3/NYSE/NYSE_daily_prices_0.csv
```

aa. rmdir
    removes the directory
    hadoop fs -rmdir /Assignment3_7/3/result

```
[Ajays-MacBook-Pro:3.1.2 ajaygoel$ hadoop fs -rmdir /Assignment3_7/3/result
2019-06-11 17:33:30,192 WARN util.NativeCodeLoader: Unable to load native-h
Ajays-MacBook-Pro:3.1.2 ajaygoel$ 
```

bb. rmr
    Recursive version of delete

cc. Setfacl
    Sets Access Control Lists (ACLs) of files and directories.

dd. setfattr
    Sets an extended attribute name and value for a file or directory.

ee. Setrep
    Changes the replication factor of a file.

ff. Stat

hadoop fs -stat /Assignment3

```
[Ajays-MacBook-Pro:3.1.2 ajaygoel$ hadoop fs -stat /Assignment3
2019-06-11 17:36:24,928 WARN util.NativeCodeLoader: Unable to load nat
2019-06-11 20:58:27
Ajays-MacBook-Pro:3.1.2 ajaygoel$
```

gg. Text

Takes a source file and outputs the file in text format

hh. Touch

Updates the access and modification times of the file specified by the URI to the current time. If the file does not exist, then a zero length file is created at URI with current time as the timestamp of that URI.

ii. Truncate

Truncate all files that match the specified file pattern to the specified length.

jj. Usage

Return the help for an individual command.

PART 5 – Programming Assignment

Copy the attached 'http_access.log' file into HDFS under /logs directory.

Using the access.log file stored in HDFS, implement MapReduce in Hadoop to find the number of times each IP accessed the website.

```
Downloads                        Music                        Public
Ajays-MacBook-Pro:~ ajaygoel$ hadoop fs -mkdir /Assignment3
2019-06-08 18:07:16,278 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Ajays-MacBook-Pro:~ ajaygoel$ hadoop fs -moveFromLocal /Users/ajaygoel/Documents/Neu/BigData/Assignments/Assignment3/accesslog.txt /Assignment3/accesslog.txt
2019-06-08 18:09:19,387 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Ajays-MacBook-Pro:~ ajaygoel$
```

Code for JAR is present in Assignment3_5 (JAVA PROJECT)

hadoop jar /Users/ajaygoel/.m2/repository/com/ajay/mr/Assignment3_5/1.0-SNAPSHOT/Assignment3_5-1.0-SNAPSHOT.jar
com.ajay.mr.assignment3_5.IpAddressMR /Assignment3/accesslog.txt
/Assignment3_5_logs

```
Ajays-MacBook-Pro:~ ajaygoel$ hadoop jar /Users/ajaygoel/.m2/repository/com/ajay/mr/Assignment3_5/1.0-SNAPSHOT/Assignment3_5-1.0-SNAPSHOT.jar com.ajay.mr.assignment3_5.IpAddressMR /Assignment3/acc
esslog.txt /Assignment3_5_logs
2019-06-08 19:00:06,375 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2019-06-08 19:00:06,847 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2019-06-08 19:00:07,337 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy th
is.
2019-06-08 19:00:07,348 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/ajaygoel/.staging/job_1560026406864_0003
2019-06-08 19:00:07,497 INFO input.FileInputFormat: Total input files to process : 1
2019-06-08 19:00:07,540 INFO mapreduce.JobSubmitter: number of splits:1
2019-06-08 19:00:07,626 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1560026406864_0003
2019-06-08 19:00:07,627 INFO mapreduce.JobSubmitter: Executing with tokens: []
2019-06-08 19:00:07,752 INFO conf.Configuration: resource-types.xml not found
2019-06-08 19:00:07,752 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2019-06-08 19:00:07,800 INFO impl.YarnClientImpl: Submitted application application_1560026406864_0003
2019-06-08 19:00:07,827 INFO mapreduce.Job: The url to track the job: http://Ajays-MacBook-Pro.local:8088/proxy/application_1560026406864_0003/
2019-06-08 19:00:07,827 INFO mapreduce.Job: Running job: job_1560026406864_0003
2019-06-08 19:00:12,894 INFO mapreduce.Job: Job job_1560026406864_0003 running in uber mode : false
2019-06-08 19:00:12,896 INFO mapreduce.Job:  map 0% reduce 0%
2019-06-08 19:00:16,948 INFO mapreduce.Job:  map 100% reduce 0%
2019-06-08 19:00:21,989 INFO mapreduce.Job:  map 100% reduce 100%
2019-06-08 19:00:22,000 INFO mapreduce.Job: Job job_1560026406864_0003 completed successfully
2019-06-08 19:00:22,084 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=709513
                FILE: Number of bytes written=1851161
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=3497779
                HDFS: Number of bytes written=31931
                HDFS: Number of read operations=8
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=2037
```

```
                Bytes Written=31931
[Ajays-MacBook-Pro:~ ajaygoel$ hadoop fs -ls /Assignment3_5_logs
2019-06-08 19:05:37,949 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes
Found 2 items
-rw-r--r--   1 ajaygoel supergroup          0 2019-06-08 19:00 /Assignment3_5_logs/_SUCCESS
-rw-r--r--   1 ajaygoel supergroup      31931 2019-06-08 19:00 /Assignment3_5_logs/part-r-00000
Ajays-MacBook-Pro:~ ajaygoel$
```

hadoop fs -cat /Assignment3_5_logs/part-r-00000

```
-rw-r--r--   1 ajaygoel supergroup        31931 2019-06-08 19:00 /Assignment3_5_
[Ajays-MacBook-Pro:~ ajaygoel$ hadoop fs -cat /Assignment3_5_logs/part-r-00000
2019-06-08 19:06:15,593 WARN util.NativeCodeLoader: Unable to load native-hado
1.162.207.87    4
1.170.44.84     83
1.192.146.100   1
1.202.184.142   1
1.202.184.145   1
1.202.89.134    2
1.234.2.41      12
1.56.79.5       4
1.59.91.151     4
1.62.189.221    4
1.85.17.247     1
10.15.10.129    2812
10.15.10.135    2108
10.15.10.144    2
10.15.10.151    4
10.15.11.112    2
10.15.8.173     3
10.15.8.20      5
```

PART 6 – Programming Assignment

Goel, Ajay: Nu Id (001897443)

Download and Copy all the files (http://msis.neu.edu/nyse/) (DailyPrices_A to DailyPrices_Z) to a folder in HDFS.

Write a MapReduce to find the Max price of stock_price_high for each stock. Capture the running time programmatically (or manually using a wristwatch or smartphone).

```
hadoop fs -copyFromLocal
/Users/ajaygoel/Documents/Neu/BigData/Assignments/Assignment3/NYSE
/Assignment3/NYSE2
```

SHELL SCRIPT in command.sh:

```
chars=( {a..z} )
i=0
for filename in `hadoop fs -ls /Assignment3/NYSE2 | awk '{print $NF}' | grep .csv$ | tr '\n'
' '`
do
        echo $filename;
        hadoop jar /Users/ajaygoel/.m2/repository/com/ajay/mr/Assignment3_6/1.0-
SNAPSHOT/Assignment3_6-1.0-SNAPSHOT.jar com.ajay.mr.assignment3_6.NYSEMR
$filename /Assignment3_6/"${chars[i++]}"
done
```

FOR JAR: PROJECT in Assignment3_6 (JAVA PROJECT)

Goel, Ajay: Nu Id (001897443)

```
/Assignment3/NYSE2/NYSE_daily_prices_Z.csv
2019-06-10 17:10:36,123 WARN util.NativeCodeLoader: Unable to load nati
2019-06-10 17:10:36,591 INFO client.RMProxy: Connecting to ResourceMana
2019-06-10 17:10:36,937 WARN mapreduce.JobResourceUploader: Hadoop comm
2019-06-10 17:10:36,947 INFO mapreduce.JobResourceUploader: Disabling E
2019-06-10 17:10:37,081 INFO input.FileInputFormat: Total input files
2019-06-10 17:10:37,522 INFO mapreduce.JobSubmitter: number of splits:1
2019-06-10 17:10:37,618 INFO mapreduce.JobSubmitter: Submitting tokens
2019-06-10 17:10:37,619 INFO mapreduce.JobSubmitter: Executing with tok
2019-06-10 17:10:37,750 INFO conf.Configuration: resource-types.xml not
2019-06-10 17:10:37,750 INFO resource.ResourceUtils: Unable to find 're
2019-06-10 17:10:37,796 INFO impl.YarnClientImpl: Submitted application
2019-06-10 17:10:37,825 INFO mapreduce.Job: The url to track the job: h
2019-06-10 17:10:37,826 INFO mapreduce.Job: Running job: job_1560192492
2019-06-10 17:10:45,921 INFO mapreduce.Job: Job job_1560192492979_0059
2019-06-10 17:10:45,923 INFO mapreduce.Job:  map 0% reduce 0%
2019-06-10 17:10:49,980 INFO mapreduce.Job:  map 100% reduce 0%
2019-06-10 17:10:55,016 INFO mapreduce.Job:  map 100% reduce 20%
2019-06-10 17:10:56,020 INFO mapreduce.Job:  map 100% reduce 40%
2019-06-10 17:10:57,027 INFO mapreduce.Job:  map 100% reduce 60%
2019-06-10 17:10:58,034 INFO mapreduce.Job:  map 100% reduce 80%
2019-06-10 17:10:59,046 INFO mapreduce.Job:  map 100% reduce 100%
2019-06-10 17:10:59,057 INFO mapreduce.Job: Job job_1560192492979_0059
2019-06-10 17:10:59,144 INFO mapreduce.Job: Counters: 50
        File System Counters
                FILE: Number of bytes read=381699
                FILE: Number of bytes written=2059867
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=2093552
                HDFS: Number of bytes written=94
                HDFS: Number of read operations=28
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=10
        Job Counters
                Killed reduce tasks=1
                Launched map tasks=1
                Launched reduce tasks=5
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=198
                Total time spent by all reduces in occupied slots (ms)=
                Total time spent by all map tasks (ms)=1980
                Total time spent by all reduce tasks (ms)=10412
                Total vcore-milliseconds taken by all map tasks=1980
                Total vcore-milliseconds taken by all reduce tasks=1041
                Total megabyte-milliseconds taken by all map tasks=2027
                Total megabyte-milliseconds taken by all reduce tasks=1
        Map-Reduce Framework
                Map input records=38821
                Map output records=38820
                Map output bytes=304029
                Map output materialized bytes=381699
                Input split bytes=128
                Combine input records=0
                Combine output records=0
                Reduce input groups=10
                Reduce shuffle bytes=381699
```

```
[Ajays-MacBook-Pro:3.1.2 ajaygoel$
[Ajays-MacBook-Pro:3.1.2 ajaygoel$ hadoop fs -cat /Assignment3_6/a
 2019-06-10 17:13:49,354 WARN util.NativeCodeLoader: Unable to load
 cat: `/Assignment3_6/a': Is a directory
[Ajays-MacBook-Pro:3.1.2 ajaygoel$ hadoop fs -cat /Assignment3_6/a/p
 2019-06-10 17:13:59,490 WARN util.NativeCodeLoader: Unable to load
 ABD      28.58
 ABX      54.74
 ACC      37.0
 ACH      111.6
 ACM      38.25
 AEA      23.94
 AEF      27.0
 AEP      53.31
 AFE      26.7
 AGD      25.5
 AGN      125.0
 AHC      16.35
 AIB      125.0
 AIG      157.19
 AIQ      15.4
 AIV      65.79
 AKT      26.25
 ALD      33.35
 ALX      471.0
```

Ajays-MacBook-Pro:3.1.2 ajaygoel$ hadoop fs -cat /Assignment3_6/a/part-r-00000

It is running in 9 Minutes 45 Seconds

PART 7 – Programming Assignment
Write one MapReduce program using each of the classes that extend
FileInputFormat<k,v>
(CombineFileInputFormat, FixedLengthInputFormat, KeyValueTextInputFormat,
NLineInputFormat, SequenceFileInputFormat, TextInputFormat)
  http://hadoop.apache.org/docs/current/api/org/apache/hadoop/mapreduce/lib/input/FileInputFormat.
html
You could use any input file of your choice. The size of the input files is not important.
The MR programs could simply do counting, or any other analysis you choose.

Goel, Ajay: Nu Id (001897443)

FIRST:
hadoop jar /Users/ajaygoel/.m2/repository/com/ajay/mr/Assignment3_7_1/1.0-
SNAPSHOT/Assignment3_7_1-1.0-SNAPSHOT.jar
com.ajay.mr.assignment3_7_1.CombineFileMR /gutenberg/sample3.txt /gutenberg/sample4.txt
/Assignment3_7/1/result

```
2019-06-14 18:33:38,826 INFO Impl.YarnClientImpl: Submitted application application_1560540941178_0004
2019-06-14 18:33:38,854 INFO mapreduce.Job: The url to track the job: http://Ajays-MacBook-Pro.local:8088/proxy/applic
2019-06-14 18:33:38,854 INFO mapreduce.Job: Running job: job_1560540941178_0004
2019-06-14 18:33:43,950 INFO mapreduce.Job: Job job_1560540941178_0004 running in uber mode : false
2019-06-14 18:33:43,952 INFO mapreduce.Job:  map 0% reduce 0%
2019-06-14 18:33:48,026 INFO mapreduce.Job:  map 100% reduce 0%
2019-06-14 18:33:52,054 INFO mapreduce.Job:  map 100% reduce 100%
2019-06-14 18:33:52,065 INFO mapreduce.Job: Job job_1560540941178_0004 completed successfully
2019-06-14 18:33:52,143 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=6676
```

THIRD:
I have created a key value of department and name and I want to count the number of
departments and the count in each department.

JAVA Project code in Assignment3_7_3

hadoop jar /Users/ajaygoel/.m2/repository/com/ajay/mr/Assignment3_7_3/1.0-
SNAPSHOT/Assignment3_7_3-1.0-SNAPSHOT.jar
com.ajay.mr.assignment3_7_3.DepartmentMR /Assignment3/part7_3 /Assignment3_7/3/result

```
plication with ToolRunner to remedy this.
2019-06-11 20:47:21,840 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for pa
09
2019-06-11 20:47:21,990 INFO input.FileInputFormat: Total input files to process : 1
2019-06-11 20:47:22,033 INFO mapreduce.JobSubmitter: number of splits:1
2019-06-11 20:47:22,125 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_156027
2019-06-11 20:47:22,126 INFO mapreduce.JobSubmitter: Executing with tokens: []
2019-06-11 20:47:22,253 INFO conf.Configuration: resource-types.xml not found
2019-06-11 20:47:22,253 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2019-06-11 20:47:22,301 INFO impl.YarnClientImpl: Submitted application application_156027
2019-06-11 20:47:22,330 INFO mapreduce.Job: The url to track the job: http://Ajays-MacBook-
2019-06-11 20:47:22,330 INFO mapreduce.Job: Running job: job_1560276490952_0009
2019-06-11 20:47:27,408 INFO mapreduce.Job: Job job_1560276490952_0009 running in uber mode
2019-06-11 20:47:27,409 INFO mapreduce.Job:  map 0% reduce 0%
2019-06-11 20:47:31,468 INFO mapreduce.Job:  map 100% reduce 0%
2019-06-11 20:47:35,498 INFO mapreduce.Job:  map 100% reduce 100%
2019-06-11 20:47:35,509 INFO mapreduce.Job: Job job_1560276490952_0009 completed successful
2019-06-11 20:47:35,600 INFO mapreduce.Job: Counters: 49
        File System Counters
```

```
2019-06-11 20:48:08,974 WARN util.NativeCodeLoader: Unable to load na
cat: `/Assignment3_7/3/result': Is a directory
[Ajays-MacBook-Pro:3.1.2 ajaygoel$ hadoop fs -cat /Assignment3_7/3/resu
2019-06-11 20:48:38,018 WARN util.NativeCodeLoader: Unable to load nat
D1      10
D2       5
D3       5
D4       5
Department       1
Ajays-MacBook-Pro:3.1.2 ajaygoel$ ▌
```

FOURTH:
I have chosen 4 lines should go to a mapper,

Number of splits:52
JAVA Project code in Assignment3_7_4
hadoop jar /Users/ajaygoel/.m2/repository/com/ajay/mr/Assignment3_7_4/1.0-
SNAPSHOT/Assignment3_7_4-1.0-SNAPSHOT.jar
com.ajay.mr.assignment3_7_4.fourLinesMR /gutenberg/sample3.txt /Assignment3_7/4/result

```
2019-06-11 22:13:03,731 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hado
15
2019-06-11 22:13:03,878 INFO input.FileInputFormat: Total input files to process : 2
2019-06-11 22:13:04,341 INFO mapreduce.JobSubmitter: number of splits:52
2019-06-11 22:13:04,840 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1560276490952_0015
2019-06-11 22:13:04,843 INFO mapreduce.JobSubmitter: Executing with tokens: []
2019-06-11 22:13:04,977 INFO conf.Configuration: resource-types.xml not found
2019-06-11 22:13:04,977 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2019-06-11 22:13:05,026 INFO impl.YarnClientImpl: Submitted application application_1560276490952_0015
2019-06-11 22:13:05,057 INFO mapreduce.Job: The url to track the job: http://Ajays-MacBook-Pro.local:808
2019-06-11 22:13:05,058 INFO mapreduce.Job: Running job: job_1560276490952_0015
2019-06-11 22:13:11,150 INFO mapreduce.Job: Job job_1560276490952_0015 running in uber mode : false
2019-06-11 22:13:11,152 INFO mapreduce.Job:  map 0% reduce 0%
2019-06-11 22:13:15,202 INFO mapreduce.Job:  map 2% reduce 0%
2019-06-11 22:13:16,208 INFO mapreduce.Job:  map 4% reduce 0%
2019-06-11 22:13:17,216 INFO mapreduce.Job:  map 6% reduce 0%
2019-06-11 22:13:18,224 INFO mapreduce.Job:  map 8% reduce 0%
2019-06-11 22:13:20,237 INFO mapreduce.Job:  map 12% reduce 0%
2019-06-11 22:13:21,242 INFO mapreduce.Job:  map 13% reduce 0%
2019-06-11 22:13:22,250 INFO mapreduce.Job:  map 19% reduce 0%
2019-06-11 22:13:24,260 INFO mapreduce.Job:  map 23% reduce 0%
2019-06-11 22:13:27,279 INFO mapreduce.Job:  map 27% reduce 0%
2019-06-11 22:13:28,283 INFO mapreduce.Job:  map 31% reduce 0%
2019-06-11 22:13:29,289 INFO mapreduce.Job:  map 33% reduce 0%
2019-06-11 22:13:31,301 INFO mapreduce.Job:  map 37% reduce 0%
2019-06-11 22:13:32,308 INFO mapreduce.Job:  map 40% reduce 0%
2019-06-11 22:13:33,313 INFO mapreduce.Job:  map 42% reduce 0%
2019-06-11 22:13:36,332 INFO mapreduce.Job:  map 50% reduce 0%
2019-06-11 22:13:37,342 INFO mapreduce.Job:  map 52% reduce 0%
2019-06-11 22:13:40,361 INFO mapreduce.Job:  map 52% reduce 17%
2019-06-11 22:13:41,370 INFO mapreduce.Job:  map 60% reduce 17%
2019-06-11 22:13:42,378 INFO mapreduce.Job:  map 62% reduce 17%
2019-06-11 22:13:46,399 INFO mapreduce.Job:  map 62% reduce 21%
2019-06-11 22:13:47,405 INFO mapreduce.Job:  map 71% reduce 21%
2019-06-11 22:13:51,428 INFO mapreduce.Job:  map 79% reduce 21%
2019-06-11 22:13:52,436 INFO mapreduce.Job:  map 81% reduce 25%
2019-06-11 22:13:55,455 INFO mapreduce.Job:  map 83% reduce 25%
2019-06-11 22:13:57,465 INFO mapreduce.Job:  map 90% reduce 25%
2019-06-11 22:13:58,473 INFO mapreduce.Job:  map 92% reduce 30%
2019-06-11 22:14:01,492 INFO mapreduce.Job:  map 100% reduce 30%
2019-06-11 22:14:02,502 INFO mapreduce.Job:  map 100% reduce 100%
2019-06-11 22:14:02,522 INFO mapreduce.Job: Job job_1560276490952_0015 completed successfully
2019-06-11 22:14:02,682 INFO mapreduce.Job: Counters: 49
        File System Counters
```

FIFTH: Sequence
hadoop jar /Users/ajaygoel/.m2/repository/com/ajay/mr/Assignment3_7_5/1.0-
SNAPSHOT/Assignment3_7_5-1.0-SNAPSHOT.jar
com.ajay.mr.assignment3_7_5.fourLinesMR /Assignment3/sequence.txt /Assignment3/5/result

```
Ajays-MacBook-Pro:~ ajaygoel$
Ajays-MacBook-Pro:~ ajaygoel$ hadoop fs -moveFromLocal /Users/ajaygoel/Documents/Neu/BigData/Assignments/Assignment3/input.txt /Assignment3/sequence2.txt
2019-06-13 17:24:03,926 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Ajays-MacBook-Pro:~ ajaygoel$ hadoop jar /Users/ajaygoel/.m2/repository/com/ajay/mr/Assignment3_7_5/1.0-SNAPSHOT/Assignment3_7_5-1.0-SNAPSHOT.jar com.aja
nce2.txt /Assignment3_7/5/result
2019-06-13 17:24:37,365 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2019-06-13 17:24:37,833 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2019-06-13 17:24:38,367 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute yo
2019-06-13 17:24:38,378 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/ajaygoel/.staging/job_15604573073
2019-06-13 17:24:38,527 INFO input.FileInputFormat: Total input files to process : 1
2019-06-13 17:24:38,558 INFO mapreduce.JobSubmitter: number of splits:1
2019-06-13 17:24:38,645 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1560457307388_0015
2019-06-13 17:24:38,646 INFO mapreduce.JobSubmitter: Executing with tokens: []
2019-06-13 17:24:38,767 INFO conf.Configuration: resource-types.xml not found
2019-06-13 17:24:38,767 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2019-06-13 17:24:38,812 INFO impl.YarnClientImpl: Submitted application application_1560457307388_0015
2019-06-13 17:24:38,839 INFO mapreduce.Job: The url to track the job: http://Ajays-MacBook-Pro.local:8088/proxy/application_1560457307388_0015/
2019-06-13 17:24:38,840 INFO mapreduce.Job: Running job: job_1560457307388_0015
2019-06-13 17:24:43,909 INFO mapreduce.Job: Job job_1560457307388_0015 running in uber mode : false
2019-06-13 17:24:43,911 INFO mapreduce.Job:  map 0% reduce 0%
2019-06-13 17:24:47,964 INFO mapreduce.Job:  map 100% reduce 0%
2019-06-13 17:24:52,000 INFO mapreduce.Job:  map 100% reduce 100%
2019-06-13 17:24:52,011 INFO mapreduce.Job: Job job_1560457307388_0015 completed successfully
2019-06-13 17:24:52,097 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=27
```

SIXTH: Text Input format
Done in case of NYSE above in part 5 and 6