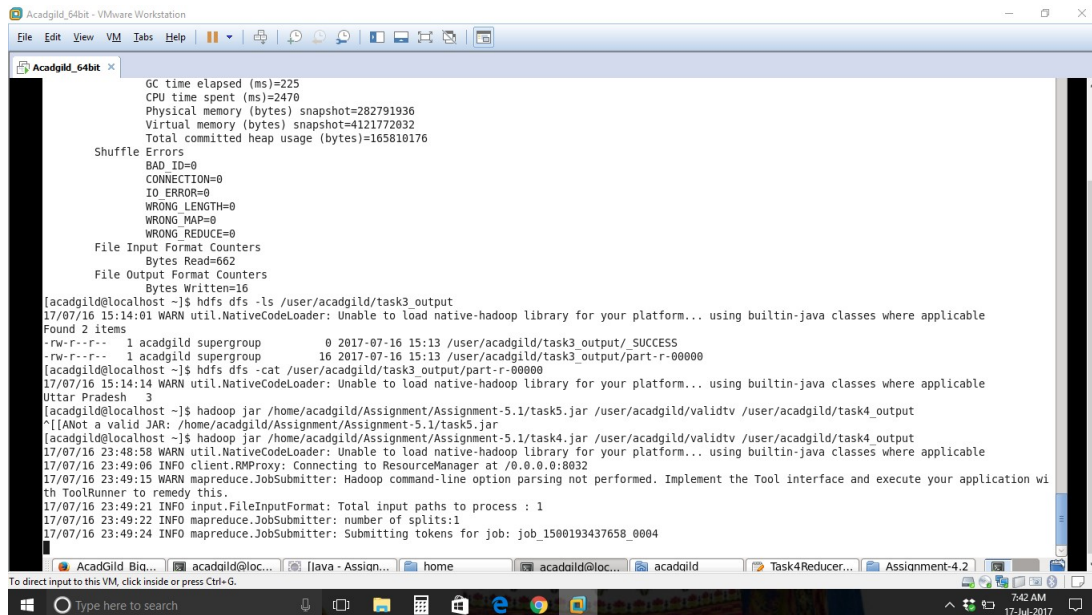


Assignment 5.1 Screenshots

Task 4:

Write a Map Reduce program to modify Task 2 (refer session 4, assignment 2) to use a custom partitioner with 4 reducers. Make sure that all records whose company name starts with A-F (upper or lower case) should go to 1st reducer, those starting with G-L to 2nd reducer, those starting with M-R to 3rd reducer and others to 4th reducer.



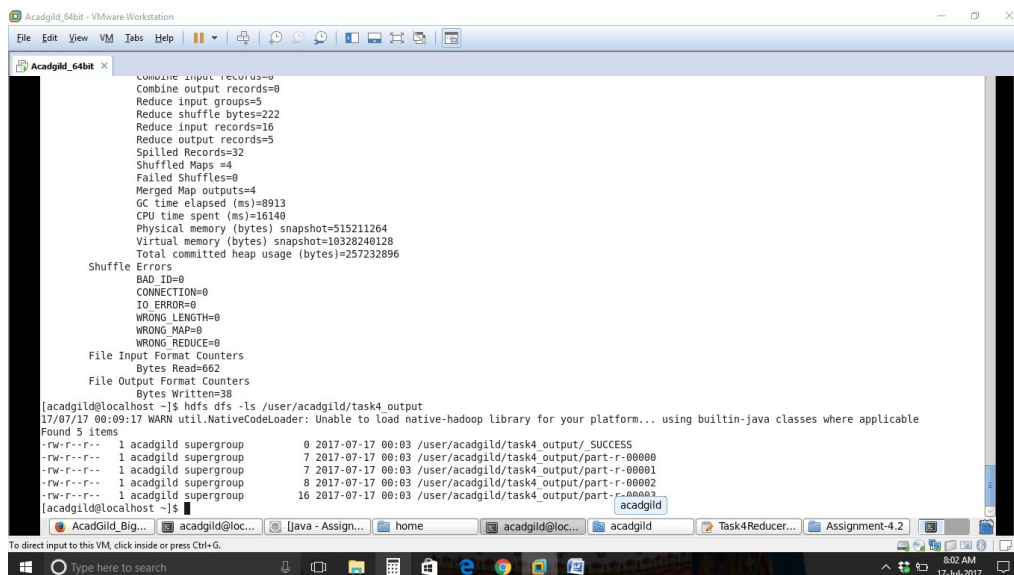
```
GC time elapsed (ms)=225
CPU time spent (ms)=2470
Physical memory (bytes) snapshot=282791936
Virtual memory (bytes) snapshot=4121772032
Total committed heap usage (bytes)=165810176

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=662
File Output Format Counters
  Bytes Written=16

[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/task3 output
17/07/16 15:14:01 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup          0 2017-07-16 15:13 /user/acadgild/task3 output/_SUCCESS
-rw-r--r-- 1 acadgild supergroup        16 2017-07-16 15:13 /user/acadgild/task3 output/part-r-00000
[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/task3 output/part-r-00000
Uttar Pradesh 3
[acadgild@localhost ~]$ hadoop jar /home/acadgild/Assignment/Assignment-5.1/task5.jar /user/acadgild/validtv /user/acadgild/task4_output
^[[Anot a valid JAR: /home/acadgild/Assignment/Assignment-5.1/task5.jar
[acadgild@localhost ~]$ hadoop jar /home/acadgild/Assignment/Assignment-5.1/task4.jar /user/acadgild/validtv /user/acadgild/task4_output
17/07/16 23:48:58 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
17/07/16 23:49:06 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
17/07/16 23:49:15 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
17/07/16 23:49:21 INFO input.FileInputFormat: Total input paths to process : 1
17/07/16 23:49:22 INFO mapreduce.JobSubmitter: number of splits:1
17/07/16 23:49:24 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1500193437658_0004
```

Ouptut hdfs folder shows four files for four reducers



```
Combine input records=0
Combine output records=0
Reduce input groups=5
Reduce shuffle bytes=222
Reduce input records=16
Reduce output records=5
Spilled Records=32
Shuffled Maps =4
Failed Shuffles=0
Merged Map outputs=4
GC time elapsed (ms)=8913
CPU time spent (ms)=16140
Physical memory (bytes) snapshot=515211264
Virtual memory (bytes) snapshot=10328240128
Total committed heap usage (bytes)=257232896

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=662
File Output Format Counters
  Bytes Written=38

[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/task4 output
17/07/17 00:09:17 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 5 items
-rw-r--r-- 1 acadgild supergroup          0 2017-07-17 00:03 /user/acadgild/task4 output/_SUCCESS
-rw-r--r-- 1 acadgild supergroup          7 2017-07-17 00:03 /user/acadgild/task4 output/part-r-00000
-rw-r--r-- 1 acadgild supergroup          7 2017-07-17 00:03 /user/acadgild/task4 output/part-r-00001
-rw-r--r-- 1 acadgild supergroup          8 2017-07-17 00:03 /user/acadgild/task4 output/part-r-00002
-rw-r--r-- 1 acadgild supergroup        16 2017-07-17 00:03 /user/acadgild/task4 output/part-r-00003
[acadgild@localhost ~]$
```

Subsequent four screenshots show four reducers output for companies A-F,G_L,M-R,S-Z respectively

Acadgild_64bit - VMware Workstation

```
Reduce shuffle bytes=222
Reduce input records=16
Reduce output records=5
Spilled Records=32
Shuffled Maps=4
Failed Shuffles=0
Merged Map outputs=4
GC time elapsed (ms)=8913
CPU time spent (ms)=16140
Physical memory (bytes) snapshot=515211264
Virtual memory (bytes) snapshot=10328240128
Total committed heap usage (bytes)=257232896

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=662
File Output Format Counters
Bytes Written=38

[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/task4 output
Found 5 items
-rw-r--r-- 1 acadgild supergroup 0 2017-07-17 00:03 /user/acadgild/task4 output/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 7 2017-07-17 00:03 /user/acadgild/task4 output/part-r-00000
-rw-r--r-- 1 acadgild supergroup 7 2017-07-17 00:03 /user/acadgild/task4 output/part-r-00001
-rw-r--r-- 1 acadgild supergroup 8 2017-07-17 00:03 /user/acadgild/task4 output/part-r-00002
-rw-r--r-- 1 acadgild supergroup 16 2017-07-17 00:03 /user/acadgild/task4 output/part-r-00003
[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/task4 output/part-r-00000
Akai 1
[acadgild@localhost ~]$
```

Task4Reducer.java (~\Assignment\Assignment-5.1) - gedit

AcadGild_Big... acadgild@loc... [Java - Assign... home acadgild@loc... acadgild Task4Reducer... Assignment-4.2

To direct input to this VM, click inside or press Ctrl+G.

Type here to search 8:02 AM 17-Jul-2017

Acadgild_64bit - VMware Workstation

```
Spilled Records=32
Shuffled Maps=4
Failed Shuffles=0
Merged Map outputs=4
GC time elapsed (ms)=8913
CPU time spent (ms)=16140
Physical memory (bytes) snapshot=515211264
Virtual memory (bytes) snapshot=10328240128
Total committed heap usage (bytes)=257232896

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=662
File Output Format Counters
Bytes Written=38

[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/task4 output
Found 5 items
-rw-r--r-- 1 acadgild supergroup 0 2017-07-17 00:03 /user/acadgild/task4 output/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 7 2017-07-17 00:03 /user/acadgild/task4 output/part-r-00000
-rw-r--r-- 1 acadgild supergroup 7 2017-07-17 00:03 /user/acadgild/task4 output/part-r-00001
-rw-r--r-- 1 acadgild supergroup 8 2017-07-17 00:03 /user/acadgild/task4 output/part-r-00002
-rw-r--r-- 1 acadgild supergroup 16 2017-07-17 00:03 /user/acadgild/task4 output/part-r-00003
[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/task4 output/part-r-00000
Akai 1
[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/task4 output/part-r-00001
Lava 3
[acadgild@localhost ~]$
```

acadgild

AcadGild_Big... acadgild@loc... [Java - Assign... home acadgild@loc... acadgild Task4Reducer... Assignment-4.2

To direct input to this VM, click inside or press Ctrl+G.

Type here to search 8:03 AM 17-Jul-2017

```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Acadgild_64bit x
merged map outputs=4
GC time elapsed (ms)=8913
CPU time spent (ms)=16140
Physical memory (bytes) snapshot=515211264
Virtual memory (bytes) snapshot=10328240128
Total committed heap usage (bytes)=257232896

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=662

File Output Format Counters
Bytes Written=38

[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/task4_output
17/07/17 00:09:17 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 5 items
-rw-r--r-- 1 acadgild supergroup 0 2017-07-17 00:03 /user/acadgild/task4_output/ SUCCESS
-rw-r--r-- 1 acadgild supergroup 7 2017-07-17 00:03 /user/acadgild/task4_output/part-r-00000
-rw-r--r-- 1 acadgild supergroup 7 2017-07-17 00:03 /user/acadgild/task4_output/part-r-00001
-rw-r--r-- 1 acadgild supergroup 8 2017-07-17 00:03 /user/acadgild/task4_output/part-r-00002
-rw-r--r-- 1 acadgild supergroup 16 2017-07-17 00:03 /user/acadgild/task4_output/part-r-00003
[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/task4_output/part-r-00000
17/07/17 00:09:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Akai 1
[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/task4_output/part-r-00001
17/07/17 00:10:07 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Lava 3
[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/task4_output/part-r-00002
17/07/17 00:10:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Onida 3
[acadgild@localhost ~]$
home
AcadGild_Big... acadgild@loc... [Java - Assign... home acadgild@loc... acadgild Task4Reducer... Assignment-4.2
```

```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Acadgild_64bit x
Virtual memory (bytes) snapshot=10328240128
Total committed heap usage (bytes)=257232896

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

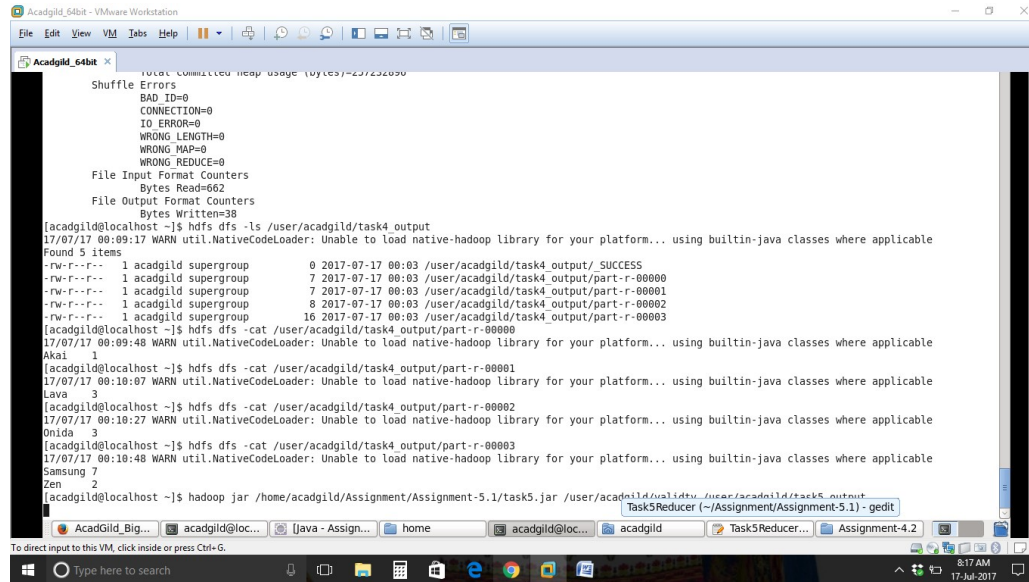
File Input Format Counters
Bytes Read=662

File Output Format Counters
Bytes Written=38

[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/task4_output
17/07/17 00:09:17 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 5 items
-rw-r--r-- 1 acadgild supergroup 0 2017-07-17 00:03 /user/acadgild/task4_output/ SUCCESS
-rw-r--r-- 1 acadgild supergroup 7 2017-07-17 00:03 /user/acadgild/task4_output/part-r-00000
-rw-r--r-- 1 acadgild supergroup 7 2017-07-17 00:03 /user/acadgild/task4_output/part-r-00001
-rw-r--r-- 1 acadgild supergroup 8 2017-07-17 00:03 /user/acadgild/task4_output/part-r-00002
-rw-r--r-- 1 acadgild supergroup 16 2017-07-17 00:03 /user/acadgild/task4_output/part-r-00003
[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/task4_output/part-r-00000
17/07/17 00:09:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Akai 1
[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/task4_output/part-r-00001
17/07/17 00:10:07 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Lava 3
[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/task4_output/part-r-00002
17/07/17 00:10:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Onida 3
[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/task4_output/part-r-00003
17/07/17 00:10:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Samsung 7
Zen 2
[acadgild@localhost ~]$
Task4Reducer.java (~/.Assignment/Assignment-5.1) - gedit
AcadGild_Big... acadgild@loc... [Java - Assign... home acadgild@loc... acadgild Task4Reducer... Assignment-4.2
```

Task 5:

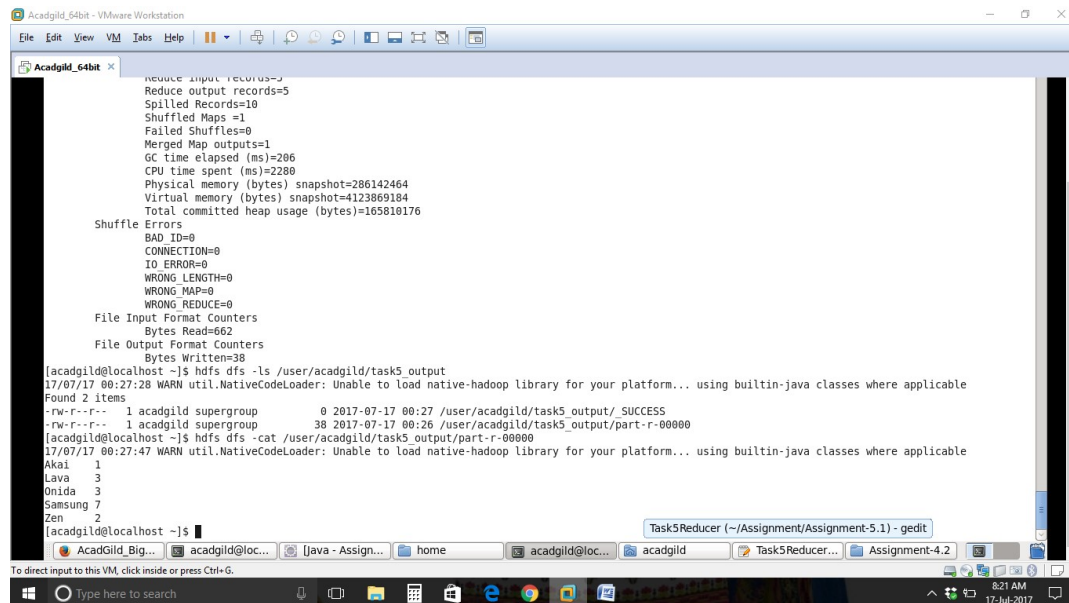
Modify Task 2 (refer session 4, assignment 2) to take advantage of Combiner.



The screenshot shows a terminal window titled 'Acadgild_64bit' within a VMware Workstation. The terminal displays the following output:

```
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=662
File Output Format Counters
Bytes Written=38
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/task4 output
Found 5 items
-rw-r--r-- 1 acadgild supergroup 0 2017-07-17 00:03 /user/acadgild/task4 output/ SUCCESS
-rw-r--r-- 1 acadgild supergroup 7 2017-07-17 00:03 /user/acadgild/task4 output/part-r-00000
-rw-r--r-- 1 acadgild supergroup 7 2017-07-17 00:03 /user/acadgild/task4 output/part-r-00001
-rw-r--r-- 1 acadgild supergroup 8 2017-07-17 00:03 /user/acadgild/task4 output/part-r-00002
-rw-r--r-- 1 acadgild supergroup 16 2017-07-17 00:03 /user/acadgild/task4 output/part-r-00003
[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/task4 output/part-r-00000
17/07/17 00:09:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Akai 1
[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/task4 output/part-r-00001
17/07/17 00:10:07 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Lava 3
[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/task4 output/part-r-00002
17/07/17 00:10:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Onida 3
[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/task4 output/part-r-00003
17/07/17 00:10:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Samsung 7
[acadgild@localhost ~]$ hadoop jar /home/acadgild/Assignment/Assignment-5.1/task5.jar /user/acadgild/output /user/acadgild/task5_output
Task5Reducer (~/.Assignment/Assignment-5.1) - gedit
```

All 4 reducers output is combined



The screenshot shows a terminal window titled 'Acadgild_64bit' within a VMware Workstation. The terminal displays the following output:

```
reduce input records=5
Reduce output records=5
Spilled Records=10
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=206
CPU time spent (ms)=2280
Physical memory (bytes) snapshot=286142464
Virtual memory (bytes) snapshot=4123869184
Total committed heap usage (bytes)=165810176
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=662
File Output Format Counters
Bytes Written=38
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/task5 output
Found 2 items
-rw-r--r-- 1 acadgild supergroup 0 2017-07-17 00:27 /user/acadgild/task5 output/ SUCCESS
-rw-r--r-- 1 acadgild supergroup 38 2017-07-17 00:26 /user/acadgild/task5 output/part-r-00000
[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/task5 output/part-r-00000
17/07/17 00:27:47 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Akai 1
Lava 3
Onida 3
Samsung 7
Zen 2
[acadgild@localhost ~]$
```