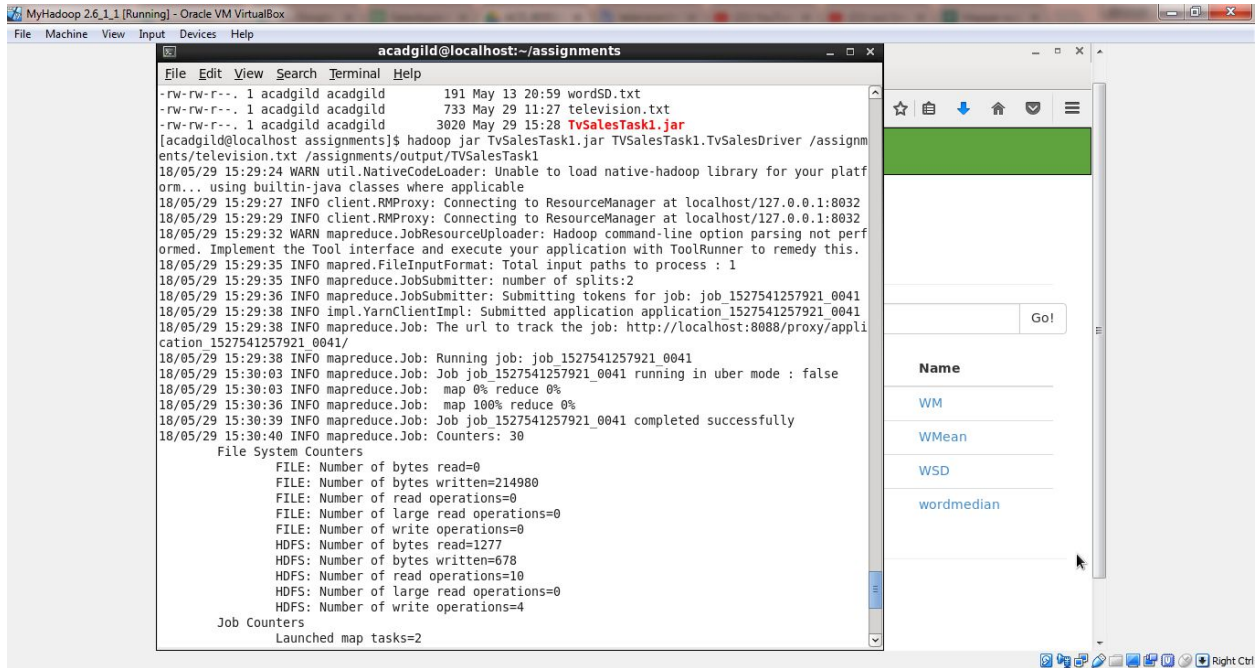
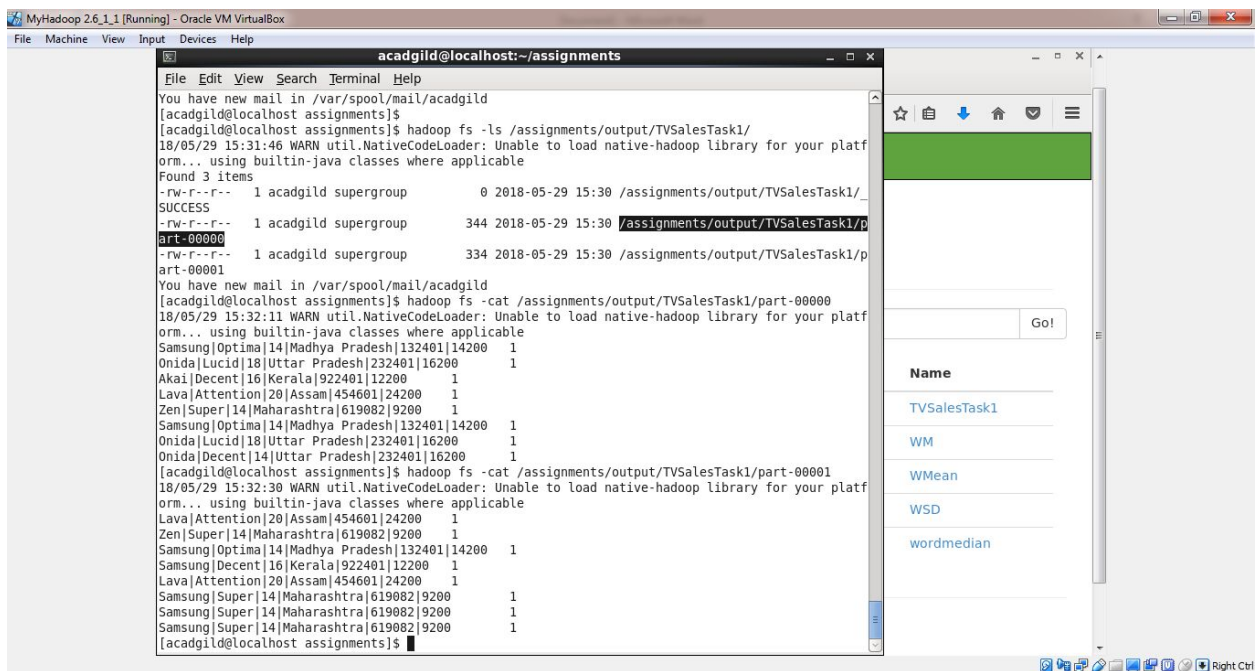


Task 1:

Write a Map Reduce program to filter out the invalid records. Map only job will fit for this context.



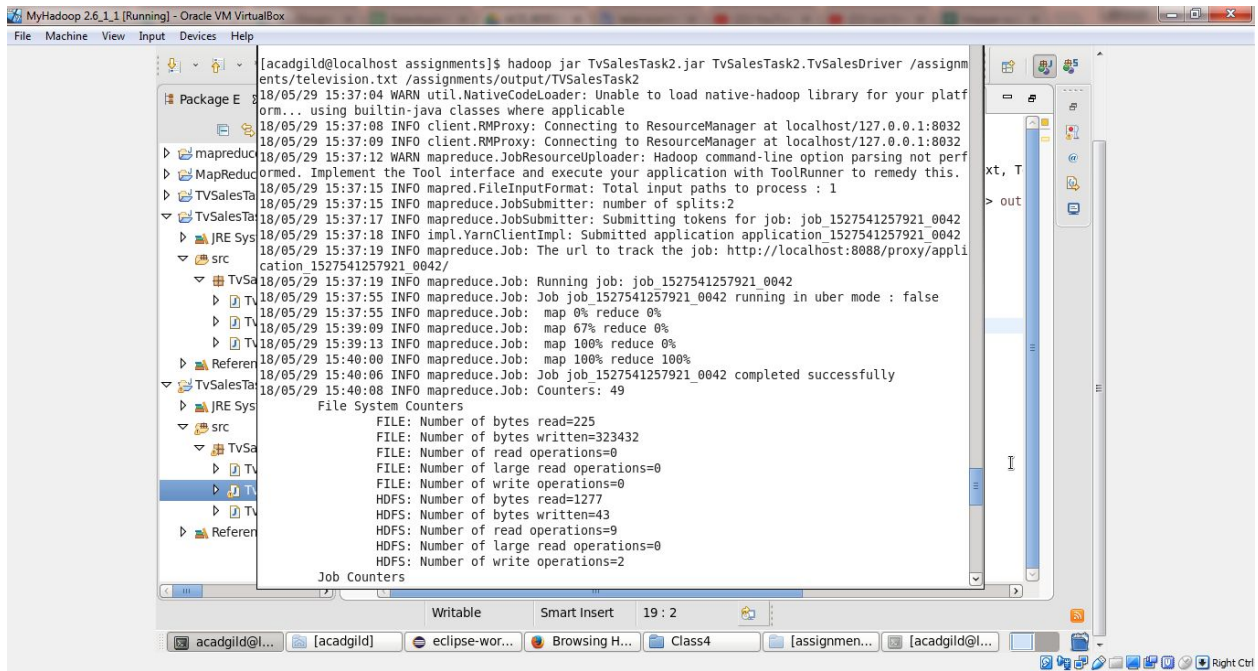
```
acacgild@localhost:~/assignments
File Edit View Search Terminal Help
-rw-rw-r-- 1 acacgild acacgild 191 May 13 20:59 wordSD.txt
-rw-rw-r-- 1 acacgild acacgild 733 May 29 11:27 television.txt
-rw-rw-r-- 1 acacgild acacgild 3020 May 29 15:28 TVSalesTask1.jar
[acacgild@localhost assignments]$ hadoop jar TVSalesTask1.jar TVSalesDriver /assignm
ents/television.txt /assignments/output/TVSalesTask1
18/05/29 15:29:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
18/05/29 15:29:27 INFO client.RMPProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/05/29 15:29:29 INFO client.RMPProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/05/29 15:29:32 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not perf
ormed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/05/29 15:29:35 INFO mapreduce.FileInputFormat: Total input paths to process : 1
18/05/29 15:29:35 INFO mapreduce.JobSubmitter: number of splits:2
18/05/29 15:29:36 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1527541257921_0041
18/05/29 15:29:38 INFO impl.YarnClientImpl: Submitted application application_1527541257921_0041
18/05/29 15:29:38 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/appli
cation_1527541257921_0041/
18/05/29 15:29:38 INFO mapreduce.Job: Running job: job_1527541257921_0041
18/05/29 15:30:03 INFO mapreduce.Job: Job job_1527541257921_0041 running in uber mode : false
18/05/29 15:30:03 INFO mapreduce.Job: map 0% reduce 0%
18/05/29 15:30:36 INFO mapreduce.Job: map 100% reduce 0%
18/05/29 15:30:39 INFO mapreduce.Job: Job job_1527541257921_0041 completed successfully
18/05/29 15:30:40 INFO mapreduce.Job: Counters: 30
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=214900
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=1277
  HDFS: Number of bytes written=678
  HDFS: Number of read operations=10
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=4
Job Counters
  Launched map tasks=2
```



```
acacgild@localhost:~/assignments
File Edit View Search Terminal Help
You have new mail in /var/spool/mail/acacgild
[acacgild@localhost assignments]$
[acacgild@localhost assignments]$ hadoop fs -ls /assignments/output/TVSalesTask1/
18/05/29 15:31:46 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
Found 3 items
-rw-r--r-- 1 acacgild supergroup 0 2018-05-29 15:30 /assignments/output/TVSalesTask1/
SUCCESS
-rw-r--r-- 1 acacgild supergroup 344 2018-05-29 15:30 /assignments/output/TVSalesTask1/p
art-00000
-rw-r--r-- 1 acacgild supergroup 334 2018-05-29 15:30 /assignments/output/TVSalesTask1/p
art-00001
You have new mail in /var/spool/mail/acacgild
[acacgild@localhost assignments]$ hadoop fs -cat /assignments/output/TVSalesTask1/part-00000
18/05/29 15:32:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
Samsung|Optima|14|Madhya Pradesh|132401|14200 1
Onida|Lucid|18|Uttar Pradesh|232401|16200 1
Akai|Decent|16|Kerala|922401|12200 1
Lava|Attention|20|Assam|454601|24200 1
Zen|Super|14|Maharashtra|619082|9200 1
Samsung|Optima|14|Madhya Pradesh|132401|14200 1
Onida|Lucid|18|Uttar Pradesh|232401|16200 1
Onida|Decent|14|Uttar Pradesh|232401|16200 1
[acacgild@localhost assignments]$ hadoop fs -cat /assignments/output/TVSalesTask1/part-00001
18/05/29 15:32:30 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
Lava|Attention|20|Assam|454601|24200 1
Zen|Super|14|Maharashtra|619082|9200 1
Samsung|Optima|14|Madhya Pradesh|132401|14200 1
Samsung|Decent|16|Kerala|922401|12200 1
Lava|Attention|20|Assam|454601|24200 1
Samsung|Super|14|Maharashtra|619082|9200 1
Samsung|Super|14|Maharashtra|619082|9200 1
Samsung|Super|14|Maharashtra|619082|9200 1
[acacgild@localhost assignments]$
```

Task 2:

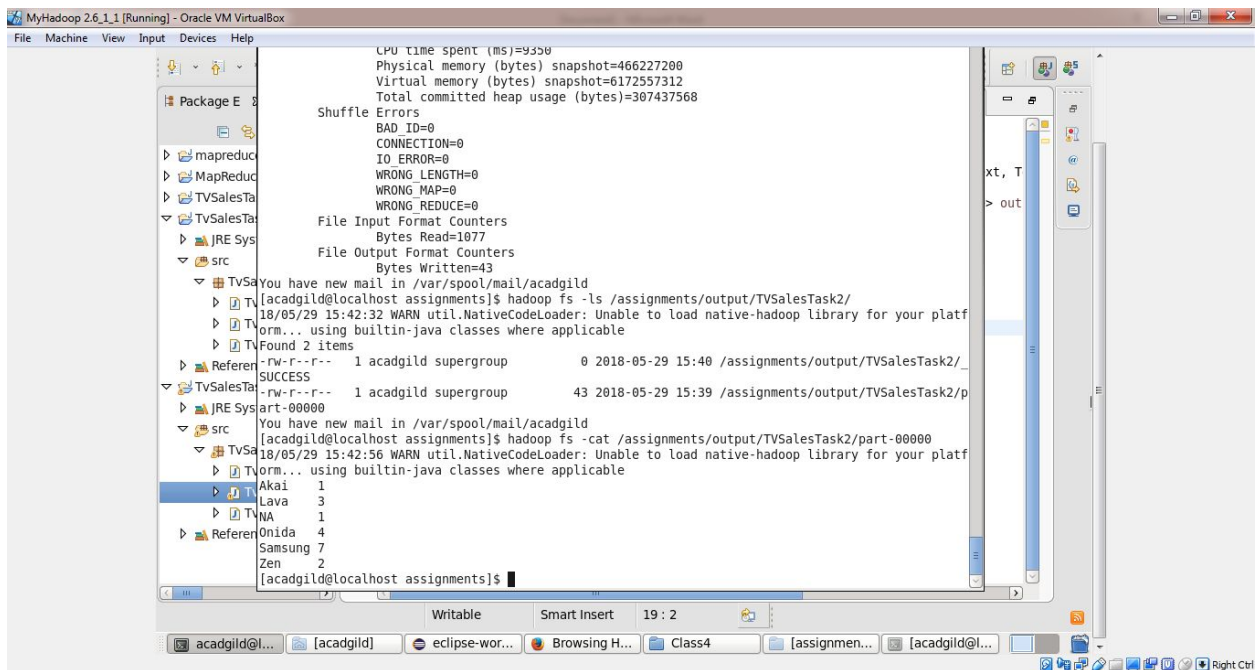
Write a Map Reduce program to calculate the total units sold for each Company.



```
[acadgild@localhost assignments]$ hadoop jar TVSalesTask2.jar TVSalesTask2.TvSalesDriver /assignments/television.txt /assignments/output/TVSalesTask2
18/05/29 15:37:04 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/05/29 15:37:08 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/05/29 15:37:09 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/05/29 15:37:12 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/05/29 15:37:15 INFO mapreduce.FileInputFormat: Total input paths to process : 1
18/05/29 15:37:15 INFO mapreduce.JobSubmitter: number of splits:2
18/05/29 15:37:17 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1527541257921_0042
18/05/29 15:37:18 INFO impl.YarnClientImpl: Submitted application application_1527541257921_0042
18/05/29 15:37:19 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1527541257921_0042/
18/05/29 15:37:19 INFO mapreduce.Job: Running job: job_1527541257921_0042
18/05/29 15:37:55 INFO mapreduce.Job: Job job_1527541257921_0042 running in uber mode : false
18/05/29 15:37:55 INFO mapreduce.Job: map 0% reduce 0%
18/05/29 15:39:09 INFO mapreduce.Job: map 67% reduce 0%
18/05/29 15:39:13 INFO mapreduce.Job: map 100% reduce 0%
18/05/29 15:40:00 INFO mapreduce.Job: map 100% reduce 100%
18/05/29 15:40:06 INFO mapreduce.Job: Job job_1527541257921_0042 completed successfully
18/05/29 15:40:08 INFO mapreduce.Job: Counters: 49

File System Counters
  FILE: Number of bytes read=225
  FILE: Number of bytes written=323432
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=1277
  HDFS: Number of bytes written=43
  HDFS: Number of read operations=9
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2

Job Counters
  MapReduce job mapreduce.counters=49
```



```
CPU time spent (ms)=9350
Physical memory (bytes) snapshot=466227200
Virtual memory (bytes) snapshot=617257312
Total committed heap usage (bytes)=307437568

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=1077

File Output Format Counters
  Bytes Written=43

You have new mail in /var/spool/mail/acadgild
[acadgild@localhost assignments]$ hadoop fs -ls /assignments/output/TVSalesTask2/
18/05/29 15:42:32 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 2018-05-29 15:40 /assignments/output/TVSalesTask2/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 43 2018-05-29 15:39 /assignments/output/TVSalesTask2/part-000000

You have new mail in /var/spool/mail/acadgild
[acadgild@localhost assignments]$ hadoop fs -cat /assignments/output/TVSalesTask2/part-000000
18/05/29 15:42:56 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Akai 1
Lava 3
NA 1
Onida 4
Samsung 7
Zen 2
[acadgild@localhost assignments]$
```

Task 3:

Write a Map Reduce program to calculate the total units sold in each state for Onida company

