Mapreduce

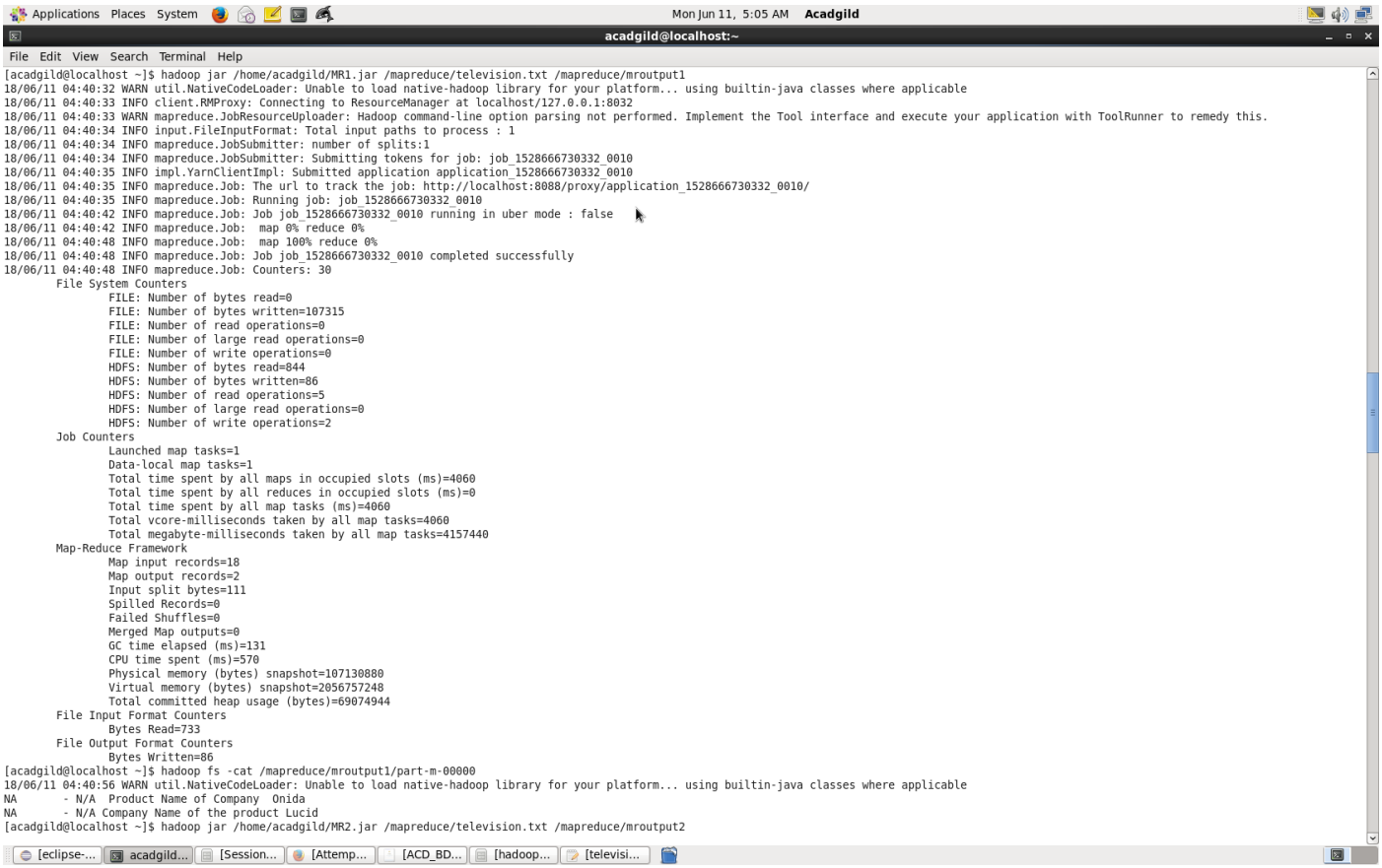
I have written Mapreduce programs according to requirement, mentioned o/p format

& i/p format classes, created jar files & executed as below.

Attaching Java programs & jar files in Github for reference

Task 1:

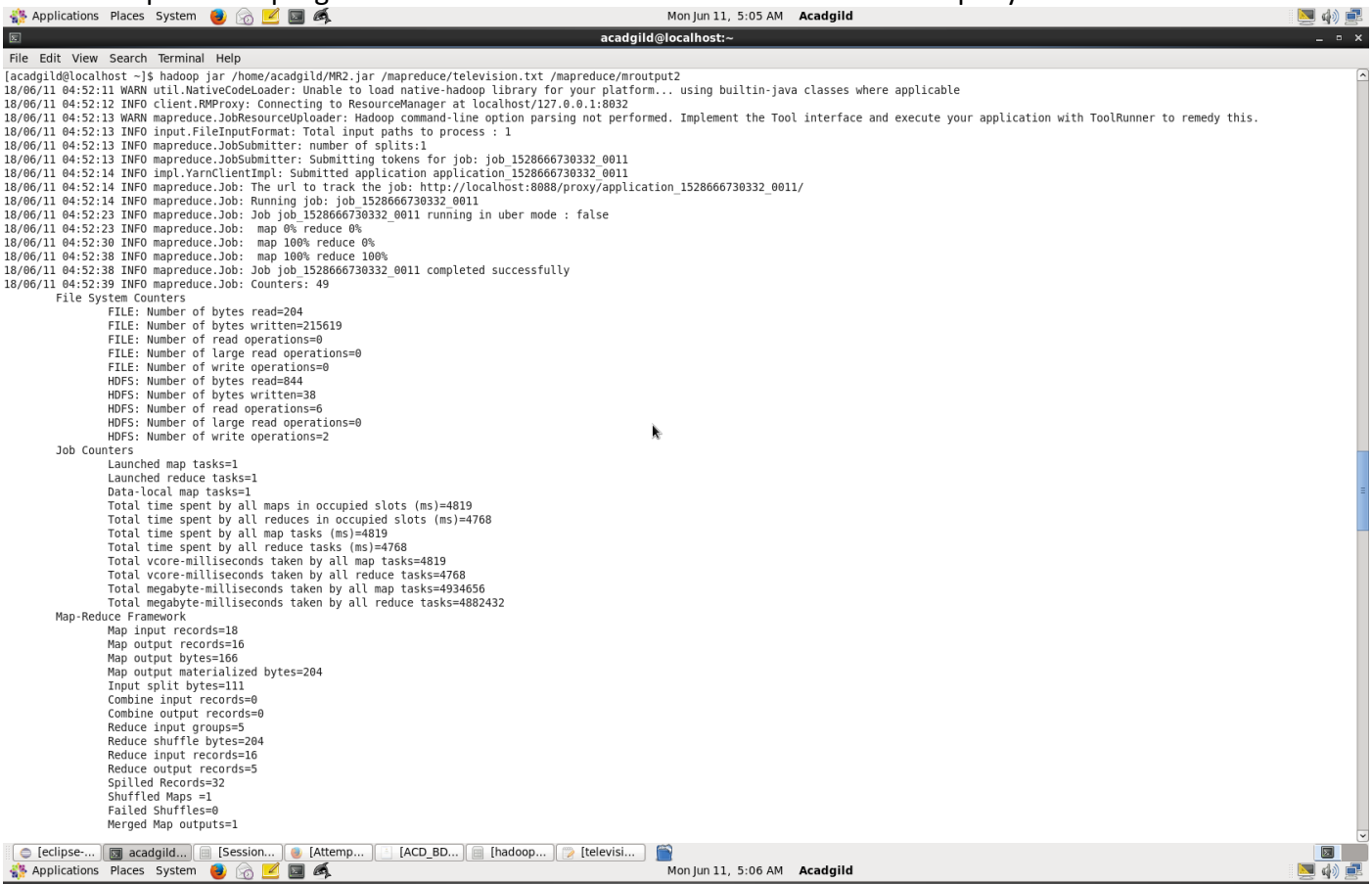
Write a Map Reduce program to filter out the invalid records. Map only job will fit for this context.



Mapreduce

Task 2:

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



```
Applications  Places  System                    Mon Jun 11, 5:05 AM  Acadgild
acadgild@localhost:~
File  Edit  View  Search  Terminal  Help
[acadgild@localhost ~]$ hadoop jar /home/acadgild/MR2.jar /mapreduce/television.txt /mapreduce/mroutput2
18/06/11 04:52:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/06/11 04:52:12 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/06/11 04:52:13 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/06/11 04:52:13 INFO input.FileInputFormat: Total input paths to process : 1
18/06/11 04:52:13 INFO mapreduce.JobSubmitter: number of splits:1
18/06/11 04:52:13 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1528666730332_0011
18/06/11 04:52:14 INFO impl.YarnClientImpl: Submitted application application_1528666730332_0011
18/06/11 04:52:14 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1528666730332_0011/
18/06/11 04:52:14 INFO mapreduce.Job: Running job: job_1528666730332_0011
18/06/11 04:52:23 INFO mapreduce.Job: Job job_1528666730332_0011 running in uber mode : false
18/06/11 04:52:23 INFO mapreduce.Job:  map 0% reduce 0%
18/06/11 04:52:30 INFO mapreduce.Job:  map 100% reduce 0%
18/06/11 04:52:38 INFO mapreduce.Job:  map 100% reduce 100%
18/06/11 04:52:38 INFO mapreduce.Job: Job job_1528666730332_0011 completed successfully
18/06/11 04:52:39 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=204
                FILE: Number of bytes written=215619
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=844
                HDFS: Number of bytes written=38
                HDFS: Number of read operations=6
                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)=4819
                Total time spent by all reduces in occupied slots (ms)=4768
                Total time spent by all map tasks (ms)=4819
                Total time spent by all reduce tasks (ms)=4768
                Total vcore-milliseconds taken by all map tasks=4819
                Total vcore-milliseconds taken by all reduce tasks=4768
                Total megabyte-milliseconds taken by all map tasks=4934656
                Total megabyte-milliseconds taken by all reduce tasks=4882432
        Map-Reduce Framework
                Map input records=18
                Map output records=16
                Map output bytes=166
                Map output materialized bytes=204
                Input split bytes=111
                Combine input records=0
                Combine output records=0
                Reduce input groups=5
                Reduce shuffle bytes=204
                Reduce input records=16
                Reduce output records=5
                Spilled Records=32
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
[eclipse-...] acadgild... [Session...] [Attemp...] [ACD_BD...] [hadoop...] [televisi...]
Applications  Places  System                    Mon Jun 11, 5:06 AM  Acadgild
acadgild@localhost:~
File  Edit  View  Search  Terminal  Help
                Map output materialized bytes=204
                Input split bytes=111
                Combine input records=0
                Combine output records=0
                Reduce input groups=5
                Reduce shuffle bytes=204
                Reduce input records=16
                Reduce output records=5
                Spilled Records=32
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=318
                CPU time spent (ms)=1530
                Physical memory (bytes) snapshot=325165056
                Virtual memory (bytes) snapshot=4118192128
                Total committed heap usage (bytes)=230424576
        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=733
        File Output Format Counters
                Bytes Written=38
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /mapreduce/mroutput2/part-m-00000
18/06/11 04:52:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
cat: `/mapreduce/mroutput2/part-m-00000': No such file or directory
[acadgild@localhost ~]$ hadoop fs -ls /mapreduce
18/06/11 04:53:10 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 7 items
drwxr-xr-x   - acadgild supergroup          0 2018-06-11 04:40 /mapreduce/mroutput1
drwxr-xr-x   - acadgild supergroup          0 2018-06-11 04:52 /mapreduce/mroutput2
-rw-r--r--   1 acadgild supergroup         73 2018-06-03 23:28 /mapreduce/musicdata.txt
drwxr-xr-x   - acadgild supergroup          0 2018-06-04 01:25 /mapreduce/output1
drwxr-xr-x   - acadgild supergroup          0 2018-06-04 01:45 /mapreduce/output2
drwxr-xr-x   - acadgild supergroup          0 2018-06-04 01:46 /mapreduce/output3
-rw-r--r--   1 acadgild supergroup        733 2018-06-11 03:19 /mapreduce/television.txt
[acadgild@localhost ~]$ hadoop fs -ls /mapreduce/mroutput2
18/06/11 04:53:29 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 2018-06-11 04:52 /mapreduce/mroutput2/_SUCCESS
-rw-r--r--   1 acadgild supergroup         38 2018-06-11 04:52 /mapreduce/mroutput2/part-r-00000
[acadgild@localhost ~]$ hadoop fs -cat /mapreduce/mroutput2/part-r-00000
18/06/11 04:53:40 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
You have new mail in /var/spool/mail/acadgild
[eclipse-...] acadgild... [Session...] [Attemp...] [ACD_BD...] [hadoop...] [televisi...]
```

Mapreduce

Task 3:

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