NOTE: All the Map reduce programs related to TASKS 1,2,3 are written and exported the programs as respective jar files. Programs are attached as Zipped files in Github.
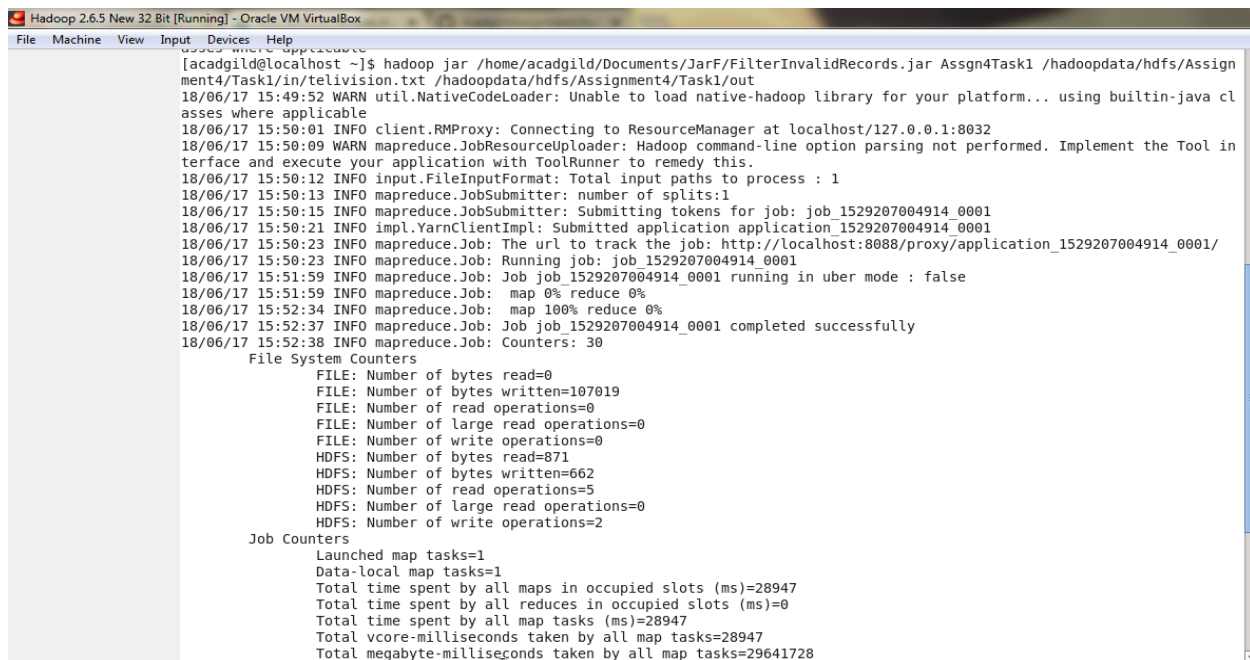
Task 1:

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

```
[acadgild@localhost ~]$ hadoop fs -mkdir /hadoopdata/hdfs/Assignment4/Task1/in
18/06/17 15:44:26 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
[acadgild@localhost ~]$ hadoop fs -put /home/acadgild/Downloads/telivision.txt /hadoopdata/hdfs/Assignment4/Task1/in
18/06/17 15:46:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
[acadgild@localhost ~]$ hadoop fs -mkdir /hadoopdata/hdfs/Assignment4/Task1/out
18/06/17 15:46:59 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
```

-----------JAR FILE-----------------

```
[acadgild@localhost ~]$ hadoop jar /home/acadgild/Documents/JarF/FilterInvalidRecords.jar Assgn4Task1 /hadoopdata/hdfs/Assign
ment4/Task1/in/television.txt /hadoopdata/hdfs/Assignment4/Task1/out
```

```
Hadoop 2.6.5 New 32 Bit [Running] - Oracle VM VirtualBox
File  Machine  View  Input  Devices  Help
asses where applicable
[acadgild@localhost ~]$ hadoop jar /home/acadgild/Documents/JarF/FilterInvalidRecords.jar Assgn4Task1 /hadoopdata/hdfs/Assign
ment4/Task1/in/television.txt /hadoopdata/hdfs/Assignment4/Task1/out
18/06/17 15:49:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/06/17 15:50:01 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/06/17 15:50:09 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool in
terface and execute your application with ToolRunner to remedy this.
18/06/17 15:50:12 INFO input.FileInputFormat: Total input paths to process : 1
18/06/17 15:50:13 INFO mapreduce.JobSubmitter: number of splits:1
18/06/17 15:50:15 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1529207004914_0001
18/06/17 15:50:21 INFO impl.YarnClientImpl: Submitted application application_1529207004914_0001
18/06/17 15:50:23 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1529207004914_0001/
18/06/17 15:50:23 INFO mapreduce.Job: Running job: job_1529207004914_0001
18/06/17 15:51:59 INFO mapreduce.Job: Job job_1529207004914_0001 running in uber mode : false
18/06/17 15:51:59 INFO mapreduce.Job:  map 0% reduce 0%
18/06/17 15:52:34 INFO mapreduce.Job:  map 100% reduce 0%
18/06/17 15:52:37 INFO mapreduce.Job: Job job_1529207004914_0001 completed successfully
18/06/17 15:52:38 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=107019
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=871
                HDFS: Number of bytes written=662
                HDFS: Number of read operations=5
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=28947
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=28947
                Total vcore-milliseconds taken by all map tasks=28947
                Total megabyte-milliseconds taken by all map tasks=29641728
```
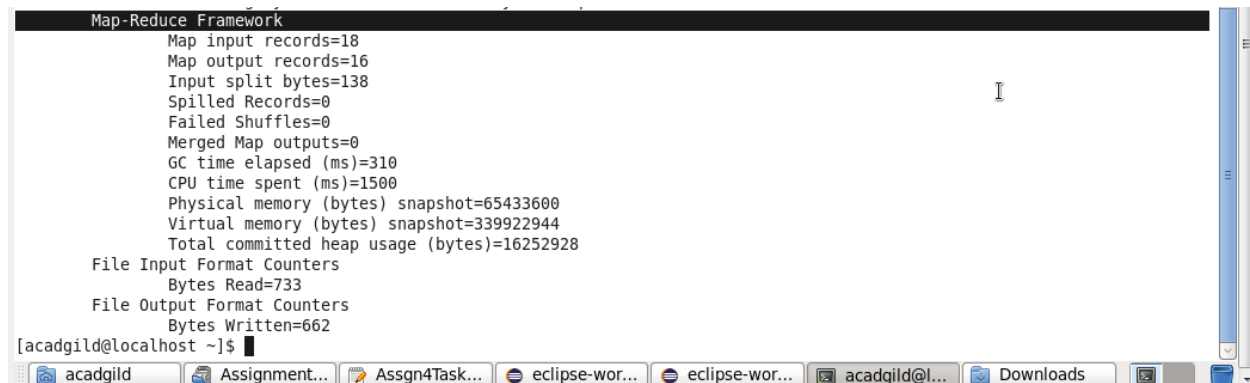
```
        Map-Reduce Framework
                Map input records=18
                Map output records=16
                Input split bytes=138
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=310
                CPU time spent (ms)=1500
                Physical memory (bytes) snapshot=65433600
                Virtual memory (bytes) snapshot=339922944
                Total committed heap usage (bytes)=16252928
        File Input Format Counters
                Bytes Read=733
        File Output Format Counters
                Bytes Written=662
[acadgild@localhost ~]$
```

```
acadgild | Assignment... | Assgn4Task... | eclipse-wor... | eclipse-wor... | acadgild@l... | Downloads
```

OUTPUT: output consists of records which do not have NA.

```
                      Bytes Written=662
[acadgild@localhost ~]$ hadoop fs -ls /hadoopdata/hdfs/Assignment4/Task1/out
18/06/17 16:09:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Found 2 items
-rw-r--r--   1 acadgild supergroup          0 2018-06-17 15:52 /hadoopdata/hdfs/Assignment4/Task1/out/_SUCCESS
-rw-r--r--   1 acadgild supergroup        662 2018-06-17 15:52 /hadoopdata/hdfs/Assignment4/Task1/out/part-m-00000
[acadgild@localhost ~]$ hadoop fs -cat /hadoopdata/hdfs/Assignment4/Task1/out/part-m-00000
18/06/17 16:10:18 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Samsung|Optima|14|Madhya Pradesh|132401|14200
Onida|Lucid|18|Uttar Pradesh|232401|16200
Akai|Decent|16|Kerala|922401|12200
Lava|Attention|20|Assam|454601|24200
Zen|Super|14|Maharashtra|619082|9200
Samsung|Optima|14|Madhya Pradesh|132401|14200
Onida|Lucid|18|Uttar Pradesh|232401|16200
Onida|Decent|14|Uttar Pradesh|232401|16200
Lava|Attention|20|Assam|454601|24200
Zen|Super|14|Maharashtra|619082|9200
Samsung|Optima|14|Madhya Pradesh|132401|14200
Samsung|Decent|16|Kerala|922401|12200
Lava|Attention|20|Assam|454601|24200
Samsung|Super|14|Maharashtra|619082|9200
Samsung|Super|14|Maharashtra|619082|9200
Samsung|Super|14|Maharashtra|619082|9200
```

TASK 2 : Write a Map Reduce program to calculate the total units sold for each Company.

------JAR FILE-----

```
Onida : Uttar Pradesh   5
[acadgild@localhost ~]$ hadoop jar /home/acadgild/Documents/JarF/Task2.jar com.acadgild.Assgn4Task2.Assgn4Task2  /hadoopdata/
hdfs/Assignment4/Task1/in/telivision.txt /hadoopdata/hdfs/Assignment4/Task1/out
```

```
[Running] - Oracle VM VirtualBox
nput  Devices  Help
   Samsung|Super|14|Maharashtra|619082|9200
   [acadgild@localhost ~]$ hadoop jar /home/acadgild/Documents/JarF/FilterInvalidRecords.jar Task1Mapper /hadoopdata/hdfs/Assign
   ment4/Task1/in/telivision.txt /hadoopdata/hdfs/Assignment4/Task1/out
   Exception in thread "main" java.lang.NoSuchMethodException: Task1Mapper.main([Ljava.lang.String;])
           at java.lang.Class.getMethod(Class.java:1786)
           at org.apache.hadoop.util.RunJar.run(RunJar.java:215)
           at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
   [acadgild@localhost ~]$ hadoop jar /home/acadgild/Documents/JarF/Task2.jar com.acadgild.Assgn4Task2.Assgn4Task2  /hadoopdata/
   hdfs/Assignment4/Task1/in/telivision.txt /hadoopdata/hdfs/Assignment4/Task1/out
   18/06/17 16:53:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
   asses where applicable
   18/06/17 16:53:58 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
   18/06/17 16:54:03 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool in
   terface and execute your application with ToolRunner to remedy this.
   18/06/17 16:54:05 INFO input.FileInputFormat: Total input paths to process : 1
   18/06/17 16:54:06 INFO mapreduce.JobSubmitter: number of splits:1
   18/06/17 16:54:07 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1529207004914_0002
   18/06/17 16:54:09 INFO impl.YarnClientImpl: Submitted application application_1529207004914_0002
   18/06/17 16:54:10 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1529207004914_0002/
   18/06/17 16:54:10 INFO mapreduce.Job: Running job: job_1529207004914_0002
   18/06/17 16:54:52 INFO mapreduce.Job: Job job_1529207004914_0002 running in uber mode : false
   18/06/17 16:54:52 INFO mapreduce.Job:  map 0% reduce 0%
   18/06/17 16:55:25 INFO mapreduce.Job:  map 100% reduce 0%
   18/06/17 16:56:02 INFO mapreduce.Job:  map 100% reduce 100%
   18/06/17 16:56:07 INFO mapreduce.Job: Job job_1529207004914_0002 completed successfully
   18/06/17 16:56:08 INFO mapreduce.Job: Counters: 49
           File System Counters
                   FILE: Number of bytes read=225
                   FILE: Number of bytes written=215831
                   FILE: Number of read operations=0
                   FILE: Number of large read operations=0
                   FILE: Number of write operations=0
                   HDFS: Number of bytes read=871
                   HDFS: Number of bytes written=43
                   HDFS: Number of read operations=6
                   HDFS: Number of large read operations=0
                   HDFS: Number of write operations=2
   acadgild    Assignment4.zip    Task2Reducer.j...    eclipse-worksp...    eclipse-worksp...    acadgild@local...
```

```
                        Total megabyte-milliseconds taken by all reduce tasks=37696512
                Map-Reduce Framework
                        Map input records=18
                        Map output records=18
                        Map output bytes=183
                        Map output materialized bytes=225
                        Input split bytes=138
                        Combine input records=0
                        Combine output records=0
                        Reduce input groups=6
                        Reduce shuffle bytes=225
                        Reduce input records=18
                        Reduce output records=6
                        Spilled Records=36
                        Shuffled Maps =1
                        Failed Shuffles=0
                        Merged Map outputs=1
                        GC time elapsed (ms)=983
                        CPU time spent (ms)=4000
                        Physical memory (bytes) snapshot=236191744
                        Virtual memory (bytes) snapshot=680849408
                        Total committed heap usage (bytes)=139776000
                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=43
```

OUTPUT : output consists of total units sold for each Company

```
[acadgild@localhost ~]$ hadoop fs -ls /hadoopdata/hdfs/Assignment4/Task1/out
18/06/17 16:57:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Found 2 items
-rw-r--r--   1 acadgild supergroup          0 2018-06-17 16:56 /hadoopdata/hdfs/Assignment4/Task1/out/_SUCCESS
-rw-r--r--   1 acadgild supergroup         43 2018-06-17 16:56 /hadoopdata/hdfs/Assignment4/Task1/out/part-r-00000
[acadgild@localhost ~]$ hadoop fs -cat /hadoopdata/hdfs/Assignment4/Task1/out/part-r-00000
18/06/17 16:58:23 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Akai    1
Lava    3
NA      1
Onida   4
Samsung 7
Zen     2
[acadgild@localhost ~]$
```

TASK 3: **Write a Map Reduce program to calculate the total units sold in each state for Onida company.**

---JAR FILE----

```
[acadgild@localhost ~]$ hadoop jar /home/acadgild/Documents/JarF/Task3.jar com.acadgild.Assgn4Task3.Assgn4Task3  /hadoopdata/
hdfs/Assignment4/Task1/in/telivision.txt /hadoopdata/hdfs/Assignment4/Task1/out█
```

```
[acadgild@localhost ~]$ hadoop jar /home/acadgild/Documents/JarF/Task3.jar com.acadgild.Assgn4Task3.Assgn4Task3  /hadoopdata/
hdfs/Assignment4/Task1/in/telivision.txt /hadoopdata/hdfs/Assignment4/Task1/out
18/06/17 17:10:05 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/06/17 17:10:12 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/06/17 17:10:17 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool in
terface and execute your application with ToolRunner to remedy this.
18/06/17 17:10:19 INFO input.FileInputFormat: Total input paths to process : 1
18/06/17 17:10:19 INFO mapreduce.JobSubmitter: number of splits:1
18/06/17 17:10:21 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1529207004914_0003
18/06/17 17:10:22 INFO impl.YarnClientImpl: Submitted application application_1529207004914_0003
18/06/17 17:10:23 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1529207004914_0003/
18/06/17 17:10:23 INFO mapreduce.Job: Running job: job_1529207004914_0003
18/06/17 17:11:07 INFO mapreduce.Job: Job job_1529207004914_0003 running in uber mode : false
18/06/17 17:11:07 INFO mapreduce.Job:  map 0% reduce 0%
18/06/17 17:11:38 INFO mapreduce.Job:  map 100% reduce 0%
18/06/17 17:12:10 INFO mapreduce.Job:  map 100% reduce 100%
18/06/17 17:12:12 INFO mapreduce.Job: Job job_1529207004914_0003 completed successfully
18/06/17 17:12:13 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=79
                FILE: Number of bytes written=215539
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=871
                HDFS: Number of bytes written=41
                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)=27215
                Total time spent by all reduces in occupied slots (ms)=28067
```

| 📁 acadgild | 📁 Assignment4.zip | 📄 Task3Reducer.java (... | ◉ eclipse-workspace -... | ▣ acadgild@localhost:~ | ▣ | 📁 |

```
it [Running] - Oracle VM VirtualBox
Input  Devices  Help
                Total vcore-milliseconds taken by all reduce tasks=28067
                Total megabyte-milliseconds taken by all map tasks=27868160
                Total megabyte-milliseconds taken by all reduce tasks=28740608
        Map-Reduce Framework
                Map input records=18
                Map output records=4
                Map output bytes=65
                Map output materialized bytes=79
                Input split bytes=138
                Combine input records=0
                Combine output records=0
                Reduce input groups=2
                Reduce shuffle bytes=79
                Reduce input records=4
                Reduce output records=2
                Spilled Records=8
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=1016
                CPU time spent (ms)=3670
                Physical memory (bytes) snapshot=236302336
                Virtual memory (bytes) snapshot=680697856
                Total committed heap usage (bytes)=139730944
        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=41
```

OUTPUT: **Output consists of total units sold in each state for Onida company.**

```
Bytes written=41
[acadgild@localhost ~]$ hadoop fs -ls /hadoopdata/hdfs/Assignment4/Task1/out
18/06/17 17:14:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Found 2 items
-rw-r--r--   1 acadgild supergroup          0 2018-06-17 17:12 /hadoopdata/hdfs/Assignment4/Task1/out/_SUCCESS
-rw-r--r--   1 acadgild supergroup         41 2018-06-17 17:12 /hadoopdata/hdfs/Assignment4/Task1/out/part-r-00000
[acadgild@localhost ~]$ hadoop fs -cat /hadoopdata/hdfs/Assignment4/Task1/out/part-r-00000
18/06/17 17:18:02 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Onida : Kerala  1
Onida : Uttar Pradesh   3
[acadgild@localhost ~]$
```