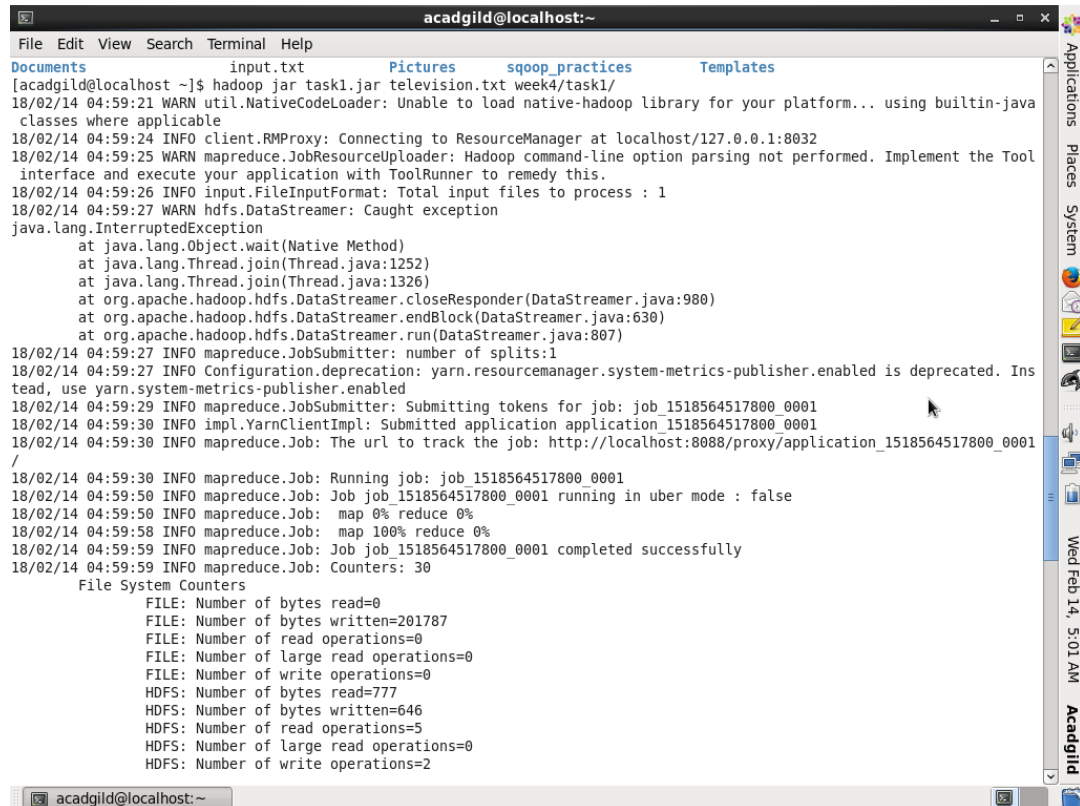


1. Write a Map Reduce program to filter out the invalid records. Map only job will fit for this context.



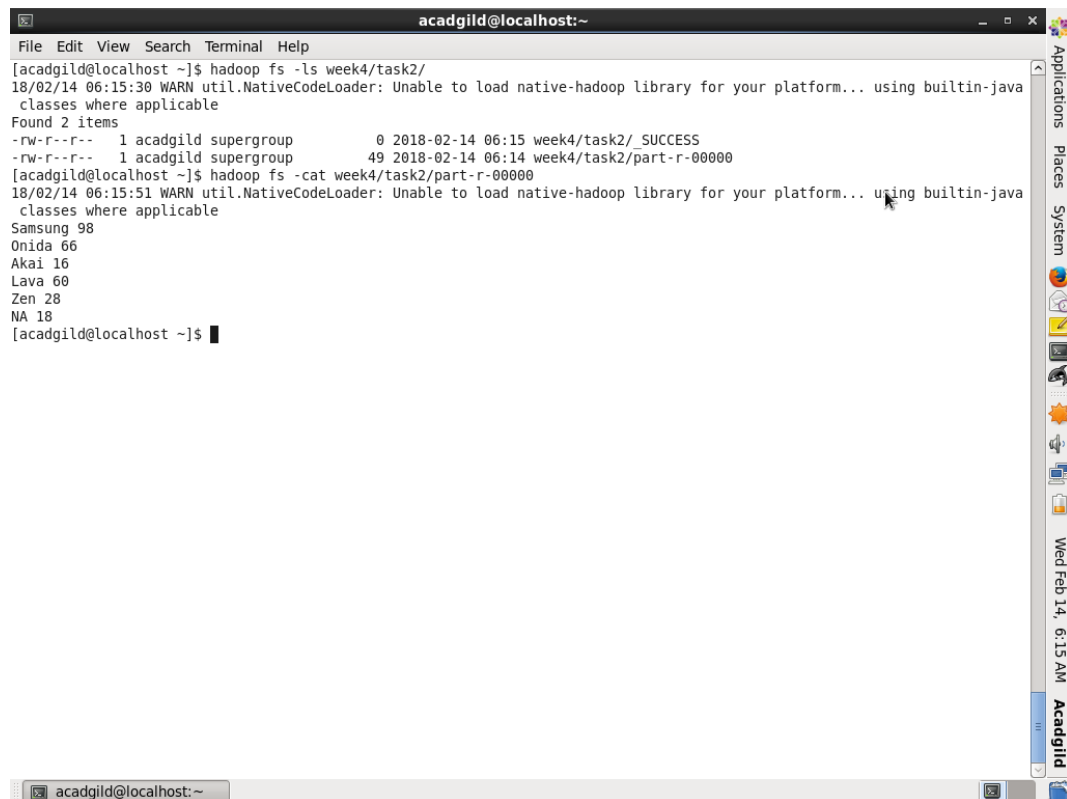
The screenshot shows a terminal window titled 'acadgild@localhost:~' with a menu bar (File, Edit, View, Search, Terminal, Help) and a toolbar. The terminal displays the output of a Hadoop MapReduce job. The logs include warnings about native code loading, information about connecting to the Resource Manager, and the submission of a job. The job is identified as 'job_1518564517800_0001'. The progress shows 0% map and reduce operations, followed by 100% completion. The final output shows the job completed successfully with counters: 30. The File System Counters section lists various metrics such as bytes read/written, read/write operations, and HDFS metrics.

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
Documents input.txt Pictures sqoop_practices Templates  
[acadgild@localhost ~]$ hadoop jar task1.jar television.txt week4/task1/  
18/02/14 04:59:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java  
classes where applicable  
18/02/14 04:59:24 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032  
18/02/14 04:59:25 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool  
interface and execute your application with ToolRunner to remedy this.  
18/02/14 04:59:26 INFO input.FileInputFormat: Total input files to process : 1  
18/02/14 04:59:27 WARN hdfs.DataStreamer: Caught exception  
java.lang.InterruptedException  
    at java.lang.Object.wait(Native Method)  
    at java.lang.Thread.join(Thread.java:1252)  
    at java.lang.Thread.join(Thread.java:1326)  
    at org.apache.hadoop.hdfs.DataStreamer.closeResponder(DataStreamer.java:980)  
    at org.apache.hadoop.hdfs.DataStreamer.endBlock(DataStreamer.java:630)  
    at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:807)  
18/02/14 04:59:27 INFO mapreduce.JobSubmitter: number of splits:1  
18/02/14 04:59:27 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Ins  
tead, use yarn.system-metrics-publisher.enabled  
18/02/14 04:59:29 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1518564517800_0001  
18/02/14 04:59:30 INFO impl.YarnClientImpl: Submitted application application_1518564517800_0001  
18/02/14 04:59:30 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1518564517800_0001  
/  
18/02/14 04:59:30 INFO mapreduce.Job: Running job: job_1518564517800_0001  
18/02/14 04:59:50 INFO mapreduce.Job: Job job_1518564517800_0001 running in uber mode : false  
18/02/14 04:59:50 INFO mapreduce.Job: map 0% reduce 0%  
18/02/14 04:59:58 INFO mapreduce.Job: map 100% reduce 0%  
18/02/14 04:59:59 INFO mapreduce.Job: Job job_1518564517800_0001 completed successfully  
18/02/14 04:59:59 INFO mapreduce.Job: Counters: 30  
    File System Counters  
        FILE: Number of bytes read=0  
        FILE: Number of bytes written=201787  
        FILE: Number of read operations=0  
        FILE: Number of large read operations=0  
        FILE: Number of write operations=0  
        HDFS: Number of bytes read=777  
        HDFS: Number of bytes written=646  
        HDFS: Number of read operations=5  
        HDFS: Number of large read operations=0  
        HDFS: Number of write operations=2
```

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
[acadgild@localhost ~]$ hadoop fs -ls week4/task1/  
18/02/14 05:00:21 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 2018-02-14 04:59 week4/task1/_SUCCESS  
-rw-r--r-- 1 acadgild supergroup 646 2018-02-14 04:59 week4/task1/part-m-000000  
[acadgild@localhost ~]$ hadoop fs -cat week4/task1/part-m-000000  
18/02/14 05:00:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java  
classes 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  
[acadgild@localhost ~]$
```

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

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
[acadgild@localhost ~]$ hadoop jar task2.jar television.txt week4/task2/  
18/02/14 06:09:56 WARN util.NativeCodeLoader: Unable to load native-hadoop libra  
ry for your platform... using builtin-java classes where applicable  
18/02/14 06:09:57 INFO client.RMPProxy: Connecting to ResourceManager at localhos  
t/127.0.0.1:8032  
18/02/14 06:09:59 WARN mapreduce.JobResourceUploader: Hadoop command-line option  
parsing not performed. Implement the Tool interface and execute your applicatio  
n with ToolRunner to remedy this.  
18/02/14 06:09:59 INFO input.FileInputFormat: Total input files to process : 1  
18/02/14 06:09:59 INFO mapreduce.JobSubmitter: number of splits:1  
18/02/14 06:09:59 INFO Configuration.deprecation: yarn.resourcemanager.system-me  
trics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publishe  
r.enabled  
18/02/14 06:10:00 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_15  
18564517800_0012  
18/02/14 06:10:00 INFO impl.YarnClientImpl: Submitted application application_15  
18564517800_0012  
18/02/14 06:10:00 INFO mapreduce.Job: The url to track the job: http://localhost  
:8088/proxy/application_1518564517800_0012/  
18/02/14 06:10:00 INFO mapreduce.Job: Running job: job_1518564517800_0012
```



The screenshot shows a terminal window titled 'acadgild@localhost:~'. The terminal contains the following text:

```
File Edit View Search Terminal Help
[acadgild@localhost ~]$ hadoop fs -ls week4/task2/
18/02/14 06:15:30 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 2018-02-14 06:15 week4/task2/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 49 2018-02-14 06:14 week4/task2/part-r-00000
[acadgild@localhost ~]$ hadoop fs -cat week4/task2/part-r-00000
18/02/14 06:15:51 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
classes where applicable
Samsung 98
Onida 66
Akai 16
Lava 60
Zen 28
NA 18
[acadgild@localhost ~]$
```

The terminal window has a menu bar with 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'. On the right side, there is a vertical toolbar with icons for 'Applications', 'Places', 'System', and a clock showing 'Wed Feb 14, 6:15 AM'. The bottom status bar shows 'acadgild@localhost:~'.

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

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
    at org.apache.hadoop.hdfs.DataStreamer.endBlock(DataStreamer.java:630)  
    at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:807)  
18/02/14 05:19:15 WARN hdfs.DataStreamer: Caught exception  
java.lang.InterruptedException  
    at java.lang.Object.wait(Native Method)  
    at java.lang.Thread.join(Thread.java:1252)  
    at java.lang.Thread.join(Thread.java:1326)  
    at org.apache.hadoop.hdfs.DataStreamer.closeResponder(DataStreamer.java:980)  
    at org.apache.hadoop.hdfs.DataStreamer.endBlock(DataStreamer.java:630)  
    at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:807)  
18/02/14 05:19:15 INFO mapreduce.JobSubmitter: number of splits:1  
18/02/14 05:19:16 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Ins  
tead, use yarn.system-metrics-publisher.enabled  
18/02/14 05:19:16 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1518564517800_0005  
18/02/14 05:19:16 INFO impl.YarnClientImpl: Submitted application application_1518564517800_0005  
18/02/14 05:19:17 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1518564517800_0005  
/  
18/02/14 05:19:17 INFO mapreduce.Job: Running job: job_1518564517800_0005  
18/02/14 05:19:31 INFO mapreduce.Job: Job job_1518564517800_0005 running in uber mode : false  
18/02/14 05:19:31 INFO mapreduce.Job: map 0% reduce 0%  
18/02/14 05:19:43 INFO mapreduce.Job: map 100% reduce 0%  
18/02/14 05:19:59 INFO mapreduce.Job: map 100% reduce 100%  
18/02/14 05:20:00 INFO mapreduce.Job: Job job_1518564517800_0005 completed successfully  
18/02/14 05:20:01 INFO mapreduce.Job: Counters: 49  
File System Counters  
    FILE: Number of bytes read=6  
    FILE: Number of bytes written=404259  
    FILE: Number of read operations=0  
    FILE: Number of large read operations=0  
    FILE: Number of write operations=0  
    HDFS: Number of bytes read=777  
    HDFS: Number of bytes written=0  
    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)=9777
```

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
[acadgild@localhost ~]$ hadoop fs -cat part-r-00000  
18/02/14 06:12:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java  
classes where applicable  
Uttar Pradesh 18  
Uttar Pradesh 18  
Uttar Pradesh 14  
Kerala 16  
[acadgild@localhost ~]$
```