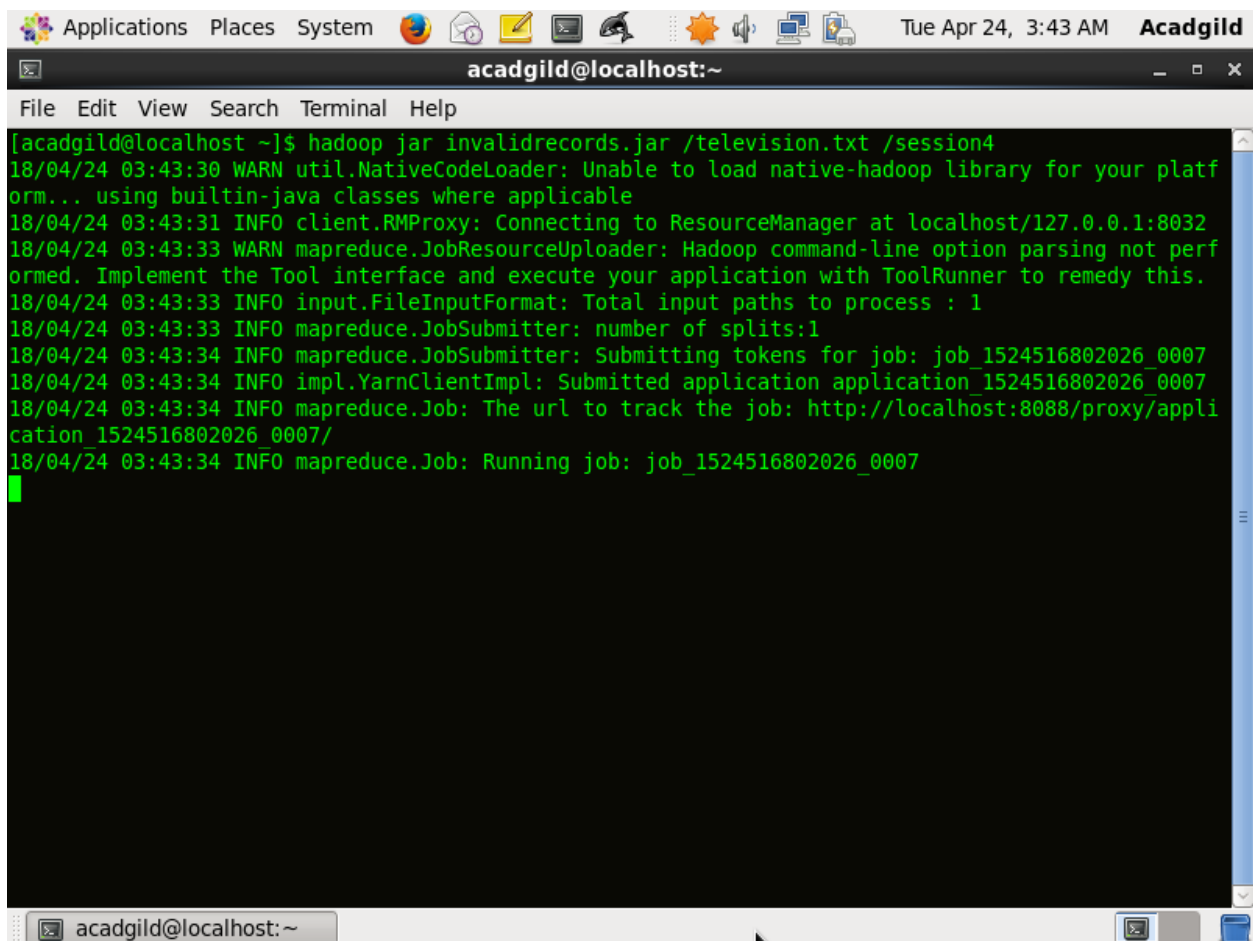


Running job to filter out invalid records. Invalid records will contain the string "NA" in company or product name fields. The following command runs the jar invalidrecords.jar using television.txt stored in hdfs and stores the output in /session4 folder in hdfs.

```
hadoop jar invalidrecords.jar /television.txt /session4
```



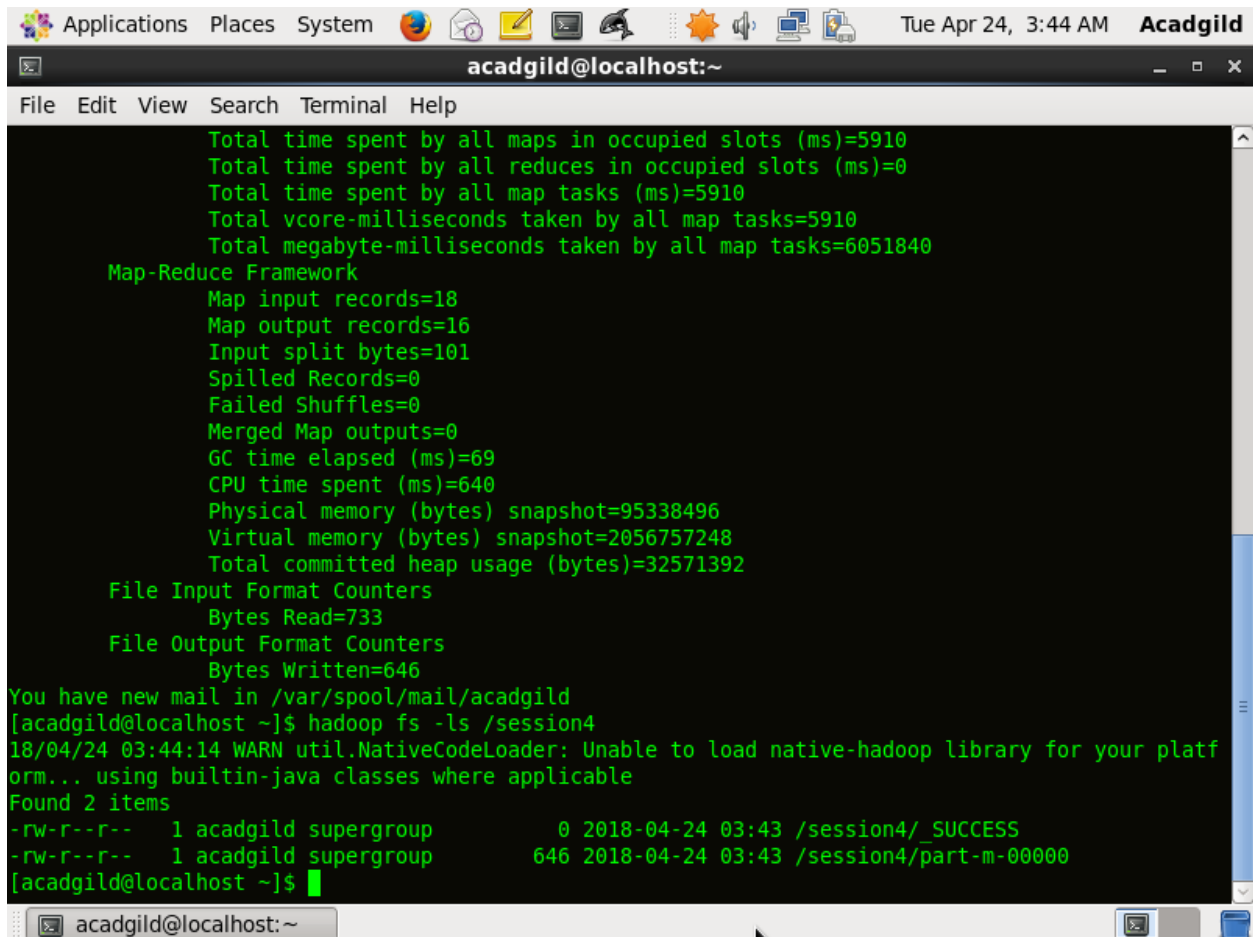
The screenshot shows a terminal window titled 'acadgild@localhost:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal displays the output of the command 'hadoop jar invalidrecords.jar /television.txt /session4'. The output includes several log messages from Hadoop, such as warnings about native code loading and information about connecting to the ResourceManager, submitting the job, and the job's progress. The job ID is 'job_1524516802026_0007'.

```
[acadgild@localhost ~]$ hadoop jar invalidrecords.jar /television.txt /session4
18/04/24 03:43:30 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
18/04/24 03:43:31 INFO client.RMPProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/04/24 03:43:33 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not perf
ormed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/04/24 03:43:33 INFO input.FileInputFormat: Total input paths to process : 1
18/04/24 03:43:33 INFO mapreduce.JobSubmitter: number of splits:1
18/04/24 03:43:34 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1524516802026_0007
18/04/24 03:43:34 INFO impl.YarnClientImpl: Submitted application application_1524516802026_0007
18/04/24 03:43:34 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/appli
cation_1524516802026_0007/
18/04/24 03:43:34 INFO mapreduce.Job: Running job: job_1524516802026_0007
```

Run the command

```
hadoop fs -ls /session4
```

To view the mapping output in session4 folder



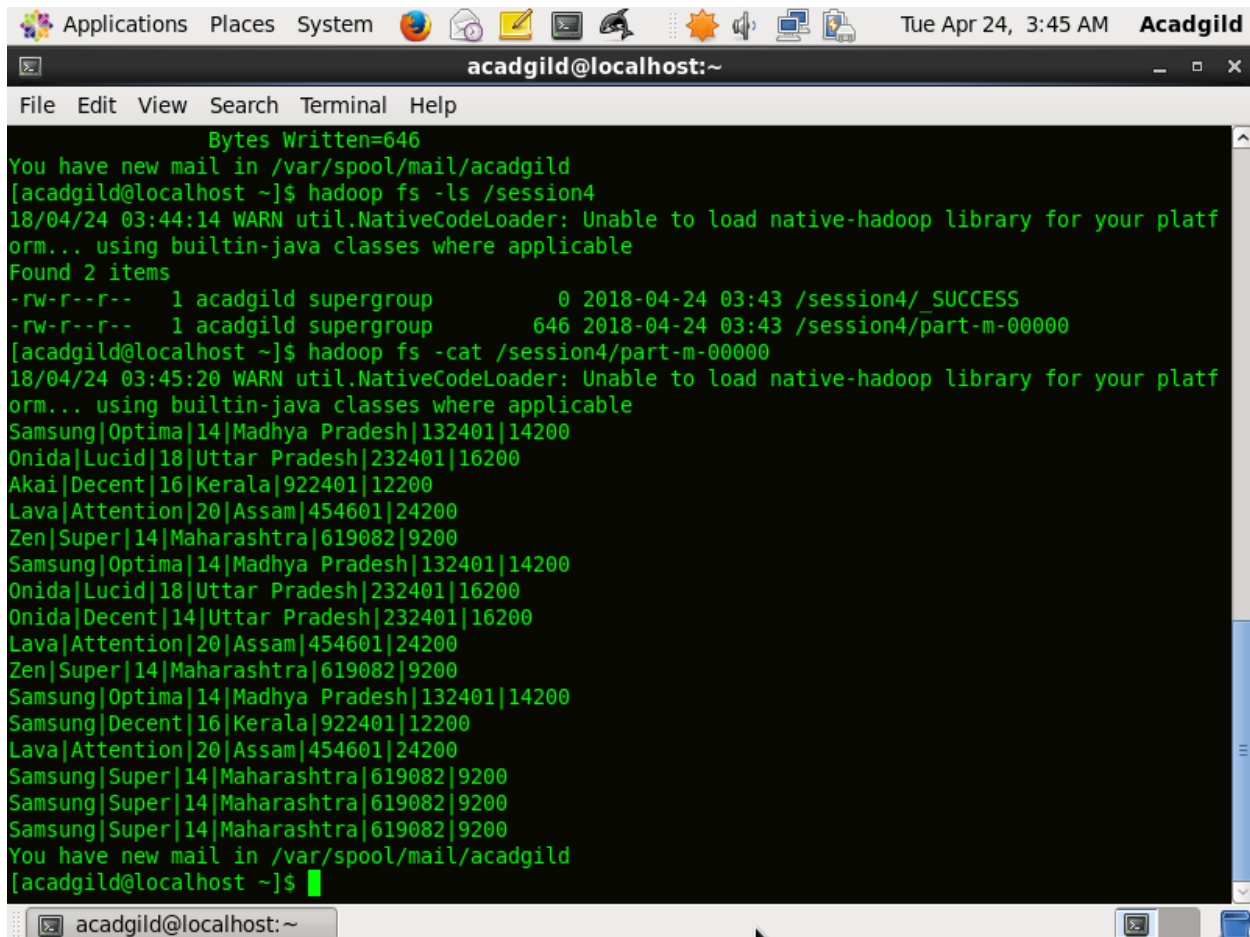
The screenshot shows a terminal window titled 'acadgild@localhost:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal displays the output of a Hadoop command, followed by a file listing command. The output includes statistics for map and reduce tasks, map-reduce framework details, and file input/output counters. A warning message about the native Hadoop library is also shown. The file listing command shows two files in the /session4 directory.

```
Total time spent by all maps in occupied slots (ms)=5910
Total time spent by all reduces in occupied slots (ms)=0
Total time spent by all map tasks (ms)=5910
Total vcore-milliseconds taken by all map tasks=5910
Total megabyte-milliseconds taken by all map tasks=6051840
Map-Reduce Framework
  Map input records=18
  Map output records=16
  Input split bytes=101
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=69
  CPU time spent (ms)=640
  Physical memory (bytes) snapshot=95338496
  Virtual memory (bytes) snapshot=2056757248
  Total committed heap usage (bytes)=32571392
File Input Format Counters
  Bytes Read=733
File Output Format Counters
  Bytes Written=646
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -ls /session4
18/04/24 03:44:14 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
Found 2 items
-rw-r--r--  1 acadgild supergroup          0 2018-04-24 03:43 /session4/_SUCCESS
-rw-r--r--  1 acadgild supergroup       646 2018-04-24 03:43 /session4/part-m-00000
[acadgild@localhost ~]$
```

To view the mapped output run the command

```
hadoop fs -cat /session4/part-m-00000
```

As you can see none of the records have any invalid fields



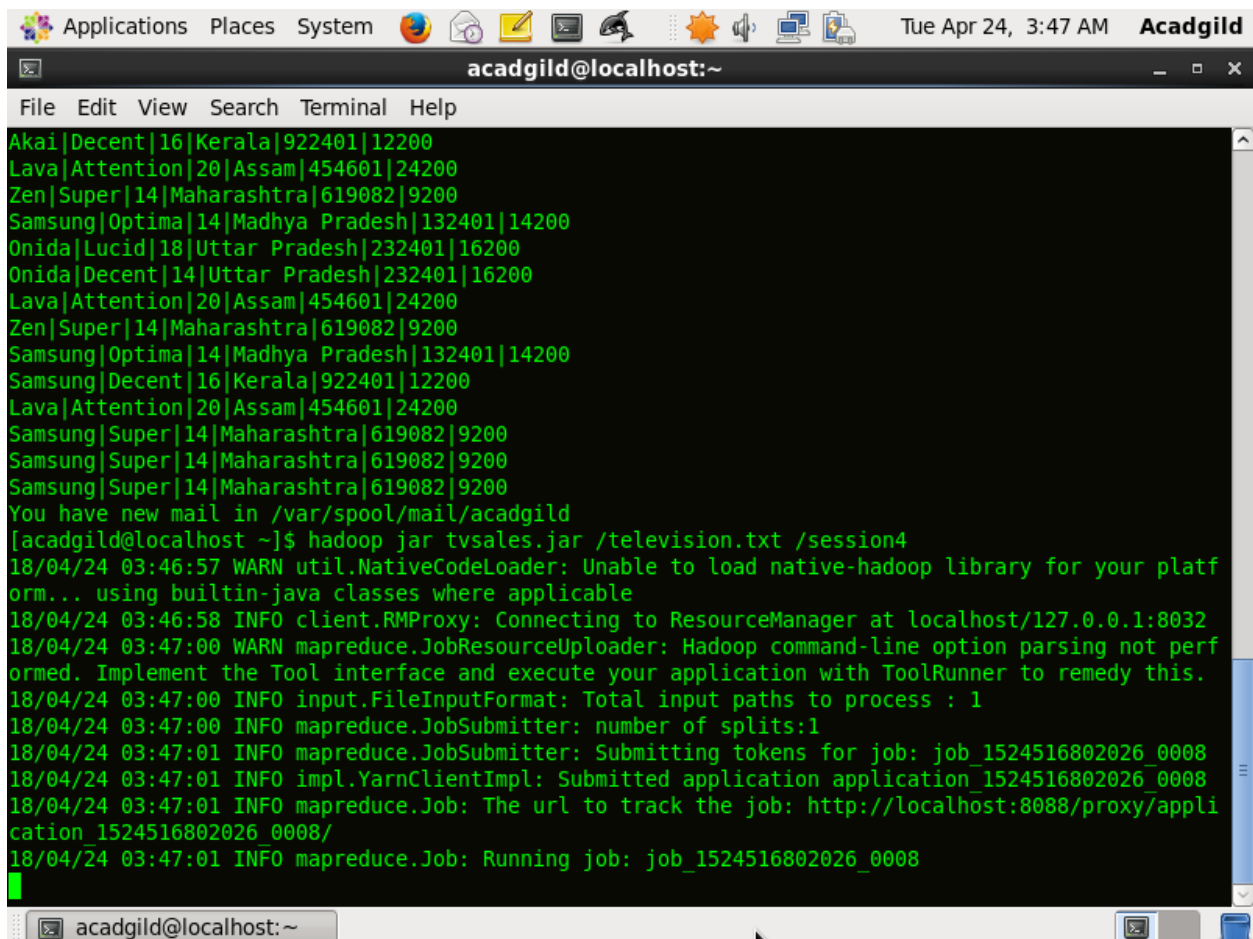
The screenshot shows a terminal window titled 'acadgild@localhost:~'. The window has a menu bar with 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'. The terminal output is as follows:

```
Bytes Written=646
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -ls /session4
18/04/24 03:44:14 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
Found 2 items
-rw-r--r--  1 acadgild supergroup          0 2018-04-24 03:43 /session4/_SUCCESS
-rw-r--r--  1 acadgild supergroup        646 2018-04-24 03:43 /session4/part-m-00000
[acadgild@localhost ~]$ hadoop fs -cat /session4/part-m-00000
18/04/24 03:45:20 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
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
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

To see the sales of tv unit of each company, I am running the jar tvsales.jar

Running the command

```
hadoop jar tvsales.jar /television.txt /session4
```



```
acadgild@localhost:~  
File Edit View Search Terminal Help  
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  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost ~]$ hadoop jar tvsales.jar /television.txt /session4  
18/04/24 03:46:57 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf  
orm... using builtin-java classes where applicable  
18/04/24 03:46:58 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032  
18/04/24 03:47:00 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not perf  
ormed. Implement the Tool interface and execute your application with ToolRunner to remedy this.  
18/04/24 03:47:00 INFO input.FileInputFormat: Total input paths to process : 1  
18/04/24 03:47:00 INFO mapreduce.JobSubmitter: number of splits:1  
18/04/24 03:47:01 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1524516802026_0008  
18/04/24 03:47:01 INFO impl.YarnClientImpl: Submitted application application_1524516802026_0008  
18/04/24 03:47:01 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/appli  
cation_1524516802026_0008/  
18/04/24 03:47:01 INFO mapreduce.Job: Running job: job_1524516802026_0008
```

Running the command to check the contents of the output folder

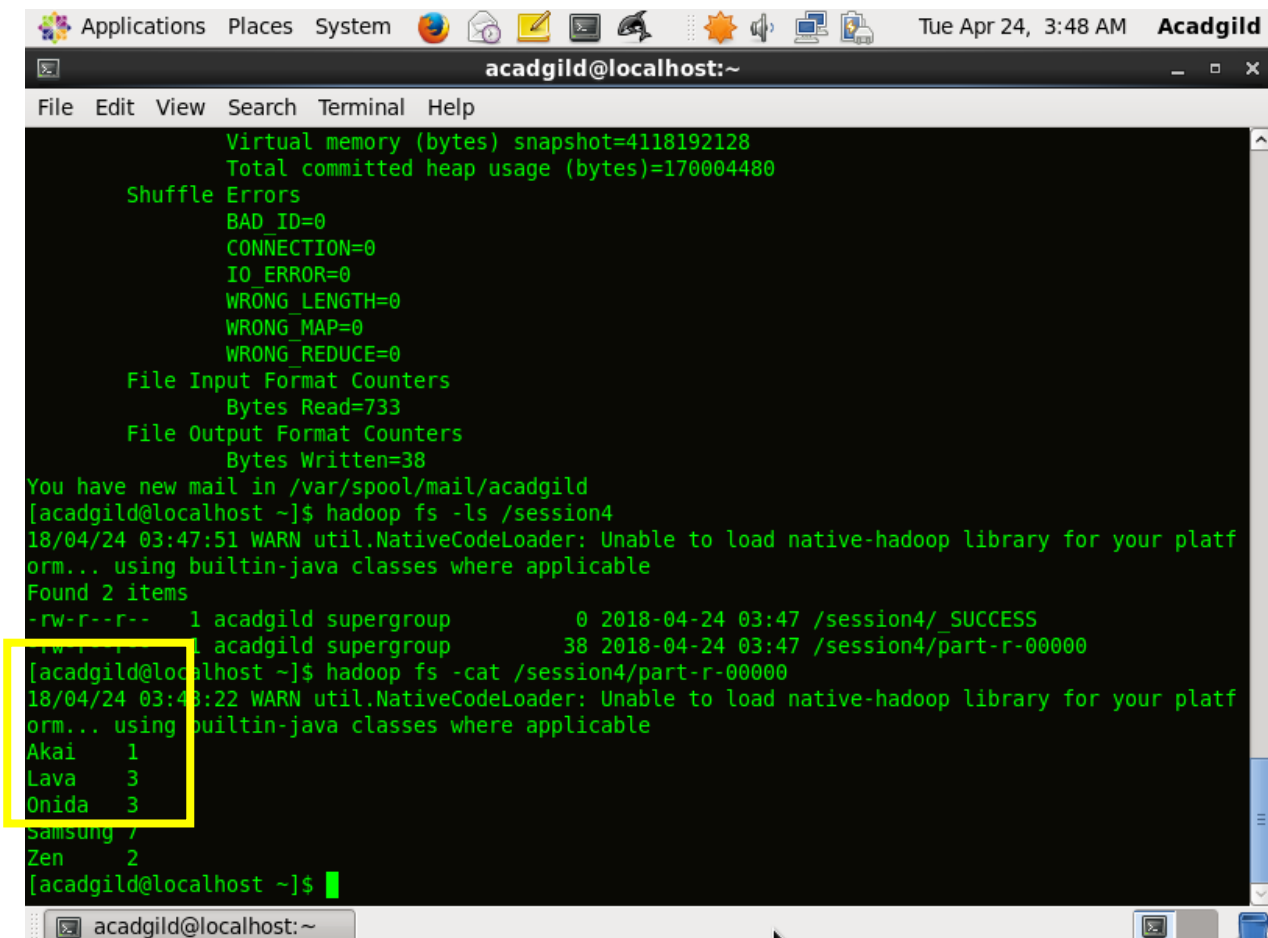
```
hadoop fs -ls /session4
```

Then running the command to view the mapreduce output

```
hadoop fs -cat /session4/part-r-00000
```

As one can see the output has the company names and the number of their tv sold. In the program the mapper code filters the record and maps the company name and digit 1 which is then picked up by the reducer to count the number of occurrences of each key and write it as the output.

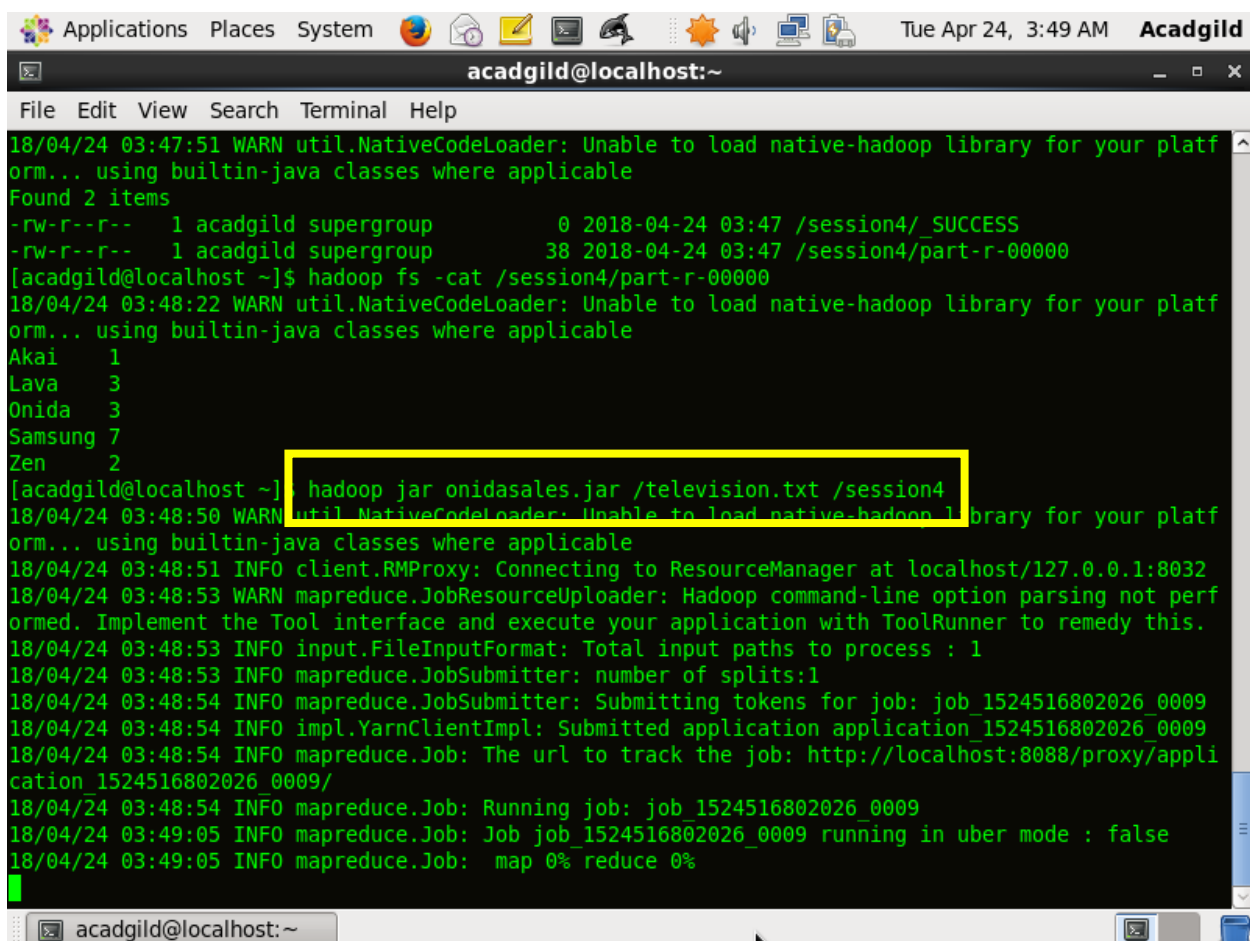
S



```
acadgild@localhost:~$ hadoop fs -ls /session4
18/04/24 03:47:51 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup 0 2018-04-24 03:47 /session4/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 38 2018-04-24 03:47 /session4/part-r-00000
[acadgild@localhost ~]$ hadoop fs -cat /session4/part-r-00000
18/04/24 03:48:22 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
Akai 1
Lava 3
Onida 3
Samsung 7
Zen 2
[acadgild@localhost ~]$
```

The third task is accomplished by running the jar onidasales.jar which gives the mapreduce output of only Onida company state-wise. To achieve the result, filtered records were checked if company name is onida and the state field of the corresponding record is mapped and given to the reducer to count the number of each unique state key and write its sum.

The command is `hadoop jar onidasales.jar /televisionsales.jar /session4`

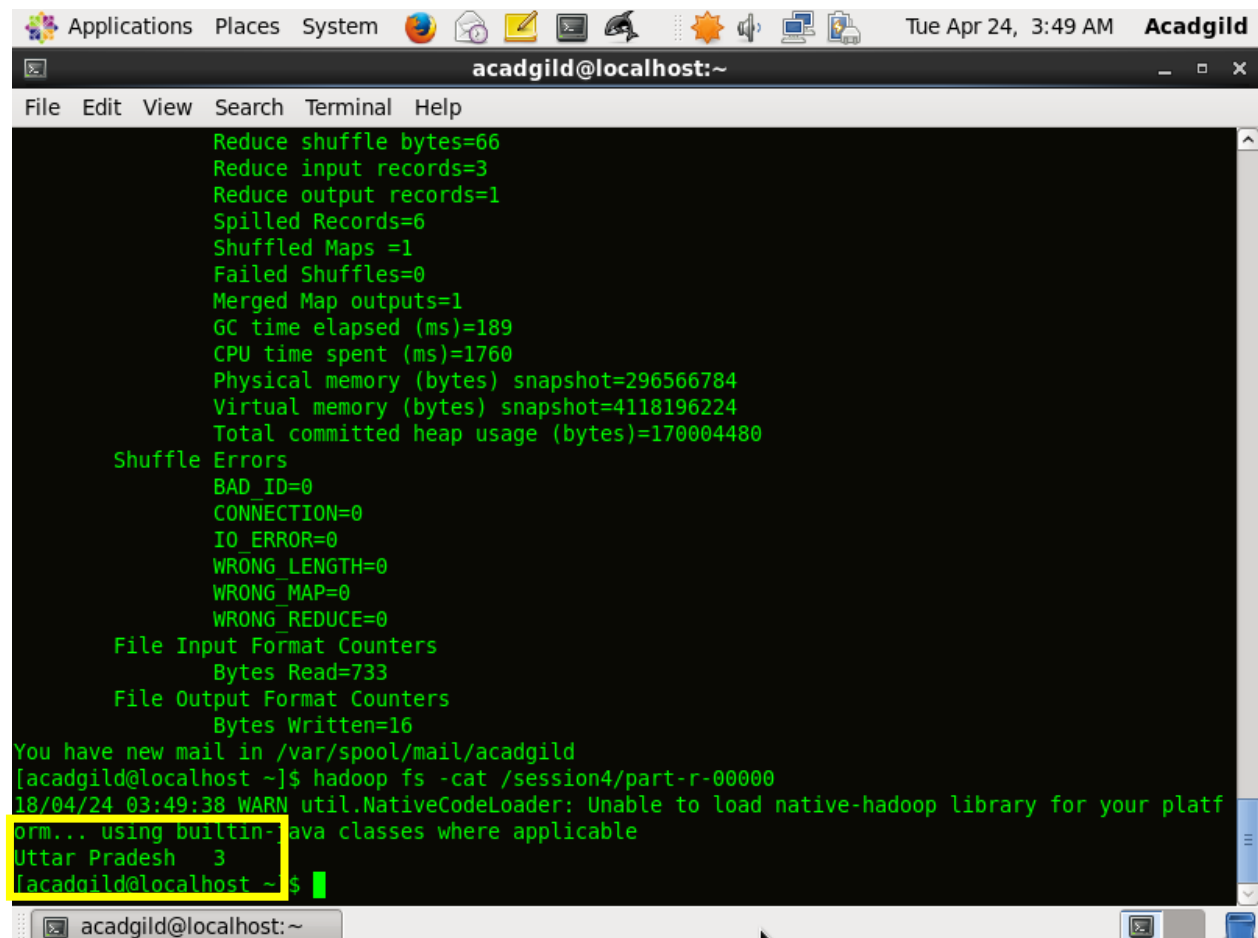


The screenshot shows a terminal window titled 'Acadgild' with the prompt 'acadgild@localhost:~'. The terminal displays the output of a Hadoop job. It starts with a warning about the native-hadoop library, followed by a file listing for '/session4/part-r-00000'. The listing shows two items: 'Akai' with a count of 1 and 'Lava' with a count of 3. The user then runs the command `hadoop jar onidasales.jar /television.txt /session4`, which is highlighted with a yellow box. The output continues with more warnings and information about the job submission and execution, including the job ID 'job_1524516802026_0009' and the status 'map 0% reduce 0%'.

```
18/04/24 03:47:51 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
Found 2 items
-rw-r--r--  1 acadgild supergroup          0 2018-04-24 03:47 /session4/_SUCCESS
-rw-r--r--  1 acadgild supergroup        38 2018-04-24 03:47 /session4/part-r-00000
[acadgild@localhost ~]$ hadoop fs -cat /session4/part-r-00000
18/04/24 03:48:22 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
Akai      1
Lava      3
Onida     3
Samsung   7
Zen       2
[acadgild@localhost ~]$ hadoop jar onidasales.jar /television.txt /session4
18/04/24 03:48:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
18/04/24 03:48:51 INFO client.RMPProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/04/24 03:48:53 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not perf
ormed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/04/24 03:48:53 INFO input.FileInputFormat: Total input paths to process : 1
18/04/24 03:48:53 INFO mapreduce.JobSubmitter: number of splits:1
18/04/24 03:48:54 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1524516802026_0009
18/04/24 03:48:54 INFO impl.YarnClientImpl: Submitted application application_1524516802026_0009
18/04/24 03:48:54 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/appli
cation_1524516802026_0009/
18/04/24 03:48:54 INFO mapreduce.Job: Running job: job_1524516802026_0009
18/04/24 03:49:05 INFO mapreduce.Job: Job job_1524516802026_0009 running in uber mode : false
18/04/24 03:49:05 INFO mapreduce.Job: map 0% reduce 0%
```

Since Onida was sold only in Uttar Pradesh but three models were sold so one can see the desired output in the part-r-00000 file in session4 folder

If one notices, an invalid record of onida from state kerala was also filtered successfully.



The screenshot shows a terminal window titled 'acadgild@localhost:~'. The window displays the output of a Hadoop job, including statistics for shuffle and map operations. Below the job output, a system message indicates new mail in the user's spool. The user then executes the command 'hadoop fs -cat /session4/part-r-00000', which outputs the contents of the file. The output shows a warning from 'util.NativeCodeLoader' and the text 'orm... using builtin- java classes where applicable' followed by 'Uttar Pradesh 3'. The terminal window has a menu bar with 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'. The top status bar shows 'Tue Apr 24, 3:49 AM' and 'Acadgild'.

```
Reduce shuffle bytes=66
Reduce input records=3
Reduce output records=1
Spilled Records=6
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=189
CPU time spent (ms)=1760
Physical memory (bytes) snapshot=296566784
Virtual memory (bytes) snapshot=4118196224
Total committed heap usage (bytes)=170004480

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=16

You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /session4/part-r-00000
18/04/24 03:49:38 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin- java classes where applicable
Uttar Pradesh 3
[acadgild@localhost ~]$
```

