

## Assignment Submission- Session 4

We have a dataset of sales of different TV sets across different locations.

Records look like:

Samsung|Optima|14|Madhya Pradesh|132401|14200

The fields are arranged like:

Company Name|Product Name|Size in inches|State|Pin Code|Price

There are some invalid records which contain 'NA' in either Company Name or Product Name.

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

**Solution:** For this, a jar file has been created by the name "TV\_Sales\_Task1.jar" . It can be run as shown in the screenshot.

```
[acadgild@localhost ~]$ hadoop jar akshat/MyMapReducePrograms/TV_Sales_Task1.jar /television.txt /TvSalesTask1 Output
18/08/26 03:37:35 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/08/26 03:37:36 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/08/26 03:37:37 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your
```

The input file -cat screenshot and output file after running job is also in the screenshot.

```
[acadgild@localhost ~]$ hadoop fs -cat /television.txt
18/08/26 03:42:41 WARN util.NativeCodeLoader: Unable to load native-hadoop lib
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
Onida|NA|16|Kerala|922401|12200
Lava|Attention|20|Assam|454601|24200
Zen|Super|14|Maharashtra|619082|9200
Samsung|Optima|14|Madhya Pradesh|132401|14200
NA|Lucid|18|Uttar Pradesh|232401|16200
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 ~]$
```

```
you have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /TvSalesTask1_Output/part-m-00000
18/08/26 03:43:40 WARN util.NativeCodeLoader: Unable to load native-hadoop librar
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 ~]$
```

### Task 2:

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

**Solution:** For this, a jar file has been created by the name “TV\_Sales\_Task2.jar” . It can be run as shown in the screenshot.

```
acadgild@localhost ~]$ hadoop jar akshat/MyMapReducePrograms/TV_Sales_Task2.jar /television.txt /TvSalesTask2_Output
18/08/26 22:40:25 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/08/26 22:40:27 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/08/26 22:40:29 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your a
18/08/26 22:40:30 INFO input.FileInputFormat: Total input paths to process : 1
```

The output file -cat screenshot is below:

There were 18 entries in the input out of which, 1 record had “NA” in the company name. Hence reducer shows details combined of 17 entries.

```
[acadgild@localhost ~]$ hadoop fs -cat /TvSalesTask2_Output/part-r-00000
18/08/26 22:43:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
Akai      1
Lava      3
Onida     4
Samsung   7
Zen       2
```

### Task 3:

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

**Solution:** For this, a jar file has been created by the name “TV\_Sales\_Task3.jar” . It can be run as shown in the screenshot.

```
acadgild@localhost ~]$ hadoop jar akshat/MyMapReducePrograms/TV_Sales_Task3.jar /television.txt /TvSalesTask3_Output
18/08/27 23:48:36 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where appl
18/08/27 23:48:39 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/08/27 23:48:41 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and exec
18/08/27 23:48:42 INFO input.FileInputFormat: Total input paths to process : 1
18/08/27 23:48:42 INFO mapreduce.JobSubmitter: number of splits:1
```

The output file -cat screenshot is below:

There were 4 entries for Onida out of which 3 sales were in Uttar Pradesh and 1 in Kerala.

```
acadgild@localhost ~]$ hadoop fs -ls /TvSalesTask3_Output
18/08/27 23:51:21 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 2018-08-27 23:49 /TvSalesTask3_Output/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 55 2018-08-27 23:49 /TvSalesTask3_Output/part-r-00000
you have new mail in /var/spool/mail/acadgild
acadgild@localhost ~]$ hadoop fs -cat /TvSalesTask3_Output/part-r-00000
18/08/27 23:51:43 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Onida sales in Kerala 1
Onida sales in Uttar Pradesh 3
```