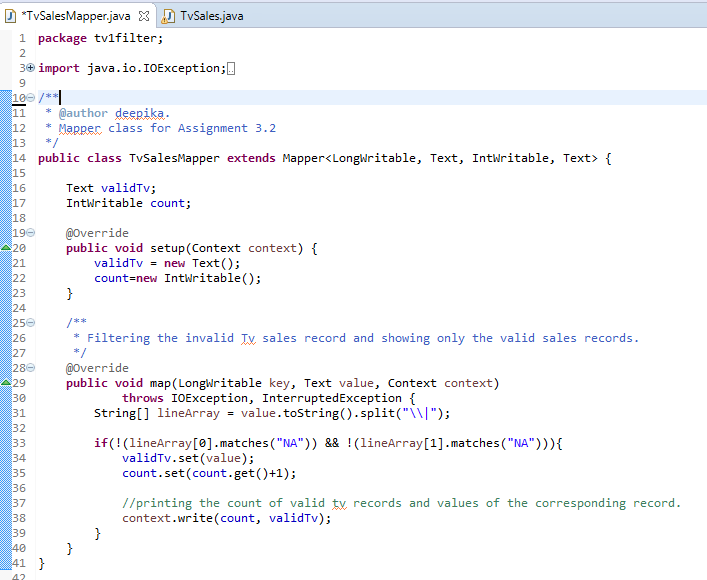
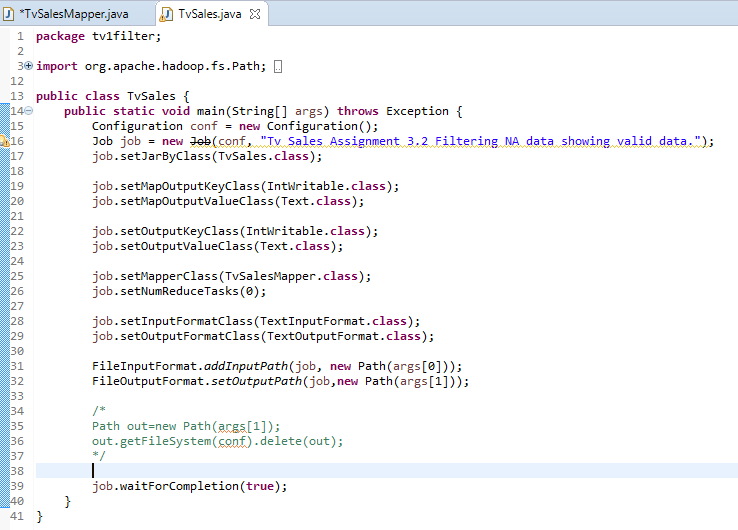
**Task : Write a Map Reduce program to filter out the invalid records. Map only job will fit for this context**

**Create a MapReduce project in Eclipse**

1. For this I’ve created the Maven project. The project package used for this task is tv1filter. This task is done by Mapper alone .



1. Once the mapper class is created based on the logic, added the Job configurations and packaged as jar file.



1. The jar file is copied to local file system of Acadgild SandBox using MobaXterm.
2. This jar file is used for processing the input and the output is stored in HDFS

* mapreduce-0.0.1-SNAPSHOT.jar 🡪 Maven MapReduce project jar
* tv1filter.TvSales 🡪 The Job Class file (which contains the configurations related to mapper).
* /user/acadgild/hadoop/tv\_sample.txt 🡪 Input File which will be used in MapReduce process
* /user/acadgild/hadoop/tv\_filter\_NA 🡪 Location where the MapReduce ouput will be stored in HDFS .

[acadgild@localhost hadoop]$ hadoop jar mapreduce-0.0.1-SNAPSHOT.jar tv1filter.TvSales /user/acadgild/hadoop/tv\_sample.txt /user/acadgild/hadoop/tv\_filter\_NA

17/09/28 16:21:27 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032

17/09/28 16:21:28 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.

17/09/28 16:21:30 INFO input.FileInputFormat: Total input paths to process : 1

17/09/28 16:21:30 INFO mapreduce.JobSubmitter: number of splits:1

17/09/28 16:21:30 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1506595794414\_0001

17/09/28 16:21:31 INFO impl.YarnClientImpl: Submitted application application\_1506595794414\_0001

17/09/28 16:21:31 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application\_1506595794414\_0001/

17/09/28 16:21:31 INFO mapreduce.Job: Running job: job\_1506595794414\_0001

17/09/28 16:21:48 INFO mapreduce.Job: Job job\_1506595794414\_0001 running in uber mode : false

17/09/28 16:21:48 INFO mapreduce.Job: map 0% reduce 0%

17/09/28 16:21:56 INFO mapreduce.Job: map 100% reduce 0%

17/09/28 16:21:56 INFO mapreduce.Job: Job job\_1506595794414\_0001 completed successfully

17/09/28 16:21:56 INFO mapreduce.Job: Counters: 30

File System Counters

FILE: Number of bytes read=0

FILE: Number of bytes written=106127

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=854

HDFS: Number of bytes written=685

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)=5294

Total time spent by all reduces in occupied slots (ms)=0

Total time spent by all map tasks (ms)=5294

Total vcore-seconds taken by all map tasks=5294

Total megabyte-seconds taken by all map tasks=5421056

Map-Reduce Framework

Map input records=18

Map output records=16

Input split bytes=121

Spilled Records=0

Failed Shuffles=0

Merged Map outputs=0

GC time elapsed (ms)=87

CPU time spent (ms)=540

Physical memory (bytes) snapshot=103415808

Virtual memory (bytes) snapshot=2055667712

Total committed heap usage (bytes)=60882944

File Input Format Counters

Bytes Read=733

File Output Format Counters

Bytes Written=685

**Navigating inside the HDFS MapReduce output folder to check the status of MapReduce process**

[acadgild@localhost hadoop]$ hadoop fs -ls /user/acadgild/hadoop/tv\_filter\_NA

Found 2 items

-rw-r--r-- 1 acadgild supergroup 0 2017-09-28 16:21 /user/acadgild/hadoop/tv\_filter\_NA/\_SUCCESS

-rw-r--r-- 1 acadgild supergroup 685 2017-09-28 16:21 /user/acadgild/hadoop/tv\_filter\_NA/part-m-00000

**To view the contents of output File**

[acadgild@localhost hadoop]$ hadoop fs -cat /user/acadgild/hadoop/tv\_filter\_NA/part-m-00000

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

2 Onida|Lucid|18|Uttar Pradesh|232401|16200

3 Akai|Decent|16|Kerala|922401|12200

4 Lava|Attention|20|Assam|454601|24200

5 Zen|Super|14|Maharashtra|619082|9200

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

7 Onida|Lucid|18|Uttar Pradesh|232401|16200

8 Onida|Decent|14|Uttar Pradesh|232401|16200

9 Lava|Attention|20|Assam|454601|24200

10 Zen|Super|14|Maharashtra|619082|9200

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

12 Samsung|Decent|16|Kerala|922401|12200

13 Lava|Attention|20|Assam|454601|24200

14 Samsung|Super|14|Maharashtra|619082|9200

15 Samsung|Super|14|Maharashtra|619082|9200

16 Samsung|Super|14|Maharashtra|619082|9200

**Comparing the Input and Output Data of MapReduce**. (Filtered the data contains NA and shows only valid data in the output**)**

|  |  |
| --- | --- |
| **Input** | **Output** |
| 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 | 1 Samsung|Optima|14|Madhya Pradesh|132401|14200  2 Onida|Lucid|18|Uttar Pradesh|232401|16200  3 Akai|Decent|16|Kerala|922401|12200  4 Lava|Attention|20|Assam|454601|24200  5 Zen|Super|14|Maharashtra|619082|9200  6 Samsung|Optima|14|Madhya Pradesh|132401|14200  7 Onida|Lucid|18|Uttar Pradesh|232401|16200  8 Onida|Decent|14|Uttar Pradesh|232401|16200  9 Lava|Attention|20|Assam|454601|24200  10 Zen|Super|14|Maharashtra|619082|9200  11 Samsung|Optima|14|Madhya Pradesh|132401|14200  12 Samsung|Decent|16|Kerala|922401|12200  13 Lava|Attention|20|Assam|454601|24200  14 Samsung|Super|14|Maharashtra|619082|9200  15 Samsung|Super|14|Maharashtra|619082|9200  16 Samsung|Super|14|Maharashtra|619082|9200 |