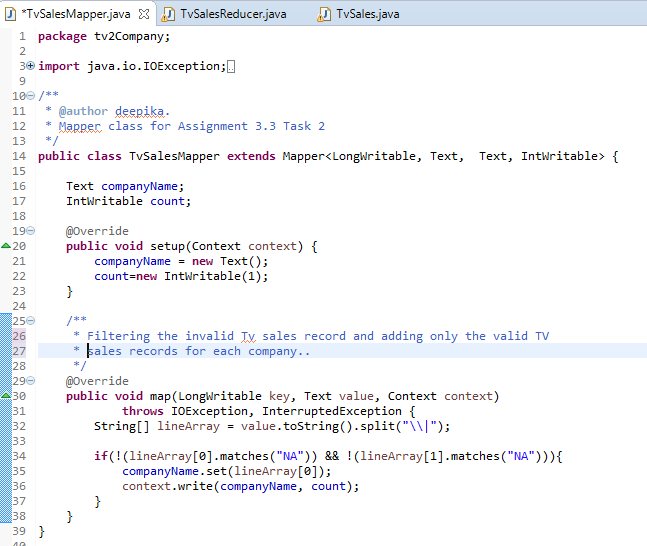
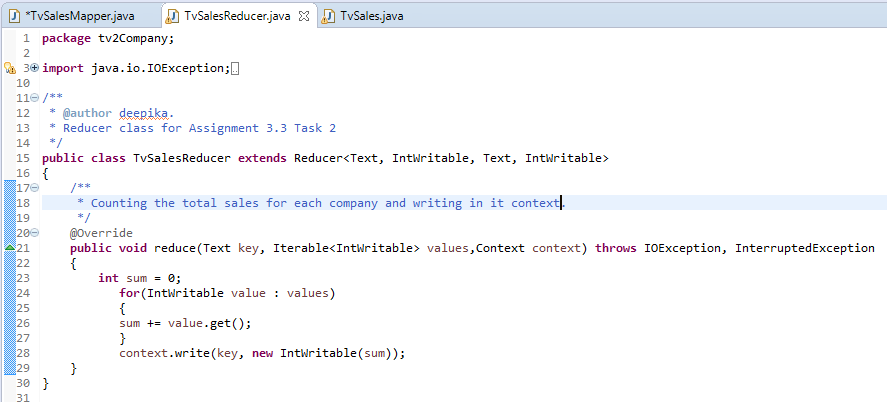
## Task 2 : Write a Map Reduce program to calculate the total units sold for each Company.

**Create a MapReduce project in Eclipse**

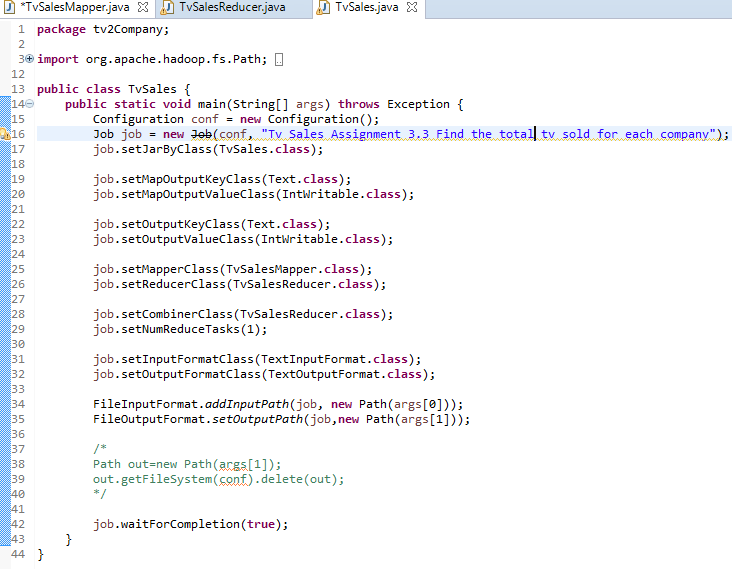
1. For this, I’ve a new package in the maven project. The project package used for this task is tv2Company.
2. This mapper used for this is as below



1. Reducer class calculates the total tv sets sold for each company.



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
* tv2Company.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\_sold\_each\_company 🡪 Location where the MapReduce ouput will be stored in HDFS .

[acadgild@localhost hadoop]$ hadoop jar mapreduce-0.0.1-SNAPSHOT.jar tv2Company.TvSales /user/acadgild/hadoop/tv\_sample.txt /user/acadgild/hadoop/tv\_sold\_each\_company

17/09/28 16:59:39 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

17/09/28 16:59:40 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032

17/09/28 16:59:41 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 16:59:42 INFO input.FileInputFormat: Total input paths to process : 1

17/09/28 16:59:42 INFO mapreduce.JobSubmitter: number of splits:1

17/09/28 16:59:42 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1506595794414\_0003

17/09/28 16:59:42 INFO impl.YarnClientImpl: Submitted application application\_1506595794414\_0003

17/09/28 16:59:42 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application\_1506595794414\_0003/

17/09/28 16:59:42 INFO mapreduce.Job: Running job: job\_1506595794414\_0003

17/09/28 16:59:52 INFO mapreduce.Job: Job job\_1506595794414\_0003 running in uber mode : false

17/09/28 16:59:52 INFO mapreduce.Job: map 0% reduce 0%

17/09/28 16:59:58 INFO mapreduce.Job: map 100% reduce 0%

17/09/28 17:00:05 INFO mapreduce.Job: map 100% reduce 100%

17/09/28 17:00:06 INFO mapreduce.Job: Job job\_1506595794414\_0003 completed successfully

17/09/28 17:00:06 INFO mapreduce.Job: Counters: 49

File System Counters

FILE: Number of bytes read=64

FILE: Number of bytes written=213311

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=38

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)=4378

Total time spent by all reduces in occupied slots (ms)=4932

Total time spent by all map tasks (ms)=4378

Total time spent by all reduce tasks (ms)=4932

Total vcore-seconds taken by all map tasks=4378

Total vcore-seconds taken by all reduce tasks=4932

Total megabyte-seconds taken by all map tasks=4483072

Total megabyte-seconds taken by all reduce tasks=5050368

Map-Reduce Framework

Map input records=18

Map output records=16

Map output bytes=166

Map output materialized bytes=64

Input split bytes=121

Combine input records=16

Combine output records=5

Reduce input groups=5

Reduce shuffle bytes=64

Reduce input records=5

Reduce output records=5

Spilled Records=10

Shuffled Maps =1

Failed Shuffles=0

Merged Map outputs=1

GC time elapsed (ms)=119

CPU time spent (ms)=1150

Physical memory (bytes) snapshot=315506688

Virtual memory (bytes) snapshot=4113727488

Total committed heap usage (bytes)=219676672

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=38

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

[acadgild@localhost hadoop]$ hadoop fs -ls /user/acadgild/hadoop/tv\_sold\_each\_company

Found 2 items

-rw-r--r-- 1 acadgild supergroup 0 2017-09-28 17:00 /user/acadgild/hadoop/tv\_sold\_each\_company/\_SUCCESS

-rw-r--r-- 1 acadgild supergroup 38 2017-09-28 17:00 /user/acadgild/hadoop/tv\_sold\_each\_company/part-r-00000

**To view the contents of output File**

[acadgild@localhost hadoop]$ hadoop fs -cat /user/acadgild/hadoop/tv\_sold\_each\_company/part-r-00000

Akai 1

Lava 3

Onida 3

Samsung 7

Zen 2

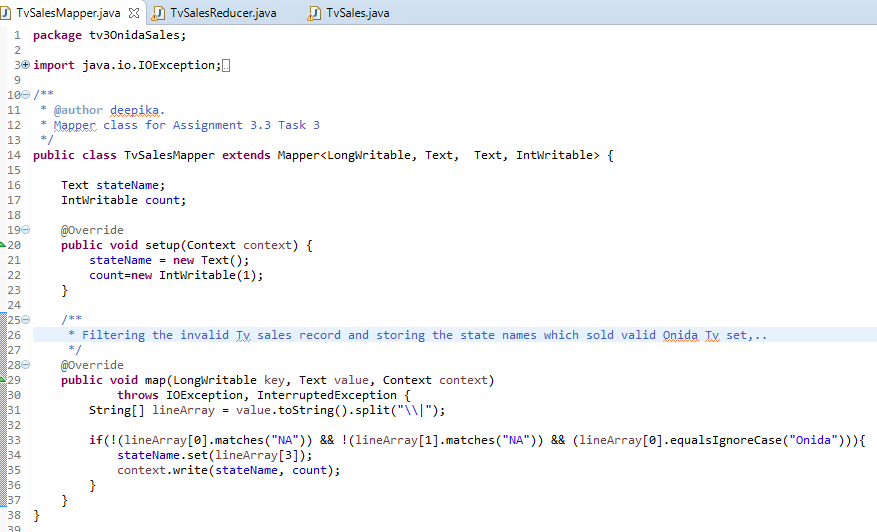
**Comparing the Input and Output Data of MapReduce**. (After filtering the data containing NA, counting the valid sets sold for each company**)**

|  |  |
| --- | --- |
| **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 | Akai 1  Lava 3  Onida 3  Samsung 7  Zen 2 |

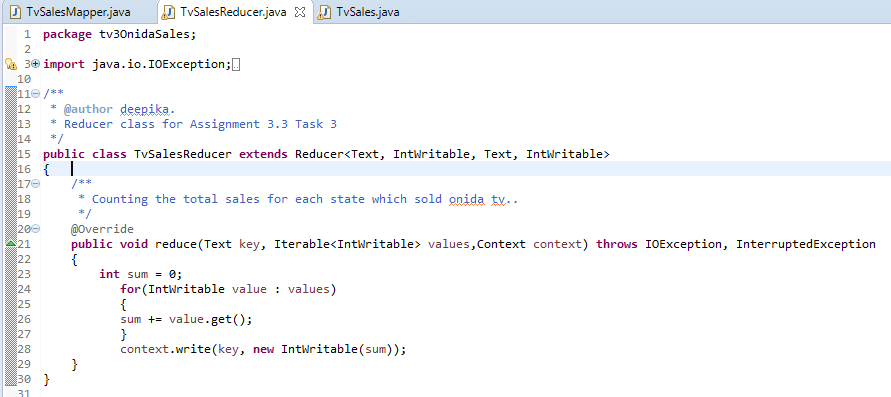
## Task 3 : Write a Map Reduce program to calculate the total units sold in each state for Onida company.

**Create a MapReduce project in Eclipse**

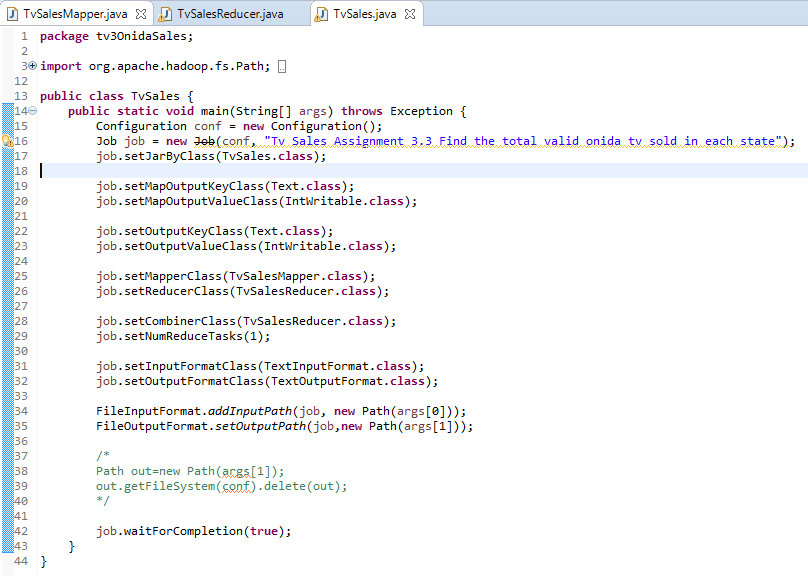
1. For this, I’ve a new package in the maven project. The project package used for this task is tv3OnidaSales.
2. This mapper used for this is as below



1. Reducer class calculates the total tv sets sold for each company.



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
* tv3OnidaSales.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\_sold\_onida\_state 🡪 Location where the MapReduce ouput will be stored in HDFS .

[acadgild@localhost hadoop]$ hadoop jar mapreduce-0.0.1-SNAPSHOT.jar tv3OnidaSales.TvSales /user/acadgild/hadoop/tv\_sample.txt /user/acadgild/hadoop/tv\_sold\_onida\_state

17/09/28 17:25:34 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

17/09/28 17:25:35 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032

17/09/28 17:25:36 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 17:25:36 INFO input.FileInputFormat: Total input paths to process : 1

17/09/28 17:25:36 INFO mapreduce.JobSubmitter: number of splits:1

17/09/28 17:25:37 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1506595794414\_0004

17/09/28 17:25:37 INFO impl.YarnClientImpl: Submitted application application\_1506595794414\_0004

17/09/28 17:25:37 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application\_1506595794414\_0004/

17/09/28 17:25:37 INFO mapreduce.Job: Running job: job\_1506595794414\_0004

17/09/28 17:25:46 INFO mapreduce.Job: Job job\_1506595794414\_0004 running in uber mode : false

17/09/28 17:25:46 INFO mapreduce.Job: map 0% reduce 0%

17/09/28 17:25:53 INFO mapreduce.Job: map 100% reduce 0%

17/09/28 17:26:00 INFO mapreduce.Job: map 100% reduce 100%

17/09/28 17:26:01 INFO mapreduce.Job: Job job\_1506595794414\_0004 completed successfully

17/09/28 17:26:01 INFO mapreduce.Job: Counters: 49

File System Counters

FILE: Number of bytes read=26

FILE: Number of bytes written=213309

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=16

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)=5147

Total time spent by all reduces in occupied slots (ms)=4732

Total time spent by all map tasks (ms)=5147

Total time spent by all reduce tasks (ms)=4732

Total vcore-seconds taken by all map tasks=5147

Total vcore-seconds taken by all reduce tasks=4732

Total megabyte-seconds taken by all map tasks=5270528

Total megabyte-seconds taken by all reduce tasks=4845568

Map-Reduce Framework

Map input records=18

Map output records=3

Map output bytes=54

Map output materialized bytes=26

Input split bytes=121

Combine input records=3

Combine output records=1

Reduce input groups=1

Reduce shuffle bytes=26

Reduce input records=1

Reduce output records=1

Spilled Records=2

Shuffled Maps =1

Failed Shuffles=0

Merged Map outputs=1

GC time elapsed (ms)=136

CPU time spent (ms)=1350

Physical memory (bytes) snapshot=315514880

Virtual memory (bytes) snapshot=4113739776

Total committed heap usage (bytes)=219676672

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

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

[acadgild@localhost hadoop]$ hadoop fs -ls /user/acadgild/hadoop/tv\_sold\_onida\_state

Found 2 items

-rw-r--r-- 1 acadgild supergroup 0 2017-09-28 17:25 /user/acadgild/hadoop/tv\_sold\_onida\_state/\_SUCCESS

-rw-r--r-- 1 acadgild supergroup 16 2017-09-28 17:25 /user/acadgild/hadoop/tv\_sold\_onida\_state/part-r-00000

**To view the contents of output File**

[acadgild@localhost hadoop]$ hadoop fs -cat /user/acadgild/hadoop/tv\_sold\_onida\_state/part-r-00000

Uttar Pradesh 3

**Comparing the Input and Output Data of MapReduce**. (After filtering the data containing NA, counting the valid Onida sets sold for each state**)**

|  |  |
| --- | --- |
| **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 | Uttar Pradesh 3 |