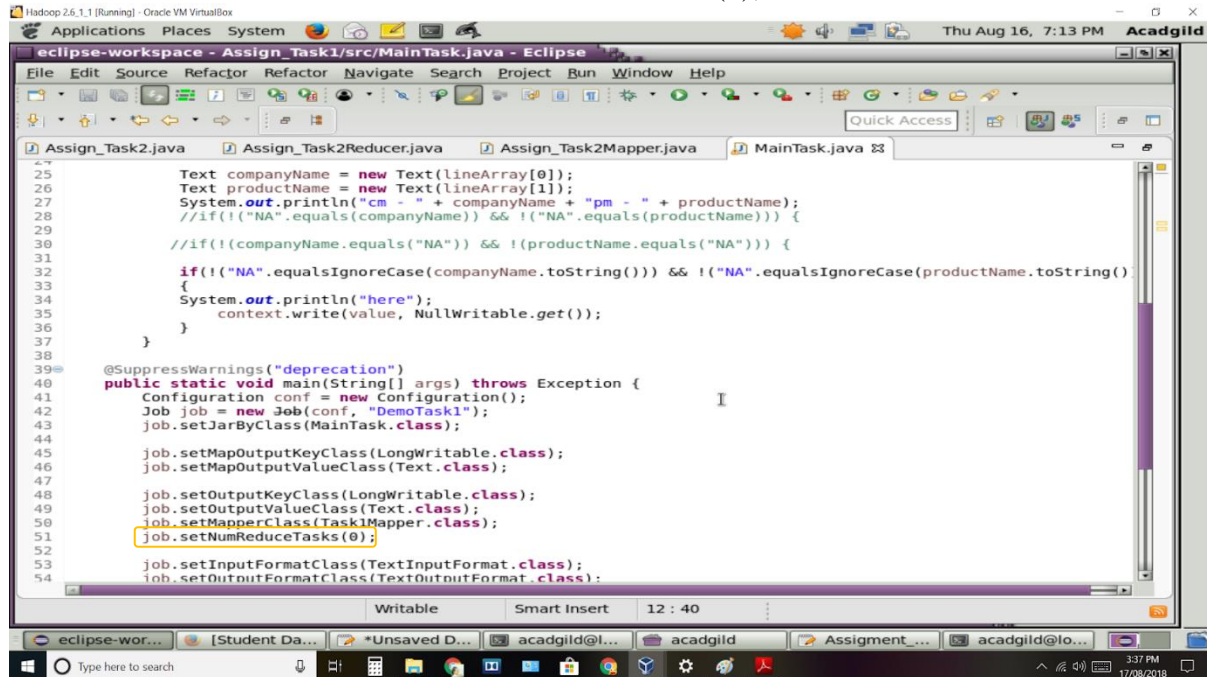


## Assignment 4

### 1) Task 1:

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

- --map only is set by setting
  - `Job.setNumReduceTasks(0);`



```
25 Text companyName = new Text(lineArray[0]);
26 Text productName = new Text(lineArray[1]);
27 System.out.println("cm - " + companyName + "pm - " + productName);
28 //if(!("NA".equals(companyName)) && !("NA".equals(productName))) {
29
30 //if(!("NA".equals(companyName)) && !("NA".equals(productName))) {
31 if(!("NA".equalsIgnoreCase(companyName.toString())) && !("NA".equalsIgnoreCase(productName.toString()))) {
32     System.out.println("here");
33     context.write(value, NullWritable.get());
34 }
35 }
36 }
37 }
38
39 @SuppressWarnings("deprecation")
40 public static void main(String[] args) throws Exception {
41     Configuration conf = new Configuration();
42     Job job = new Job(conf, "DemoTask1");
43     job.setJarByClass(MainTask.class);
44
45     job.setMapOutputKeyClass(LongWritable.class);
46     job.setMapOutputValueClass(Text.class);
47
48     job.setOutputKeyClass(LongWritable.class);
49     job.setOutputValueClass(Text.class);
50     job.setMapperClass(Task1Mapper.class);
51     job.setNumReduceTasks(0);
52
53     job.setInputFormatClass(TextInputFormat.class);
54     job.setOutputFormatClass(TextOutputFormat.class);
```

Output after filtering 'NA' Records

```
Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox
Applications Places System Thu Aug 16, 7:19 PM Acadgild

acadgild@localhost:~$
File Edit View Search Terminal Help

Merged Map outputs=0
GC time elapsed (ms)=125
CPU time spent (ms)=620
Physical memory (bytes) snapshot=84885504
Virtual memory (bytes) snapshot=2056757248
Total committed heap usage (bytes)=32571392

File Input Format Counters
  Bytes Read=717
File Output Format Counters
  Bytes Written=646

You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -ls /Task1_output
18/08/16 19:01:49 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 2018-08-16 19:00 /Task1_output/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 646 2018-08-16 19:00 /Task1_output/part-m-000000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /Task1_output/part-m-000000
18/08/16 19:02:13 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) Task 2:

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

### o Output of total Units sold for each Company

```
Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox
Applications Places System Thu Aug 16, 6:15 PM Acadgild

acadgild@localhost:~$
File Edit View Search Terminal Help

Physical memory (bytes) snapshot=468738048
Virtual memory (bytes) snapshot=6172827648
Total committed heap usage (bytes)=307437568

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=1076
File Output Format Counters
  Bytes Written=43

You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop dfs -ls /Task2_output2
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

18/08/16 18:15:07 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jav
a classes where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup 0 2018-08-16 18:14 /Task2_output2/ _SUCCESS
-rw-r--r-- 1 acadgild supergroup 43 2018-08-16 18:14 /Task2_output2/part-000000
[acadgild@localhost ~]$ hadoop dfs -cat /Task2_output2/part-000000
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

18/08/16 18:15:45 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jav
a classes where applicable
Akai 1
Lava 3
NA 1
Onida 4
Samsung 7
Zen 2
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

3) Task 3:

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