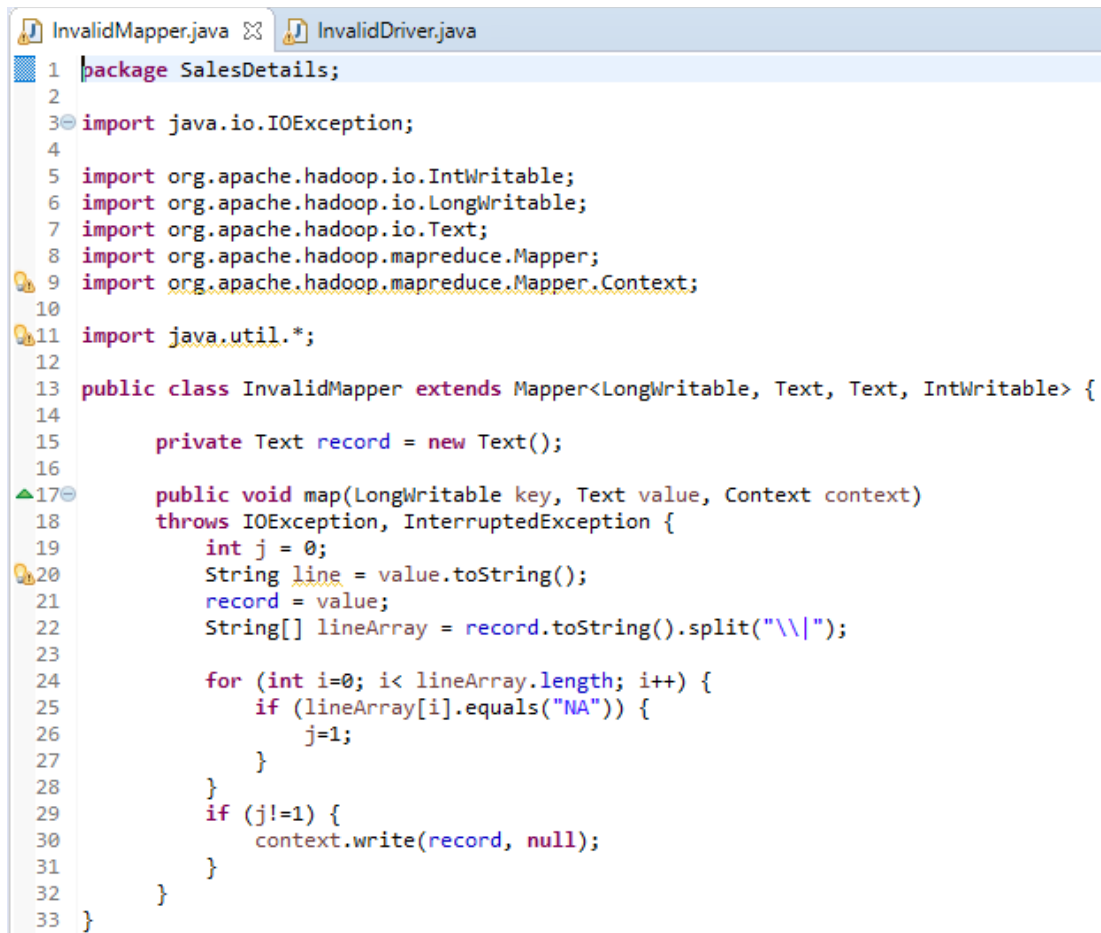


Assignment - 4.1

Task1: Write a Map Reduce program to filter out the invalid records. Map only job will fit for this context.

Answer1:

Java code screen shots are as below.(Mapper code + Driver code)



```
1 package SalesDetails;
2
3 import java.io.IOException;
4
5 import org.apache.hadoop.io.IntWritable;
6 import org.apache.hadoop.io.LongWritable;
7 import org.apache.hadoop.io.Text;
8 import org.apache.hadoop.mapreduce.Mapper;
9 import org.apache.hadoop.mapreduce.Mapper.Context;
10
11 import java.util.*;
12
13 public class InvalidMapper extends Mapper<LongWritable, Text, Text, IntWritable> {
14
15     private Text record = new Text();
16
17     public void map(LongWritable key, Text value, Context context)
18     throws IOException, InterruptedException {
19         int j = 0;
20         String line = value.toString();
21         record = value;
22         String[] lineArray = record.toString().split("\\|");
23
24         for (int i=0; i< lineArray.length; i++) {
25             if (lineArray[i].equals("NA")) {
26                 j=1;
27             }
28         }
29         if (j!=1) {
30             context.write(record, null);
31         }
32     }
33 }
```

```

InvalidMapper.java InvalidDriver.java
1 package SalesDetails;
2
3 import java.io.IOException;
4
5 import org.apache.hadoop.conf.Configuration;
6 import org.apache.hadoop.fs.Path;
7 import org.apache.hadoop.io.*;
8 import org.apache.hadoop.mapred.*;
9 import org.apache.hadoop.mapreduce.Job;
10 import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
11 import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
12 import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
13 import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;
14
15 public class InvalidDriver {
16
17     public static void main(String[] args) throws Exception {
18         if (args.length != 2) {
19             System.err.println("Usage: InvalidDriver <input path> <output path>");
20             System.exit(-1);
21         }
22
23         //Job Related Configurations
24         Configuration conf = new Configuration();
25         Job job = new Job(conf, "Invalid TV records");
26         job.setJarByClass(InvalidDriver.class);
27
28
29         // Specify the number of reducer to 2
30         job.setNumReduceTasks(0);
31
32         //Provide paths to pick the input file for the job
33         FileInputFormat.setInputPaths(job, new Path(args[0]));
34
35
36         //Provide paths to pick the output file for the job, and delete it if already present
37         Path outputPath = new Path(args[1]);
38         FileOutputFormat.setOutputPath(job, outputPath);
39         outputPath.getFileSystem(conf).delete(outputPath, true);
40
41         //To set the mapper and reducer of this job
42         job.setMapperClass(InvalidMapper.class);
43         // job.setReducerClass(WordCountReducer.class);
44
45         //Set the combiner
46         // job.setCombinerClass(WordCountReducer.class);
47
48         //set the input and output format class
49         job.setInputFormatClass(TextInputFormat.class);
50         job.setOutputFormatClass(TextOutputFormat.class);
51
52         //set up the output key and value classes
53         job.setOutputKeyClass(Text.class);
54         job.setOutputValueClass(IntWritable.class);
55
56         //execute the job
57         System.exit(job.waitForCompletion(true) ? 0 : 1);
58
59     }

```

Input Command:

hadoop jar InvalidTVrec.jar /television.txt /TVInvalidout

```
[acadgild@localhost ~]$ hadoop jar InvalidTVrec.jar /television.txt /TVInvalidout
18/03/12 22:37:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/03/12 22:37:50 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/03/12 22:37:51 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/03/12 22:37:51 INFO input.FileInputFormat: Total input paths to process : 1
18/03/12 22:37:51 INFO mapreduce.JobSubmitter: number of splits:1
18/03/12 22:37:51 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1520874036548_0002
18/03/12 22:37:52 INFO impl.YarnClientImpl: Submitted application application_1520874036548_0002
18/03/12 22:37:52 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1520874036548_0002/
18/03/12 22:37:52 INFO mapreduce.Job: Running job: job_1520874036548_0002
18/03/12 22:38:01 INFO mapreduce.Job: Job job_1520874036548_0002 running in uber mode : false
18/03/12 22:38:01 INFO mapreduce.Job: map 0% reduce 0%
18/03/12 22:38:07 INFO mapreduce.Job: map 100% reduce 0%
18/03/12 22:38:08 INFO mapreduce.Job: Job job_1520874036548_0002 completed successfully
18/03/12 22:38:08 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=107371
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=834
    HDFS: Number of bytes written=646
    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)=4565
```

Output: Since this is a mapper only job – the mapper output is checked.

Also here all the records with “NA” has been removed as a part of mapper logic.

```
[acadgild@localhost ~]$ hadoop fs -cat /TVInvalidout/part-m-00000
18/03/12 22:38:29 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
```

Task2: Write a Map Reduce program to calculate the total units sold for each company.

Answer2:

Java code screen shots are as below.(Mapper code + Reducer code + Driver code)

CompanyCountMapper.java CompanyCountReducer.java CompanyCountDriver.java

```
1 package CompanyCount;
2
3 import java.io.IOException;
4
5 import org.apache.hadoop.io.IntWritable;
6 import org.apache.hadoop.io.LongWritable;
7 import org.apache.hadoop.io.Text;
8 import org.apache.hadoop.mapreduce.Mapper;
9 import org.apache.hadoop.mapreduce.Mapper.Context;
10
11 import java.util.*;
12
13 public class CompanyCountMapper extends Mapper<LongWritable, Text, Text, IntWritable> {
14
15     private final static IntWritable one = new IntWritable(1);
16     private Text record = new Text();
17
18     public void map(LongWritable key, Text value, Context context)
19     throws IOException, InterruptedException {
20         int j=0;
21         String line = value.toString();
22         record = value;
23         String[] lineArray = record.toString().split("\\|");
24
25         if (lineArray[0].equals(null) || lineArray[0].equals("NA")) {
26             j=1;
27         }
28
29         if (j!=1) {
30             Text t1 = new Text(lineArray[0]);
31             context.write(t1, one);
32         }
33     }
34 }
```

```
CompanyCountMapper.java CompanyCountReducer.java Compa

1 package CompanyCount;
2
3 import java.io.IOException;
4
5 import org.apache.hadoop.io.IntWritable;
6 import org.apache.hadoop.io.Text;
7 import org.apache.hadoop.mapreduce.Reducer;
8
9
10 public class CompanyCountReducer
11     extends Reducer<Text, IntWritable, Text, IntWritable> {
12
13     @Override
14     public void reduce(Text key, Iterable<IntWritable> values,
15         Context context)
16         throws IOException, InterruptedException {
17         System.out.println("From The Reducer=>" + key) ;
18
19         int sum = 0;
20         for (IntWritable value : values) {
21             sum += value.get();
22         }
23         context.write(key, new IntWritable(sum));
24     }
25 }
```

```
CompanyCountMapper.java CompanyCountReducer.java CompanyCountDriver.java
1 package CompanyCount;
2
3 import java.io.IOException;
4
5 import org.apache.hadoop.conf.Configuration;
6 import org.apache.hadoop.fs.Path;
7 import org.apache.hadoop.io.*;
8 import org.apache.hadoop.mapred.*;
9 import org.apache.hadoop.mapreduce.Job;
10 import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
11 import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
12 import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
13 import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;
14
15 import CompanyCount.CompanyCountDriver;
16
17 public class CompanyCountDriver {
18
19     public static void main(String[] args) throws Exception {
20         // TODO Auto-generated method stub
21         if (args.length != 2) {
22             System.err.println("Usage: CompanyCountDriver <input path> <output path>");
23             System.exit(-1);
24         }
25
26         //Job Related Configurations
27         Configuration conf = new Configuration();
28         Job job = new Job(conf, "Total Units sold for each Company");
29         job.setJarByClass(CompanyCountDriver.class);
30
31         // Specify the number of reducer to 1
32         job.setNumReduceTasks(1);
33
34         //Provide paths to pick the input file for the job
35         FileInputFormat.setInputPaths(job, new Path(args[0]));
36
37         //Provide paths to pick the output file for the job, and delete it if already present
38         Path outputPath = new Path(args[1]);
39         FileOutputFormat.setOutputPath(job, outputPath);
40         outputPath.getFileSystem(conf).delete(outputPath, true);
41
42         //To set the mapper and reducer of this job
43         job.setMapperClass(CompanyCount.CompanyCountMapper.class);
44         job.setReducerClass(CompanyCount.CompanyCountReducer.class);
45
46         //set the input and output format class
47         job.setInputFormatClass(TextInputFormat.class);
48         job.setOutputFormatClass(TextOutputFormat.class);
49
50         //set up the output key and value classes
51         job.setOutputKeyClass(Text.class);
52         job.setOutputValueClass(IntWritable.class);
53
54         //execute the job
55         System.exit(job.waitForCompletion(true) ? 0 : 1);
56     }
57 }
58 }
```

Input Command:

```
hadoop jar CompanyCount.jar /television.txt /CompCountout
```

```
[acadgild@localhost ~]$ hadoop jar CompanyCount.jar /television.txt /CompCountout
18/03/12 22:55:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/03/12 22:55:49 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/03/12 22:55:50 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/03/12 22:55:50 INFO input.FileInputFormat: Total input paths to process : 1
18/03/12 22:55:50 INFO mapreduce.JobSubmitter: number of splits:1
18/03/12 22:55:51 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1520874036548_0004
18/03/12 22:55:51 INFO impl.YarnClientImpl: Submitted application application_1520874036548_0004
18/03/12 22:55:51 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1520874036548_0004/
18/03/12 22:55:51 INFO mapreduce.Job: Running job: job_1520874036548_0004
18/03/12 22:56:00 INFO mapreduce.Job: Job job_1520874036548_0004 running in uber mode : false
18/03/12 22:56:00 INFO mapreduce.Job: map 0% reduce 0%
18/03/12 22:56:07 INFO mapreduce.Job: map 100% reduce 0%
18/03/12 22:56:15 INFO mapreduce.Job: map 100% reduce 100%
18/03/12 22:56:15 INFO mapreduce.Job: Job job_1520874036548_0004 completed successfully
18/03/12 22:56:15 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=216
    FILE: Number of bytes written=215815
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=834
    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)=4371
    Total time spent by all reduces in occupied slots (ms)=4883
    Total time spent by all map tasks (ms)=4371
```

Output: Here the Company counts are mentioned for valid records only.

```
[acadgild@localhost ~]$ hadoop fs -cat /CompCountout/part-r-00000
18/03/12 22:56:38 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Akai 1
Lava 3
Onida 4
Samsung 7
Zen 2
```

Task3: Write a Map Reduce program to calculate the total units sold in each state for Onida company.

Answer3:

Java code screen shots are as below.(Mapper code + Reducer code + Driver code)

```
StateCountMapper.java  StateCountReducer.java  StateCountDriver.java
1 package StateCount;
2
3 import java.io.IOException;
4
5 import org.apache.hadoop.io.IntWritable;
6 import org.apache.hadoop.io.LongWritable;
7 import org.apache.hadoop.io.Text;
8 import org.apache.hadoop.mapreduce.Mapper;
9 import org.apache.hadoop.mapreduce.Mapper.Context;
10
11 import java.util.*;
12
13 public class StateCountMapper extends Mapper<LongWritable, Text, Text, IntWritable> {
14
15     private final static IntWritable one = new IntWritable(1);
16     private Text record = new Text();
17
18     public void map(LongWritable key, Text value, Context context)
19         throws IOException, InterruptedException {
20         int j=0;
21         String line = value.toString();
22         record = value;
23         String[] lineArray = record.toString().split("\\|");
24
25         if (lineArray[0].equals("Onida")) {
26             // Gets the State name only for Onida company.
27             Text t1 = new Text(lineArray[3]);
28             context.write(t1, one);
29         }
30     }
31 }
```

```
StateCountMapper.java  StateCountReducer.java  StateCountDriver.java
1 package StateCount;
2
3 import java.io.IOException;
4
5 import org.apache.hadoop.io.IntWritable;
6 import org.apache.hadoop.io.Text;
7 import org.apache.hadoop.mapreduce.Reducer;
8
9
10 public class StateCountReducer
11     extends Reducer<Text, IntWritable, Text, IntWritable> {
12
13     @Override
14     public void reduce(Text key, Iterable<IntWritable> values,
15         Context context)
16         throws IOException, InterruptedException {
17         System.out.println("From The Reducer=>" + key);
18
19         int sum = 0;
20         for (IntWritable value : values) {
21             sum += value.get();
22         }
23         context.write(key, new IntWritable(sum));
24     }
25 }
```



```

StateCountMapper.java  StateCountReducer.java  StateCountDriver.java
1 package StateCount;
2
3 import java.io.IOException;
4
5 import org.apache.hadoop.conf.Configuration;
6 import org.apache.hadoop.fs.Path;
7 import org.apache.hadoop.io.*;
8 import org.apache.hadoop.mapred.*;
9 import org.apache.hadoop.mapreduce.Job;
10 import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
11 import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
12 import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
13 import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;
14
15 import StateCount.StateCountDriver;
16
17 public class StateCountDriver {
18
19     public static void main(String[] args) throws IOException, ClassNotFoundException, InterruptedException {
20         // TODO Auto-generated method stub
21         if (args.length != 2) {
22             System.err.println("Usage: StateCountDriver <input path> <output path>");
23             System.exit(-1);
24         }
25
26         //Job Related Configurations
27         Configuration conf = new Configuration();
28         Job job = new Job(conf, "Total Units sold in each State for Onida Company");
29         job.setJarByClass(StateCountDriver.class);
30
31         // Specify the number of reducer to 2
32         job.setNumReduceTasks(1);
33
34         //Provide paths to pick the input file for the job
35         FileInputFormat.setInputPaths(job, new Path(args[0]));
36
37         //Provide paths to pick the output file for the job, and delete it if already present
38         Path outputPath = new Path(args[1]);
39         FileOutputFormat.setOutputPath(job, outputPath);
40         outputPath.getFileSystem(conf).delete(outputPath, true);
41
42         //To set the mapper and reducer of this job
43         job.setMapperClass(StateCount.StateCountMapper.class);
44         job.setReducerClass(StateCount.StateCountReducer.class);
45
46         //set the input and output format class
47         job.setInputFormatClass(TextInputFormat.class);
48         job.setOutputFormatClass(TextOutputFormat.class);
49
50         //set up the output key and value classes
51         job.setOutputKeyClass(Text.class);
52         job.setOutputValueClass(IntWritable.class);
53
54         //execute the job
55         System.exit(job.waitForCompletion(true) ? 0 : 1);
56     }
57 }
58
59 }

```

Input Command:

hadoop jar StateCount.jar /television.txt /StateCountout

```
[acadgild@localhost ~]$ hadoop jar StateCount.jar /television.txt /StateCountout
18/03/12 23:07:29 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/03/12 23:07:30 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/03/12 23:07:31 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/03/12 23:07:31 INFO input.FileInputFormat: Total input paths to process : 1
18/03/12 23:07:31 INFO mapreduce.JobSubmitter: number of splits:1
18/03/12 23:07:32 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1520874036548_0005
18/03/12 23:07:32 INFO impl.YarnClientImpl: Submitted application application_1520874036548_0005
18/03/12 23:07:32 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1520874036548_0005/
18/03/12 23:07:32 INFO mapreduce.Job: Running job: job_1520874036548_0005
18/03/12 23:07:41 INFO mapreduce.Job: Job job_1520874036548_0005 running in uber mode : false
18/03/12 23:07:41 INFO mapreduce.Job: map 0% reduce 0%
18/03/12 23:07:49 INFO mapreduce.Job: map 100% reduce 0%
18/03/12 23:07:58 INFO mapreduce.Job: map 100% reduce 100%
18/03/12 23:07:58 INFO mapreduce.Job: Job job_1520874036548_0005 completed successfully
18/03/12 23:07:58 INFO mapreduce.Job: Counters: 49
    File System Counters
      FILE: Number of bytes read=79
      FILE: Number of bytes written=215557
      FILE: Number of read operations=0
      FILE: Number of large read operations=0
      FILE: Number of write operations=0
      HDFS: Number of bytes read=834
      HDFS: Number of bytes written=25
      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)=4680
      Total time spent by all reduces in occupied slots (ms)=5589
      Total time spent by all map tasks (ms)=4680
```

Output: Here the counts are only for valid records having company name as “Onida”.

```
[acadgild@localhost ~]$ hadoop fs -cat /StateCountout/part-r-00000
18/03/12 23:08:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Kerala 1
Uttar Pradesh 3
[acadgild@localhost ~]$
```