ASSIGNMENT 4

MR -INTRODUCTION

PROBLEM STATEMENT –

1. Write a Map Reduce program to filter out the invalid records. Map only job will fit for this context.
2. Write a Map Reduce program to calculate the total units sold for each Company.
3. Write a Map Reduce program to calculate the total units sold in each state for Onida company.

SOLUTION -

1. **Map method -**

```
public void map(LongWritable key,Text value,Context context) throws IOException, InterruptedI
{
    String line=value.toString();
    String[]linearray=line.split("\\|");//split the string and store in linearray

    if(!(linearray[0].equals("NA")||linearray[1].equals("NA")))//if field 0 or 1 not "NA"
    {
        context.write(new Text(line), new Text());//write valid records
    }
}
```

Driver code –

```
public static void main(String[] args) throws Exception

{
    Configuration conf=new Configuration();
    Job job=new Job(conf, "Invalid Data");// the job runs under this

    job.setJarByClass(InvalidRecords.class);

    job.setMapOutputKeyClass(Text.class);//mapper key output
    job.setMapOutputValueClass(Text.class);//mapper output value

    job.setMapperClass(InvalidRecordsMapper.class);// mapper class
    job.setNumReduceTasks(0);

    job.setInputFormatClass(TextInputFormat.class);
    job.setOutputFormatClass(TextOutputFormat.class);

    FileInputFormat.addInputPath(job,new Path(args[0]));
    FileOutputFormat.setOutputPath(job,new Path(args[1]));

    job.waitForCompletion(true);

}
```

OUTPUT –

```
[acadgild@localhost hadoop]$ hadoop jar InvalidRecords.jar invalidRecords.InvalidRecords /television.txt /InvalidRecordsOutput
18/10/01 23:14:19 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes wh
18/10/01 23:14:22 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/10/01 23:14:24 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface
th ToolRunner to remedy this.
18/10/01 23:14:26 INFO input.FileInputFormat: Total input paths to process : 1
18/10/01 23:14:27 INFO mapreduce.JobSubmitter: number of splits:1
18/10/01 23:14:28 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1538415776141_0002
18/10/01 23:14:29 INFO impl.YarnClientImpl: Submitted application application_1538415776141_0002
18/10/01 23:14:30 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1538415776141_0002/
18/10/01 23:14:30 INFO mapreduce.Job: Running job: job_1538415776141_0002
18/10/01 23:14:55 INFO mapreduce.Job: Job job_1538415776141_0002 running in uber mode : false
18/10/01 23:14:55 INFO mapreduce.Job:  map 0% reduce 0%
18/10/01 23:15:12 INFO mapreduce.Job:  map 100% reduce 0%
18/10/01 23:15:13 INFO mapreduce.Job: Job job_1538415776141_0002 completed successfully
18/10/01 23:15:14 INFO mapreduce.Job: Counters: 30
```

```
[acadgild@localhost hadoop]$ hadoop fs -ls /InvalidRecordsOutput
18/10/01 23:15:59 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform.
Found 2 items
-rw-r--r--   1 acadgild supergroup          0 2018-10-01 23:15 /InvalidRecordsOutput/_SUCCESS
-rw-r--r--   1 acadgild supergroup        662 2018-10-01 23:15 /InvalidRecordsOutput/part-m-00000
[acadgild@localhost hadoop]$ hadoop fs -cat /InvalidRecordsOutput/part-m-00000
18/10/01 23:16:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform.
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
```

2. Map method –

```
lic class TotalUnitSaleMapper extends Mapper<LongWritable, Text, Text, IntWritable>

 private final static IntWritable unit = new IntWritable(1); // declaring the Mapper value
 private Text CompanyName = new Text();          //declaring the Mapper key

 public void map(LongWritable key, Text value, Context context ) throws IOException, InterruptedException
 {

     String[] Linearray = value.toString().split("\\|");
     StringTokenizer tokenizer=new StringTokenizer(Linearray[0]); //we have used the String Tokenizer cla
     while(tokenizer.hasMoreTokens()) // the while loop checks for the more tokens/words, if we have next
     {
         CompanyName.set(tokenizer.nextToken());

 }
         context.write(CompanyName, unit); // output of the Mapper Key and Value
```

Reduce method –

```
public class TotalUnitSaleReducer extends Reducer<Text, IntWritable, Text, IntWritable>
{
    public void reduce(Text CompanyName, Iterable<IntWritable> values, Context context) throws IOExceptio
    {
        int sum=0;                // declaring a variable sum
        for(IntWritable value:values) // the for loop get the iterable values and counting the values
        {
            sum+=value.get();
        }
        context.write(CompanyName, new IntWritable(sum)); // output of the the Key and value
    }

}
```

Driver code –

```
public class TotalUnitSale
{

    public static void main(String[] args) throws Exception
    {
        Configuration conf = new Configuration();
        Job job = new Job(conf, "TV TotalUnitSale");// the job runs under this

        job.setJarByClass(TotalUnitSale.class);

        job.setMapOutputKeyClass(Text.class); //mapper key output
        job.setMapOutputValueClass(IntWritable.class); //mapper output value

        job.setOutputKeyClass(Text.class);// output key of the mapreduce
        job.setOutputValueClass(IntWritable.class);//output value of the mapreduce

        job.setMapperClass(TotalUnitSaleMapper.class);// Mapper class
        job.setReducerClass(TotalUnitSaleReducer.class);//reducer class

        job.setNumReduceTasks(2);

        job.setInputFormatClass(TextInputFormat.class);
        job.setOutputFormatClass(TextOutputFormat.class);

        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));

        job.waitForCompletion(true);

    }
```

OUTPUT –

```
[acadgild@localhost hadoop]$ hadoop jar TotalUnits.jar totalUnits.TotalUnitSale /television.txt /TotalUnitsOutput
18/10/01 23:17:20 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
18/10/01 23:17:23 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/10/01 23:17:25 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool
th ToolRunner to remedy this.
18/10/01 23:17:26 INFO input.FileInputFormat: Total input paths to process : 1
18/10/01 23:17:26 INFO mapreduce.JobSubmitter: number of splits:1
18/10/01 23:17:27 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1538415776141_0003
18/10/01 23:17:27 INFO impl.YarnClientImpl: Submitted application application_1538415776141_0003
18/10/01 23:17:27 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1538415776141_0003/
18/10/01 23:17:27 INFO mapreduce.Job: Running job: job_1538415776141_0003
18/10/01 23:17:44 INFO mapreduce.Job: Job job_1538415776141_0003 running in uber mode : false
18/10/01 23:17:44 INFO mapreduce.Job:  map 0% reduce 0%
18/10/01 23:17:58 INFO mapreduce.Job:  map 100% reduce 0%
18/10/01 23:18:17 INFO mapreduce.Job:  map 100% reduce 50%
18/10/01 23:18:18 INFO mapreduce.Job:  map 100% reduce 100%
18/10/01 23:18:19 INFO mapreduce.Job: Job job_1538415776141_0003 completed successfully
```

```
[acadgild@localhost hadoop]$ hadoop fs -cat /TotalUnitsOutput/part-r-00000
18/10/01 23:25:28 WARN util.NativeCodeLoader: Unable to load native-hadoop
NA      1
Onida   4
Zen     2
[acadgild@localhost hadoop]$ hadoop fs -cat /TotalUnitsOutput/part-r-00001
18/10/01 23:25:39 WARN util.NativeCodeLoader: Unable to load native-hadoop
Akai    1
Lava    3
Samsung 7
[acadgild@localhost hadoop]$
```

3. **Map method –**

```
class OnidaMapper extends Mapper<LongWritable, Text, Text, IntWritable>

lic void map(LongWritable key, Text value, Context context) throws IOException, InterruptedException


 String[] Linearray = value.toString().split("\\|"); //the array is split into string value and stored in
 if(Linearray[0].equals("Onida")) // checking the word Onida in the linearray[0], if it is Onida print th
 {
     Text State = new Text(Linearray[3]);
     IntWritable unit= new IntWritable(1);
     context.write(State, unit);
 }
```

Reduce method –

```
ic class OnidaReducer extends Reducer<Text, IntWritable, Text, IntWritable>

public void reduce(Text State, Iterable<IntWritable> values, Context context) throws IOException, Interru
{
    int sum = 0;  // declaring the variable sum
    for(IntWritable value:values)  // the for loop get the iterable values and counting the values
    {
        sum+= value.get();
    }

    context.write(State, new IntWritable(sum)); // print the state name which is the key and the number o
}
```

Driver code –

```
public static void main(String[] args) throws Exception

{

    Configuration conf = new Configuration();
    Job job = new Job(conf, "Onida Total Unit");// the job runs under this

    job.setJarByClass(OnidaTotalUnit.class);

    job.setMapOutputKeyClass(Text.class); //mapper key output
    job.setMapOutputValueClass(IntWritable.class); //mapper output value

    job.setOutputKeyClass(Text.class);//output key of the mapreduce
    job.setOutputValueClass(IntWritable.class); //output value of the mapreduce

    job.setMapperClass(OnidaMapper.class); // mapper class
    job.setReducerClass(OnidaReducer.class);// reducer class

    job.setNumReduceTasks(2);

    job.setInputFormatClass(TextInputFormat.class);
    job.setOutputFormatClass(TextOutputFormat.class);

    FileInputFormat.addInputPath(job, new Path(args[0]));
    FileOutputFormat.setOutputPath(job, new Path(args[1]));

    job.waitForCompletion(true);
}
```

OUTPUT –

```
[acadgild@localhost hadoop]$ hadoop jar OnidaUnits.jar totalUnitsOnida.OnidaTotalUnit /television.txt /OnidaOutput
18/10/01 23:26:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
18/10/01 23:26:51 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/10/01 23:26:53 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool
th ToolRunner to remedy this.
18/10/01 23:26:54 INFO input.FileInputFormat: Total input paths to process : 1
18/10/01 23:26:54 INFO mapreduce.JobSubmitter: number of splits:1
18/10/01 23:26:55 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1538415776141_0004
18/10/01 23:26:55 INFO impl.YarnClientImpl: Submitted application application_1538415776141_0004
18/10/01 23:26:55 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1538415776141_0004/
18/10/01 23:26:55 INFO mapreduce.Job: Running job: job_1538415776141_0004
18/10/01 23:27:13 INFO mapreduce.Job: Job job_1538415776141_0004 running in uber mode : false
18/10/01 23:27:13 INFO mapreduce.Job:  map 0% reduce 0%
18/10/01 23:27:26 INFO mapreduce.Job:  map 100% reduce 0%
18/10/01 23:27:45 INFO mapreduce.Job:  map 100% reduce 100%
18/10/01 23:27:47 INFO mapreduce.Job: Job job_1538415776141_0004 completed successfully
18/10/01 23:27:47 INFO mapreduce.Job: Counters: 49
```

```
[acadgild@localhost hadoop]$ hadoop fs -cat /OnidaOutput/part-r-00000
18/10/01 23:28:34 WARN util.NativeCodeLoader: Unable to load native-hadoop
Uttar Pradesh   3
[acadgild@localhost hadoop]$ hadoop fs -cat /OnidaOutput/part-r-00001
18/10/01 23:28:43 WARN util.NativeCodeLoader: Unable to load native-hadoop
Kerala  1
```