

Assignment 4

Problem Statement:

We have a dataset of sales of different TV sets across different locations.

Records look like:

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

The fields are arranged like:

Company Name|Product Name|Size in inches|State|Pin Code|Price

There are some invalid records which contain 'NA' in either Company Name or Product Name.

Dataset:

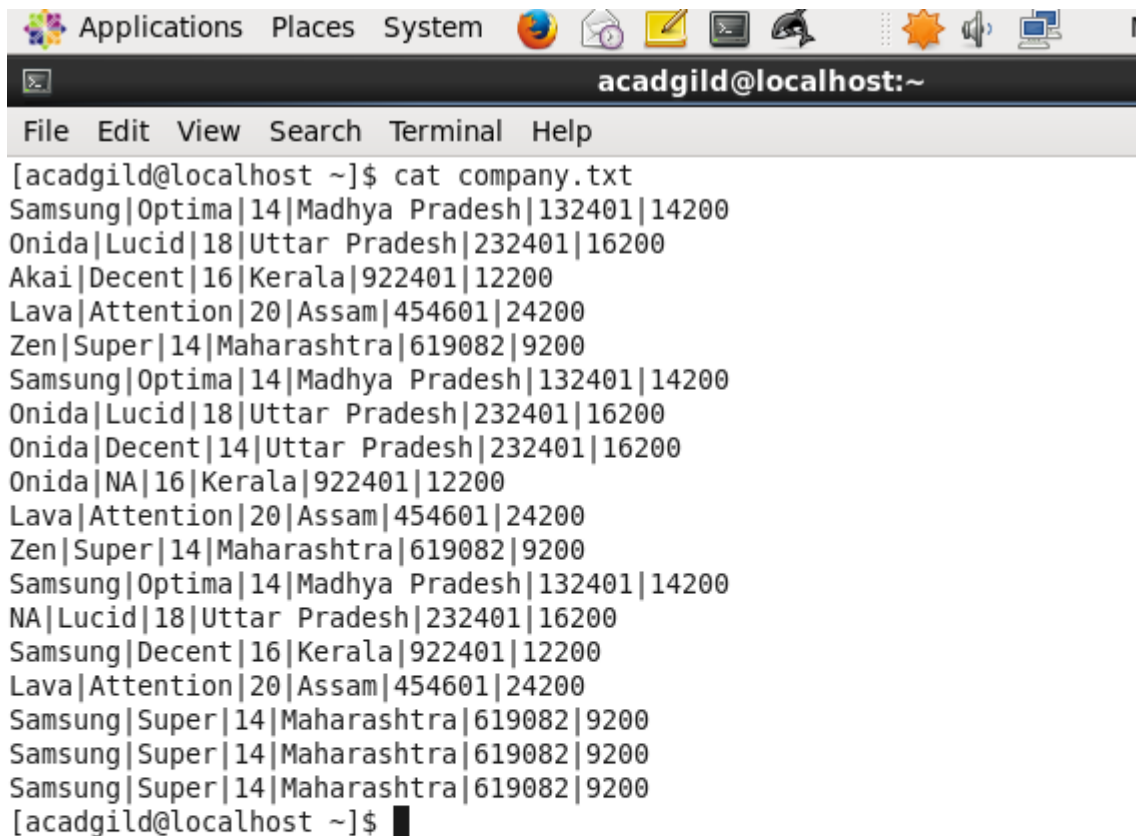
```
Samsung|Optima|14|Madhya Pradesh|132401|14200
Onida|Lucid|18|Uttar Pradesh|232401|16200
Akai|Decent|16|Kerala|922401|12200
Lava|Attention|20|Assam|454601|24200
Zen|Super|14|Maharashtra|619082|9200
Samsung|Optima|14|Madhya Pradesh|132401|14200
Onida|Lucid|18|Uttar Pradesh|232401|16200
Onida|Decent|14|Uttar 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
```

Create a file:

```
cat > company.txt
```

To View a file:

```
cat company.txt
```



A terminal window titled 'acadgild@localhost:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal displays the output of the command 'cat company.txt'. The output consists of 18 lines of text, each representing a record with fields separated by vertical bars. The records are as follows:

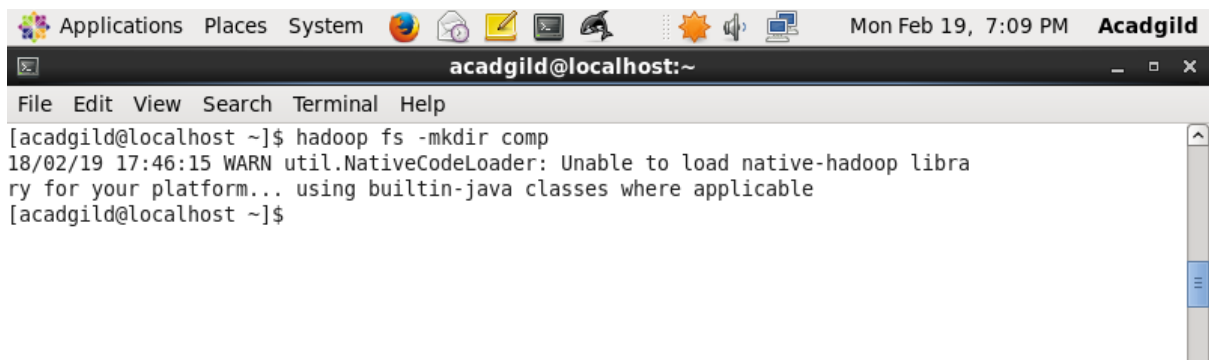
Brand	Model	Year	State	Pin	Phone
Samsung	Optima	14	Madhya Pradesh	132401	14200
Onida	Lucid	18	Uttar Pradesh	232401	16200
Akai	Decent	16	Kerala	922401	12200
Lava	Attention	20	Assam	454601	24200
Zen	Super	14	Maharashtra	619082	9200
Samsung	Optima	14	Madhya Pradesh	132401	14200
Onida	Lucid	18	Uttar Pradesh	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

The terminal prompt is '[acadgild@localhost ~]\$'.



Make a directory:

Hadoop fs -mkdir comp

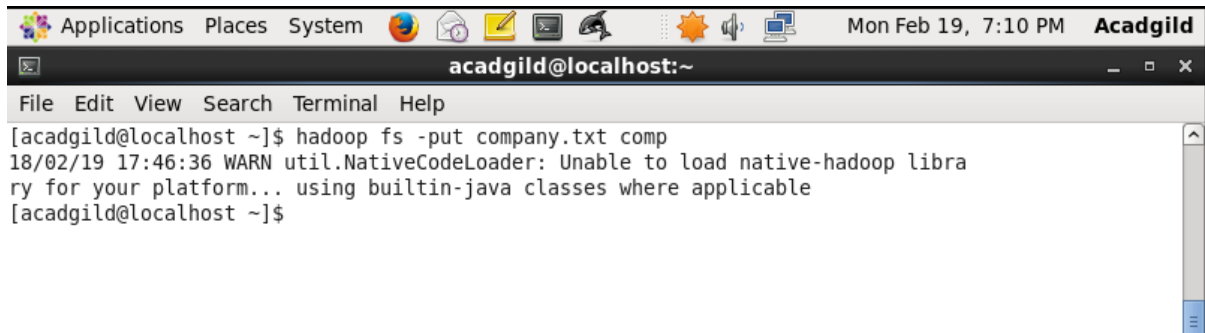


A terminal window titled 'acadgild@localhost:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal displays the output of the command 'hadoop fs -mkdir comp'. The output is:

```
[acadgild@localhost ~]$ hadoop fs -mkdir comp
18/02/19 17:46:15 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[acadgild@localhost ~]$
```

Push the file into directory:

Hadoop fs -put company.txt comp

A screenshot of a terminal window titled 'acadgild@localhost:~'. The window shows the command 'hadoop fs -put company.txt comp' being executed. The output includes a warning message: '18/02/19 17:46:36 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable'. The terminal interface includes a menu bar with 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'. The system tray at the top shows various icons and the date/time 'Mon Feb 19, 7:10 PM'.

Task 1:

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

Driver Code:

```
package Company;

import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;
import org.w3c.dom.Text;

public class company
{

    public static void main(String[] args) throws Exception
    {
        Configuration conf=new Configuration();
        Job job=new Job(conf, "Invalid Data");

        job.setJarByClass(InvalidRecord.class);

        job.setMapOutputKeyClass(Text.class);
        job.setMapOutputValueClass(Text.class);

        job.setMapperClass(InvalidRecordsMapper.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);
    }
}
```

Mapper code

```
package companymapper;

import java.io.IOException;

import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;

public class companymapper extends Mapper<LongWritable,Text,Text,Text>
{
    public void map(LongWritable key,Text value,Context context) throws
IOException, InterruptedException
    {
        String line=value.toString();
        String[]linearray=line.split("\\|");

        if(!(linearray[0].equals("NA")||linearray[1].equals("NA")))
        {
            context.write(new Text(line), new Text());
        }
    }
}
```

Output:

Hadoop jar company.jar Company.company comp/company.txt comp/company11

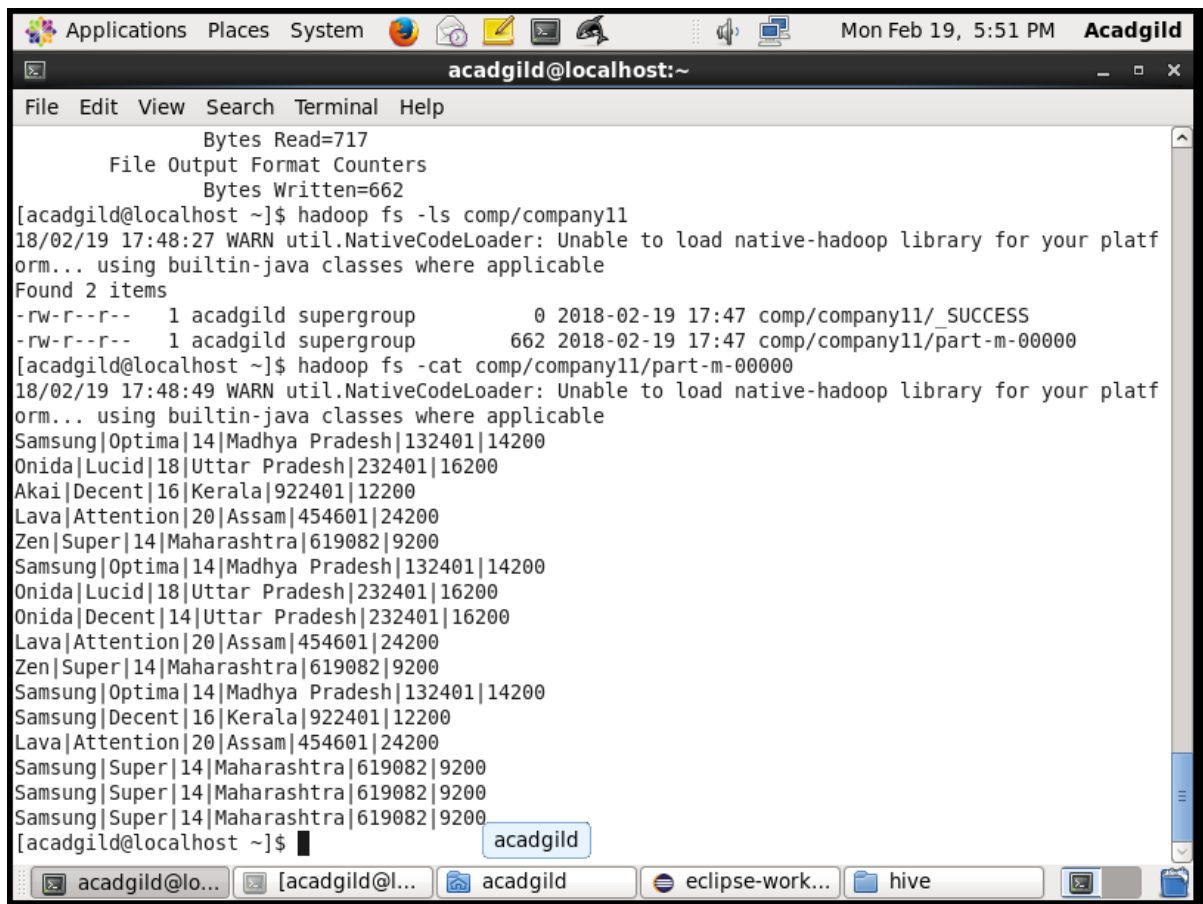
```
Applications Places System Mon Feb 19, 5:49 PM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help

[acadgild@localhost ~]$ hadoop jar company.jar Company.company comp/company.txt
comp/company11
18/02/19 17:47:21 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
18/02/19 17:47:23 INFO client.RMProxy: Connecting to ResourceManager at localhos
t/127.0.0.1:8032
18/02/19 17:47:25 WARN mapreduce.JobResourceUploader: Hadoop command-line option
 parsing not performed. Implement the Tool interface and execute your applicatio
n with ToolRunner to remedy this.
18/02/19 17:47:25 INFO input.FileInputFormat: Total input files to process : 1
18/02/19 17:47:25 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1252)
    at java.lang.Thread.join(Thread.java:1326)
    at org.apache.hadoop.hdfs.DataStreamer.closeResponder(DataStreamer.java:
980)
    at org.apache.hadoop.hdfs.DataStreamer.endBlock(DataStreamer.java:630)
    at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:807)
18/02/19 17:47:25 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1252)
    at java.lang.Thread.join(Thread.java:1326)
    at org.apache.hadoop.hdfs.DataStreamer.closeResponder(DataStreamer.java:
980)
    at org.apache.hadoop.hdfs.DataStreamer.endBlock(DataStreamer.java:630)
    at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:807)
```

```
Applications Places System Mon Feb 19, 5:50 PM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help

18/02/19 17:47:28 INFO mapreduce.Job: Running job: job_1519042378766_0001
18/02/19 17:47:50 INFO mapreduce.Job: Job job_1519042378766_0001 running in uber mode : false
18/02/19 17:47:50 INFO mapreduce.Job: map 0% reduce 0%
18/02/19 17:48:00 INFO mapreduce.Job: map 100% reduce 0%
18/02/19 17:48:01 INFO mapreduce.Job: Job job_1519042378766_0001 completed successfully
18/02/19 17:48:02 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=201785
    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=662
    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)=7966
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=7966
    Total vcore-milliseconds taken by all map tasks=7966
    Total megabyte-milliseconds taken by all map tasks=8157184
  Map-Reduce Framework
    Map input records=18
    Map output records=16
I acadgild@localhost:~
```

```
hadoop fs -ls comp/company11
hadoop fs -cat comp/company11/part-r-00000
```



The screenshot shows a terminal window titled 'acadgild@localhost:~'. The window contains the following text:

```
Bytes Read=717
File Output Format Counters
Bytes Written=662
[acadgild@localhost ~]$ hadoop fs -ls comp/company11
18/02/19 17:48:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
Found 2 items
-rw-r--r--  1 acadgild supergroup          0 2018-02-19 17:47 comp/company11/_SUCCESS
-rw-r--r--  1 acadgild supergroup      662 2018-02-19 17:47 comp/company11/part-m-00000
[acadgild@localhost ~]$ hadoop fs -cat comp/company11/part-m-00000
18/02/19 17:48:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... 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 ~]$
```

Task 2:

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

Driver Code:

```
package Company;

import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;

public class company
{
    public static void main(String[] args) throws Exception
    {
        Configuration conf = new Configuration();
```

```

this      Job job = new Job(conf, "TV TotalUnitSale");// the job runs under

        job.setJarByClass(company.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(companymapper.class); // Mapper class
        job.setReducerClass(companyreducer.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);

    }

}

```

Mapper Code:

```

package Company;

import java.io.IOException;
import java.util.StringTokenizer;

import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;

public class companymapper 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 class which takes array into single word/token.
        while(tokenizer.hasMoreTokens()) // the while loop checks for the
more tokens/words, if we have next token it will continue the loop
        {

```

```

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

```

Reducer Code:

```

package Company;

import java.io.IOException;

import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;

public class companyreducer extends Reducer<Text, IntWritable, Text, IntWritable>
{
    public void reduce(Text CompanyName, Iterable<IntWritable> values, Context
context) throws IOException, InterruptedException
    {
        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
Key and value
    }
}

```

Output:

Hadoop jar company.jar Company.company comp/company.txt comp/company12

Applications Places System Mon Feb 19, 6:02 PM Acadgild

acadgild@localhost:~

File Edit View Search Terminal Help

```
[acadgild@localhost ~]$ hadoop jar company.jar Company.company comp/company.txt comp/company12
18/02/19 18:01:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
18/02/19 18:01:52 INFO client.RMPProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/02/19 18:01:53 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not perf
ormed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/02/19 18:01:54 INFO input.FileInputFormat: Total input files to process : 1
18/02/19 18:01:54 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedExce
ption
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1252)
    at java.lang.Thread.join(Thread.java:1326)
    at org.apache.hadoop.hdfs.DataStreamer.closeResponder(DataStreamer.java:980)
    at org.apache.hadoop.hdfs.DataStreamer.endBlock(DataStreamer.java:630)
    at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:807)
18/02/19 18:01:54 INFO mapreduce.JobSubmitter: number of splits:1
18/02/19 18:01:54 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.e
nabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
18/02/19 18:01:55 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1519042378766_0002
18/02/19 18:01:56 INFO impl.YarnClientImpl: Submitted application application_1519042378766_0002
18/02/19 18:01:56 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/appli
cation_1519042378766_0002/
18/02/19 18:01:56 INFO mapreduce.Job: Running job: job_1519042378766_0002
18/02/19 18:02:12 INFO mapreduce.Job: Job job_1519042378766_0002 running in uber mode : false
18/02/19 18:02:12 INFO mapreduce.Job:  map 0% reduce 0%
18/02/19 18:02:24 INFO mapreduce.Job:  map 100% reduce 0%
```

acadgild

acadgild... [acadgil... acadgild eclipse-... hive company...

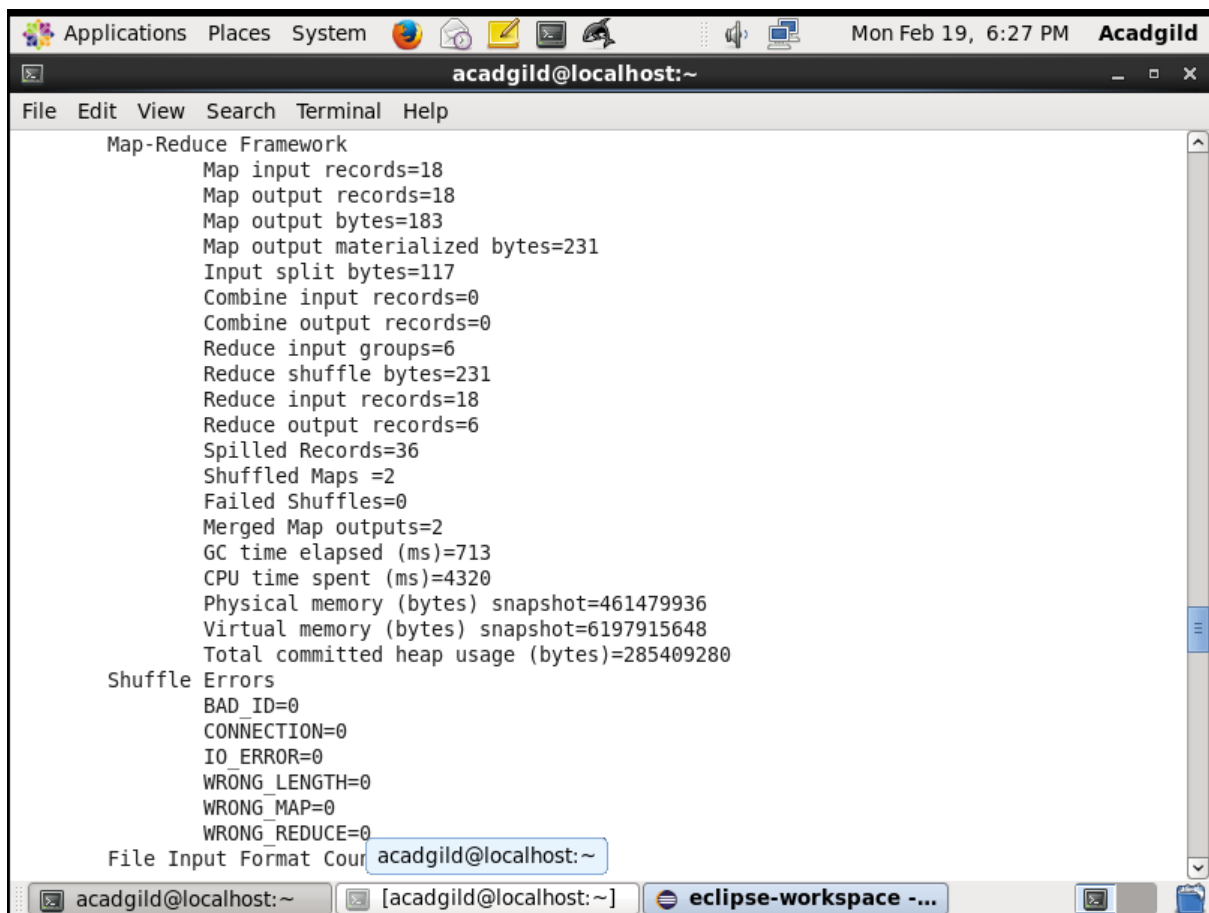
Applications Places System Mon Feb 19, 6:26 PM Acadgild

acadgild@localhost:~

File Edit View Search Terminal Help

```
18/02/19 18:02:12 INFO mapreduce.Job:  map 0% reduce 0%
18/02/19 18:02:24 INFO mapreduce.Job:  map 100% reduce 0%
18/02/19 18:02:46 INFO mapreduce.Job:  map 100% reduce 100%
18/02/19 18:02:47 INFO mapreduce.Job: Job job_1519042378766_0002 completed successfully
18/02/19 18:02:47 INFO mapreduce.Job: Counters: 50
  File System Counters
    FILE: Number of bytes read=231
    FILE: Number of bytes written=608017
    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=43
    HDFS: Number of read operations=9
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=4
  Job Counters
    Killed reduce tasks=1
    Launched map tasks=1
    Launched reduce tasks=2
    Data-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=9376
    Total time spent by all reduces in occupied slots (ms)=37167
    Total time spent by all map tasks (ms)=9376
    Total time spent by all reduce tasks (ms)=37167
    Total vcore-milliseconds taken by all map tasks=9376
    Total vcore-milliseconds taken by all reduce tasks=37167
    Total megabyte-milliseconds taken by all map tasks=9601024
    Total megabyte-milliseconds taken by all reduce tasks=38059008
```

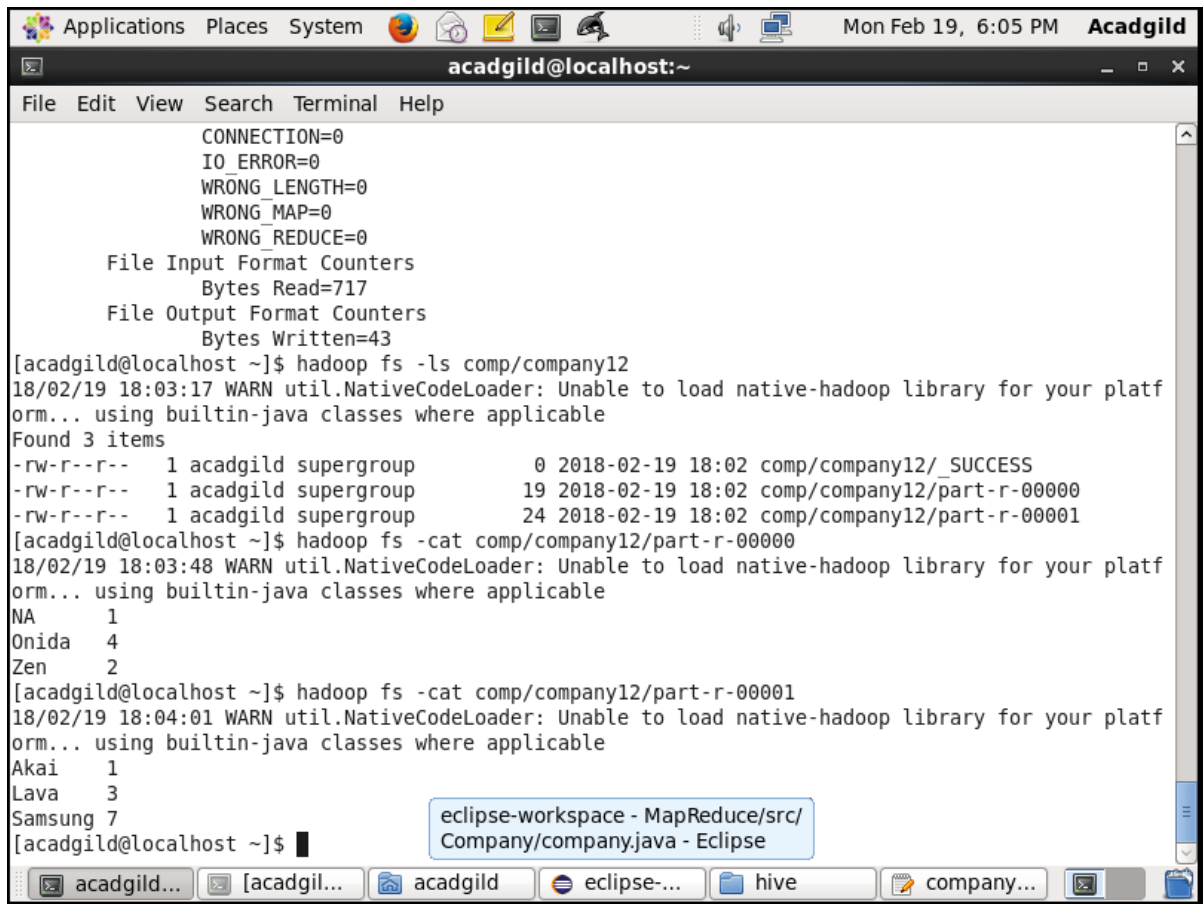
acadgild@localhost:~ [acadgild@localhost:~] eclipse-workspace -...



The screenshot shows an Eclipse IDE window with a terminal titled 'acadgild@localhost:~'. The terminal displays the output of a Hadoop Map-Reduce job. The output is organized into sections: 'Map-Reduce Framework' and 'Shuffle Errors'. The 'Map-Reduce Framework' section lists various statistics such as input/output records, bytes, materialized bytes, split bytes, combine records, reduce groups, shuffle bytes, spilled records, shuffled maps, failed shuffles, merged map outputs, GC time, CPU time, physical memory, virtual memory, and total committed heap usage. The 'Shuffle Errors' section lists several error types, all of which are set to 0, indicating no errors occurred. The terminal window is part of an Eclipse workspace, as indicated by the 'eclipse-workspace -...' tab at the bottom.

```
acadmild@localhost:~  
File Edit View Search Terminal Help  
Map-Reduce Framework  
  Map input records=18  
  Map output records=18  
  Map output bytes=183  
  Map output materialized bytes=231  
  Input split bytes=117  
  Combine input records=0  
  Combine output records=0  
  Reduce input groups=6  
  Reduce shuffle bytes=231  
  Reduce input records=18  
  Reduce output records=6  
  Spilled Records=36  
  Shuffled Maps =2  
  Failed Shuffles=0  
  Merged Map outputs=2  
  GC time elapsed (ms)=713  
  CPU time spent (ms)=4320  
  Physical memory (bytes) snapshot=461479936  
  Virtual memory (bytes) snapshot=6197915648  
  Total committed heap usage (bytes)=285409280  
Shuffle Errors  
  BAD_ID=0  
  CONNECTION=0  
  IO_ERROR=0  
  WRONG_LENGTH=0  
  WRONG_MAP=0  
  WRONG_REDUCE=0  
File Input Format Cour acadmild@localhost:~  
acadmild@localhost:~ [acadmild@localhost:~] eclipse-workspace -...
```

```
hadoop fs -ls comp/company12  
hadoop fs -cat comp/company12/part-r-00000
```



```
Applications Places System Mon Feb 19, 6:05 PM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=717
File Output Format Counters
  Bytes Written=43
[acadgild@localhost ~]$ hadoop fs -ls comp/company12
18/02/19 18:03:17 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
Found 3 items
-rw-r--r--  1 acadgild supergroup          0 2018-02-19 18:02 comp/company12/_SUCCESS
-rw-r--r--  1 acadgild supergroup        19 2018-02-19 18:02 comp/company12/part-r-00000
-rw-r--r--  1 acadgild supergroup        24 2018-02-19 18:02 comp/company12/part-r-00001
[acadgild@localhost ~]$ hadoop fs -cat comp/company12/part-r-00000
18/02/19 18:03:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
NA      1
Onida   4
Zen     2
[acadgild@localhost ~]$ hadoop fs -cat comp/company12/part-r-00001
18/02/19 18:04:01 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
Akai    1
Lava    3
Samsung 7
[acadgild@localhost ~]$
```

Task 3:

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

Driver Code

```
package Company;

import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;

import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;

public class company
{
    public static void main(String[] args) throws Exception
    {
        Configuration conf = new Configuration();
```

```

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

        job.setJarByClass(company.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(companymapper.class); // mapper class
        job.setReducerClass(companyreducer.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);
    }
}

```

Mapper Code

```

package Company;

import java.io.IOException;

import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;

public class companymapper extends Mapper<LongWritable, Text, Text, IntWritable>
{
    public 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 Linearray
        if(Linearray[0].equals("Onida")) // checking the word Onida in the
linearray[0], if it is Onida print the state name in linearray[3]and unit value
        {
            Text State = new Text(Linearray[3]);
            IntWritable unit= new IntWritable(1);
            context.write(State, unit);
        }
    }
}

```

```
}
```

Reducer Code

```
package Company;

import java.io.IOException;

import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;

public class companyreducer extends Reducer<Text, IntWritable, Text, IntWritable>
{
    public void reduce(Text State, Iterable<IntWritable> values, Context
context) throws IOException, InterruptedException
    {
        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 of units stored in the sum
    }
}
```

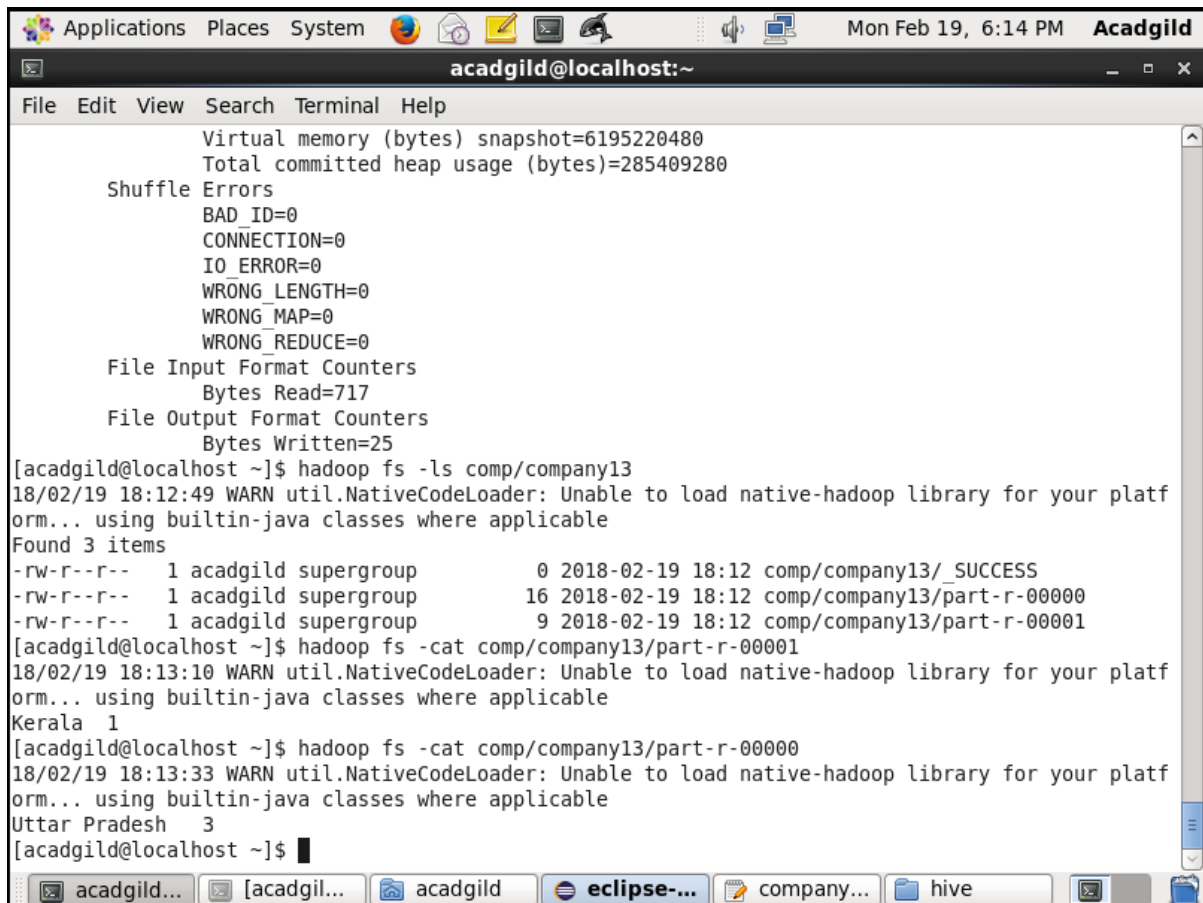
Output:

```
Hadoop jar company.jar Company.company comp/company.txt comp/company13
```

```
Applications Places System Mon Feb 19, 6:28 PM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhost ~]$ hadoop jar company.jar Company.company comp/company.txt comp/company13
18/02/19 18:11:35 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
18/02/19 18:11:37 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/02/19 18:11:39 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not perf
ormed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/02/19 18:11:40 INFO input.FileInputFormat: Total input files to process : 1
18/02/19 18:11:40 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1252)
    at java.lang.Thread.join(Thread.java:1326)
    at org.apache.hadoop.hdfs.DataStreamer.closeResponder(DataStreamer.java:980)
    at org.apache.hadoop.hdfs.DataStreamer.endBlock(DataStreamer.java:630)
    at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:807)
18/02/19 18:11:40 INFO mapreduce.JobSubmitter: number of splits:1
18/02/19 18:11:41 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.e
nabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
18/02/19 18:11:41 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1519042378766_0003
18/02/19 18:11:42 INFO impl.YarnClientImpl: Submitted application application_1519042378766_0003
18/02/19 18:11:42 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/appli
cation_1519042378766_0003/
18/02/19 18:11:42 INFO mapreduce.Job: Running job: job_1519042378766_0003
18/02/19 18:12:00 INFO mapreduce.Job: Job job_1519042378766_0003 running in uber mode : false
18/02/19 18:12:00 INFO mapreduce.Job: map 0% reduce 0%
18/02/19 18:12:14 INFO mapreduce.Job: map 100% reduce 0%
18/02/19 18:12:37 INFO mapreduce.Job: map 100% reduce 50%
18/02/19 18:12:38 INFO mapreduce.Job: map 100% reduce 100%
18/02/19 18:12:38 INFO mapreduce.Job: Job job_1519042378766_0003 completed successfully
acadgild@localhost:~ [acadgild@localhost:~] eclipse-workspace -...
```

```
Applications Places System Mon Feb 19, 6:29 PM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
18/02/19 18:12:38 INFO mapreduce.Job: Counters: 50
  File System Counters
    FILE: Number of bytes read=85
    FILE: Number of bytes written=607725
    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=9
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=4
  Job Counters
    Killed reduce tasks=1
    Launched map tasks=1
    Launched reduce tasks=2
    Data-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=11166
    Total time spent by all reduces in occupied slots (ms)=38888
    Total time spent by all map tasks (ms)=11166
    Total time spent by all reduce tasks (ms)=38888
    Total vcore-milliseconds taken by all map tasks=11166
    Total vcore-milliseconds taken by all reduce tasks=38888
    Total megabyte-milliseconds taken by all map tasks=11433984
    Total megabyte-milliseconds taken by all reduce tasks=39821312
  Map-Reduce Framework
    Map input records=18
    Map output records=4
    Map output bytes=65
eclipse-workspace - MapReduce/src/
Company/company.java - Eclipse
acadgild@localhost:~ [acadgild@localhost:~] eclipse-workspace -...
```

```
hadoop fs -ls comp/company13
hadoop fs -cat comp/company13/part-r-00000
```



```
Applications Places System Mon Feb 19, 6:14 PM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
Virtual memory (bytes) snapshot=6195220480
Total committed heap usage (bytes)=285409280
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=717
File Output Format Counters
Bytes Written=25
[acadgild@localhost ~]$ hadoop fs -ls comp/company13
18/02/19 18:12:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
Found 3 items
-rw-r--r-- 1 acadgild supergroup 0 2018-02-19 18:12 comp/company13/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 16 2018-02-19 18:12 comp/company13/part-r-00000
-rw-r--r-- 1 acadgild supergroup 9 2018-02-19 18:12 comp/company13/part-r-00001
[acadgild@localhost ~]$ hadoop fs -cat comp/company13/part-r-00001
18/02/19 18:13:10 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
Kerala 1
[acadgild@localhost ~]$ hadoop fs -cat comp/company13/part-r-00000
18/02/19 18:13:33 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
Uttar Pradesh 3
[acadgild@localhost ~]$
```