Minor project-Telecom problem

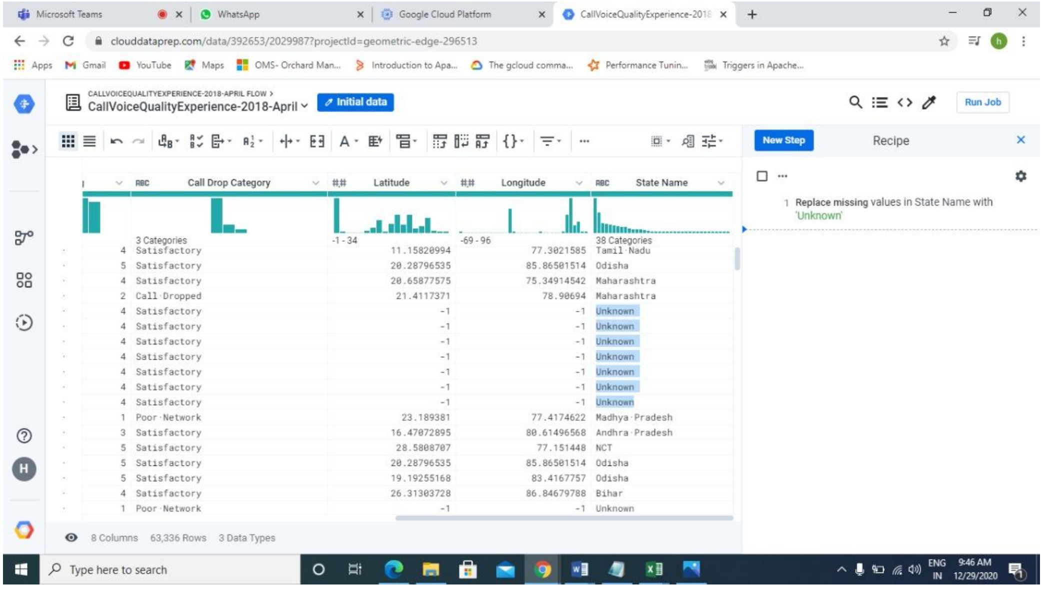
Telecom Dataset

Use-case: Customers Feedback using TRAI (Telecom regulatory authority of India) MyCall App.

Customers rates their experience about voice call quality in real time and help TRAI gather customer experience data along with Network data.

Problem Statement:

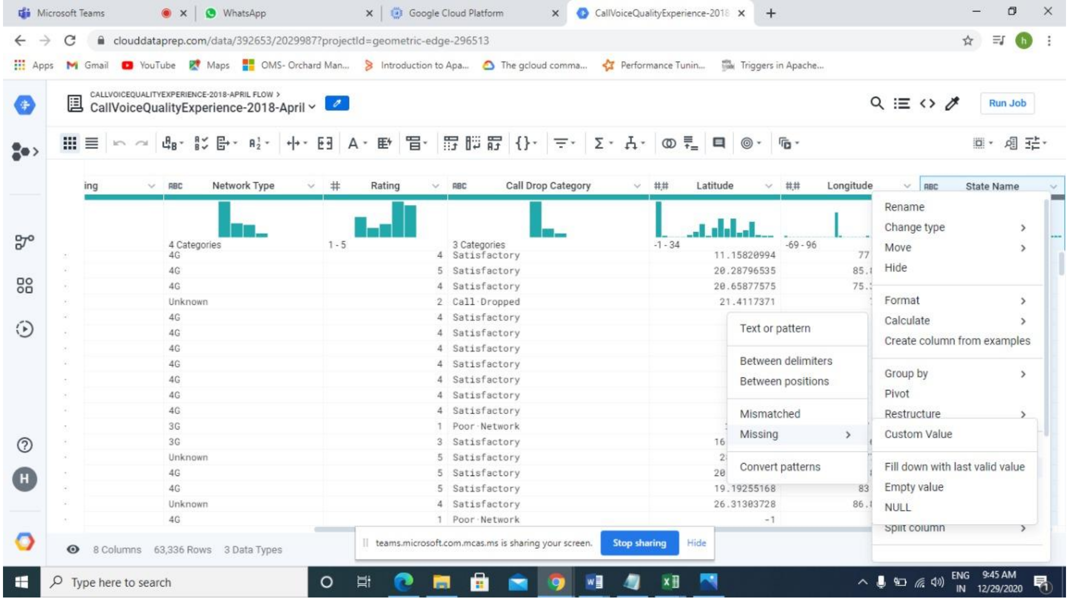
1. Clean and transform data for proper processing and getting complete insights without any garbage values [Data Prep]


Microsoft Teams
WhatsApp
Google Cloud Platform
CallVoiceQualityExperience-2018
clouddataprep.com/data/392653/2029987?projectId=geometric-edge-296513
Apps Gmail YouTube Maps OMS- Orchard Man... Introduction to Apa... The gcloud comma... Performance Tunin... Triggers in Apache...
CALLVOICEQUALITYEXPERIENCE-2018-APRIL FLOW >
CallVoiceQualityExperience-2018-April
Initial data
Run Job
New Step
Recipe
Type here to search
ENG IN 9:46 AM 12/29/2020

1 Replace missing values in State Name with 'Unknown'

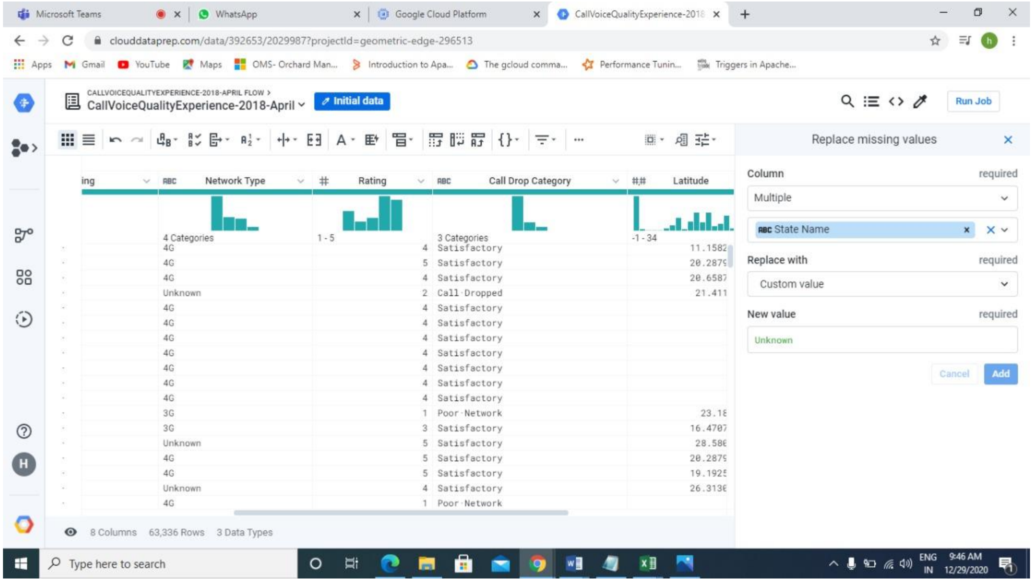
	Call Drop Category	Latitude	Longitude	State Name
	3 Categories	-1 - 34	-69 - 96	38 Categories
4	Satisfactory	11.15820994	77.3021585	Tamil Nadu
5	Satisfactory	20.28796535	85.86501514	Odisha
4	Satisfactory	20.65877575	75.34914542	Maharashtra
2	Call Dropped	21.4117371	78.90694	Maharashtra
4	Satisfactory	-1	-1	Unknown
4	Satisfactory	-1	-1	Unknown
4	Satisfactory	-1	-1	Unknown
4	Satisfactory	-1	-1	Unknown
4	Satisfactory	-1	-1	Unknown
4	Satisfactory	-1	-1	Unknown
4	Satisfactory	-1	-1	Unknown
1	Poor Network	23.189381	77.4174622	Madhya Pradesh
3	Satisfactory	16.47072895	80.61496568	Andhra Pradesh
5	Satisfactory	28.5808707	77.151448	NCT
5	Satisfactory	20.28796535	85.86501514	Odisha
5	Satisfactory	19.19255168	83.4167757	Odisha
4	Satisfactory	26.31303728	86.84679788	Bihar
1	Poor Network	-1	-1	Unknown

8 Columns 63,336 Rows 3 Data Types


Microsoft Teams
WhatsApp
Google Cloud Platform
CallVoiceQualityExperience-2018
clouddataprep.com/data/392653/2029987?projectId=geometric-edge-296513
Apps Gmail YouTube Maps OMS- Orchard Man... Introduction to Apa... The gcloud comma... Performance Tunin... Triggers in Apache...
CALLVOICEQUALITYEXPERIENCE-2018-APRIL FLOW >
CallVoiceQualityExperience-2018-April
Run Job
Rename
Change type
Move
Hide
Format
Calculate
Create column from examples
Group by
Pivot
Restructure
Split column
Text or pattern
Between delimiters
Between positions
Mismatched
Missing
Convert patterns
Custom Value
Fill down with last valid value
Empty value
NULL
teams.microsoft.com.mcas.ms is sharing your screen.
Stop sharing
Hide
Type here to search
ENG IN 9:45 AM 12/29/2020

ing	Network Type	Rating	Call Drop Category	Latitude	Longitude	State Name
	4 Categories	1 - 5	3 Categories	-1 - 34	-69 - 96	
	4G	4	Satisfactory	11.15820994	77	
	4G	5	Satisfactory	20.28796535	85.	
	4G	4	Satisfactory	20.65877575	75.	
	Unknown	2	Call Dropped	21.4117371		
	4G	4	Satisfactory			
	4G	4	Satisfactory			
	4G	4	Satisfactory			
	4G	4	Satisfactory			
	4G	4	Satisfactory			
	4G	4	Satisfactory			
	4G	4	Satisfactory			
	3G	1	Poor Network			
	3G	3	Satisfactory	16		
	Unknown	5	Satisfactory	2		
	4G	5	Satisfactory	20		
	4G	5	Satisfactory	19.19255168	83	
	Unknown	4	Satisfactory	26.31303728	86.	
	4G	1	Poor Network	-1		

8 Columns 63,336 Rows 3 Data Types



CALLVOICEQUALITYEXPERIENCE-2018-APRIL FLOW >
CallVoiceQualityExperience-2018-April
Initial data

ing	Network Type	Rating	Call Drop Category	Latitude
	4 Categories	1 - 5	3 Categories	-1 - 34
	4G	4	Satisfactory	11.1582
	4G	5	Satisfactory	20.2879
	4G	4	Satisfactory	20.6587
	Unknown	2	Call Dropped	21.411
	4G	4	Satisfactory	
	4G	4	Satisfactory	
	4G	4	Satisfactory	
	4G	4	Satisfactory	
	4G	4	Satisfactory	
	4G	4	Satisfactory	
	4G	4	Satisfactory	
	3G	1	Poor Network	23.18
	3G	3	Satisfactory	16.4707
	Unknown	5	Satisfactory	28.586
	4G	5	Satisfactory	20.2879
	4G	5	Satisfactory	19.1925
	Unknown	4	Satisfactory	26.3136
	4G	1	Poor Network	

8 Columns 63,336 Rows 3 Data Types

Replace missing values

Column required
Multiple
State Name

Replace with required
Custom value

New value required
Unknown

Cancel Add

2) List the states having poor connectivity for 4g network of jio for both indoor and outdoor conditions [Appache beam with sql]

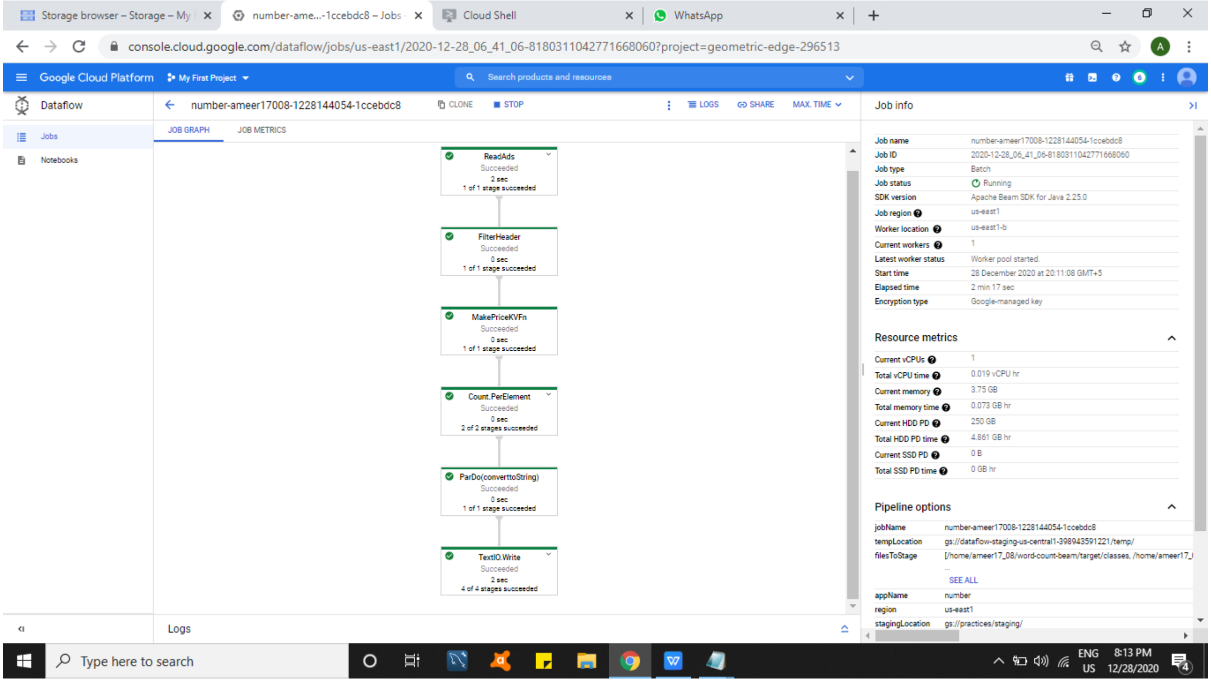
Code:

```
package org.apache.beam.examples.project;


import org.apache.beam.sdk.schemas.Schema;

import java.util.stream.Collectors;

import java.util.*;
import org.apache.beam.sdk.Pipeline;
import org.apache.beam.sdk.extensions.sql.SqlTransform;
import org.apache.beam.sdk.options.PipelineOptions;
import org.apache.beam.sdk.options.PipelineOptionsFactory;
import org.apache.beam.sdk.io.TextIO;
import org.apache.beam.sdk.transforms.*;
import org.apache.beam.sdk.values.KV;
import org.apache.beam.sdk.values.PCollection;
import org.apache.beam.sdk.transforms.Sum;
import org.apache.beam.sdk.values.Row;

public class poorconnectivity {
   private static final String CSV_HEADER ="Operator,Indoor_Outdoor_Travelling,Network Type,Rating,Call Drop Category,Latitude,Longitude,State Name";
   public static List<String> stateArray = new ArrayList<String>();
```

```
  private static final  Schema schema = Schema.builder()
        .addStringField("Operator")
        .addStringField("Indoor_Outdoor_Travelling")
        .addStringField("Network_Type")
        .addStringField("Rating")
        .addStringField("Call_Drop_Category")
        .addStringField("Latitude")
        .addStringField("Longitude")
        .addStringField("State_Name")
        .build();

  public static void main(String[] args) {



    PipelineOptions options =
PipelineOptionsFactory.fromArgs(args).withValidation().create();
    Pipeline pipeline = Pipeline.create(options);

  PCollection<Row> cvrt =
pipeline.apply(TextIO.read().from("src/main/project_resources/CallVoiceQualityExperience-
2018-April.csv"))
  .apply(ParDo.of(new RowParDo())).setRowSchema(schema);

  PCollection<Row> out = cvrt.apply(
      SqlTransform.query("SELECT State_Name,Operator,Indoor_Outdoor_Travelling FROM
PCOLLECTION WHERE  Call_Drop_Category like 'Poor Network' and Network_Type='4G' and
Operator like 'RJio'  and Indoor_Outdoor_Travelling in('Outdoor','Indoor')  group by
State_Name, Operator,Indoor_Outdoor_Travelling ORDER BY State_Name limit 1000"));
      out.apply(ParDo.of(new RowToString()))
      .apply("fightdetails", ParDo.of(new fightKVFn()))
      .apply("MakeModelGrouping", GroupByKey.create())
      .apply("grouping",ParDo.of(new itera()))
      .apply(MapElements.via(new SimpleFunction<String, Void>() {
          @Override
          public Void apply(String input) {
            System.out.println(input);
            return null;
          }

        }));
  pipeline.run().waitUntilFinish();


  }
  public static class RowParDo extends DoFn<String, Row> {
    @ProcessElement
    public void processElement(ProcessContext c) {
      if (!c.element().equalsIgnoreCase(CSV_HEADER)) {
```

```
      String[] vals = c.element().split(",");

      Row appRow = Row
          .withSchema(schema)
          .addValues(vals[0],vals[1],vals[2],vals[3],vals[4],vals[5],vals[6],vals[7])
          .build();
      c.output(appRow);
      }

    }
  }



public static class RowToString extends DoFn<Row, String> {
  @ProcessElement
  public void processElement(ProcessContext c) {
    String line = c.element().getValues()
        .stream()
        .map(Object::toString)
        .collect(Collectors.joining(","));
    c.output(line);
  }
}
private static class fightKVFn extends DoFn<String, KV<String, Integer>> {

  @ProcessElement
  public void processElement(
      @Element String element,
      OutputReceiver<KV<String, Integer>> out) {
    String[] fields = element.split(",");
    out.output(KV.of(fields[0], 1));
    }
  }
private static class itera extends DoFn<KV<String,Iterable<Integer>>,String> {

  @ProcessElement
  public void processElement(
      @Element KV<String,Iterable<Integer>> element,
      OutputReceiver<String> out) {
     int sum_result=0;

    String state = element.getKey();
       for(Integer val:element.getValue())
               sum_result+=val;

    if(sum_result>1){
    out.output(state);
```

```
            }
        }
    }
}
```



Job info

Job name	number-ameer17008-1228144054-1ccebdc8
Job ID	2020-12-28_06_41_06-8180311042771668060
Job type	Batch
Job status	Running
SDK version	Apache Beam SDK for Java 2.25.0
Job region	us-east1
Worker location	us-east1-b
Current workers	1
Latest worker status	Worker pool started.
Start time	28 December 2020 at 20:11:08 GMT+5
Elapsed time	2 min 17 sec
Encryption type	Google-managed key

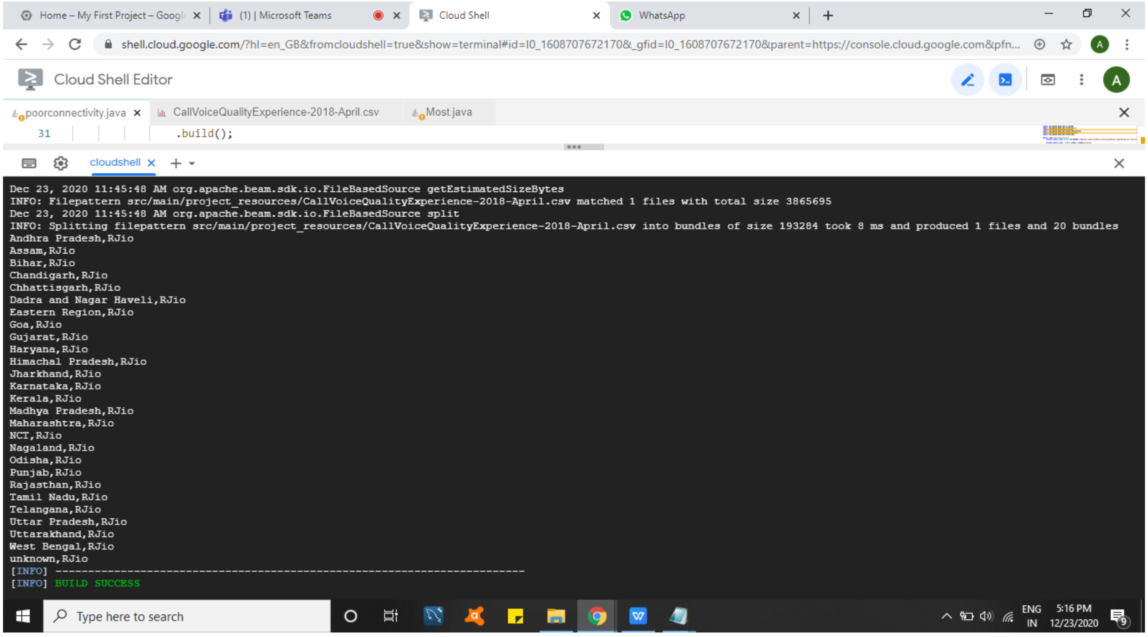
Resource metrics

Current vCPUs	1
Total vCPU time	0.019 vCPU hr
Current memory	3.75 GB
Total memory time	0.073 GB hr
Current HDD PD	250 GB
Total HDD PD time	4.861 GB hr
Current SSD PD	0 B
Total SSD PD time	0 GB hr

Pipeline options

jobName	number-ameer17008-1228144054-1ccebdc8
tempLocation	gs://dataflow-staging-us-central1-398943591221/temp/
filesToStage	[/home/ameer17_08/word-count-beam/target/classes, /home/ameer17_
appName	number
region	us-east1
stagingLocation	gs://practices/staging/

Output:



shell.cloud.google.com/?hl=en_GB&fromcloudshell=true&show=terminal#id=I0_1608707672170&_gfid=I0_1608707672170&parent=https://console.cloud.google.com&pfn...
Cloud Shell Editor
poorconnectivity.java
CallVoiceQualityExperience-2018-April.csv
Most.java
.build();
cloudshell

```
Dec 23, 2020 11:45:48 AM org.apache.beam.sdk.io.FileBasedSource getEstimatedSizeBytes
INFO: Filepattern src/main/project_resources/CallVoiceQualityExperience-2018-April.csv matched 1 files with total size 3865695
Dec 23, 2020 11:45:48 AM org.apache.beam.sdk.io.FileBasedSource split
INFO: Splitting filepattern src/main/project_resources/CallVoiceQualityExperience-2018-April.csv into bundles of size 193284 took 8 ms and produced 1 files and 20 bundles
Andhra Pradesh,RJio
Assam,RJio
Bihar,RJio
Chandigarh,RJio
Chhattisgarh,RJio
Dadra and Nagar Haveli,RJio
Eastern Region,RJio
Goa,RJio
Gujarat,RJio
Haryana,RJio
Himachal Pradesh,RJio
Jharkhand,RJio
Karnataka,RJio
Kerala,RJio
Madhya Pradesh,RJio
Maharashtra,RJio
NCT,RJio
Nagaland,RJio
Odisha,RJio
Punjab,RJio
Rajasthan,RJio
Tamil Nadu,RJio
Telangana,RJio
Uttar Pradesh,RJio
Uttarakhand,RJio
West Bengal,RJio
unknown,RJio
[INFO] ------------------------------------------------------------------------
[INFO] BUILD SUCCESS
```

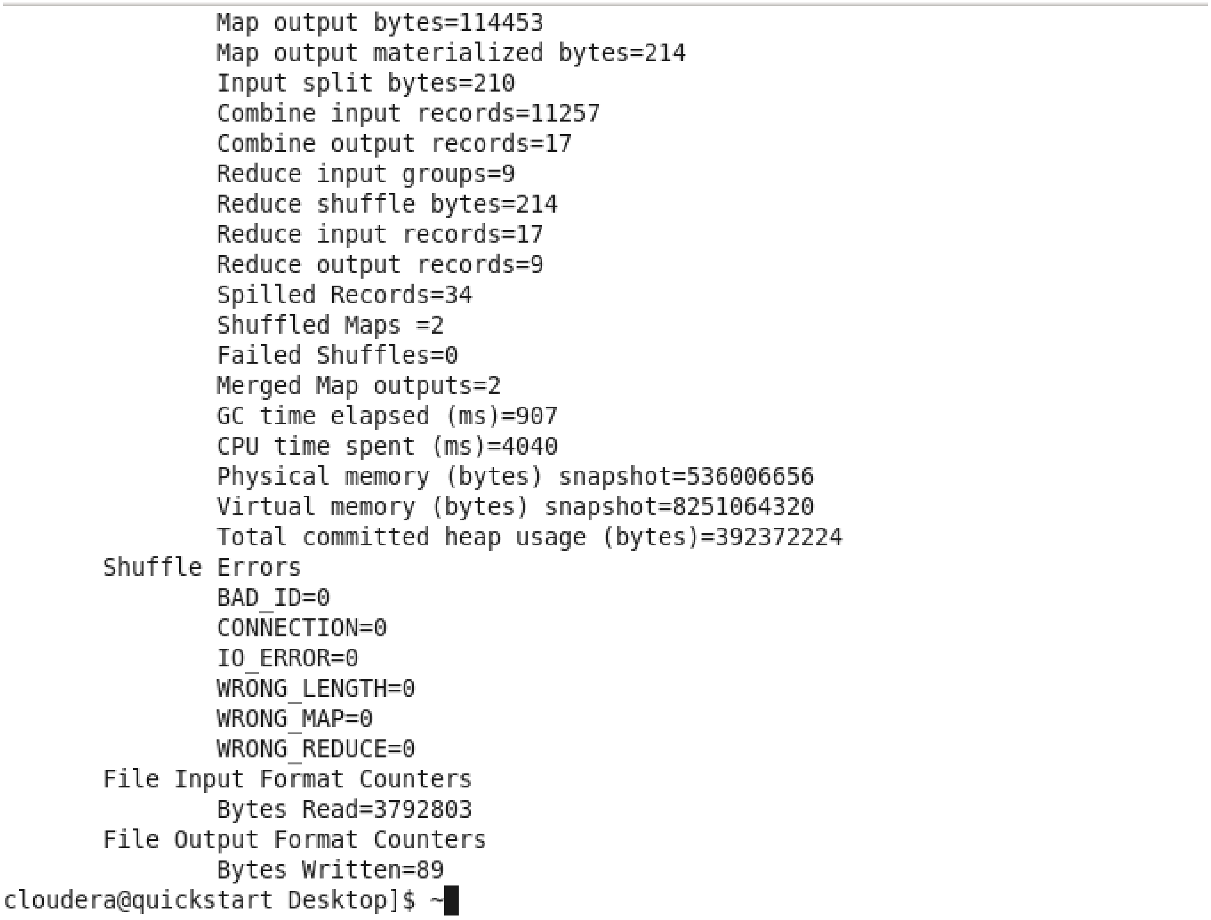
3.Find out the number of poor network compliants registered to each network providers [Mapreduce]

Code:

```
package mypack;

import java.io.IOException;
import java.util.Iterator;

import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.FileInputFormat;
import org.apache.hadoop.mapred.FileOutputFormat;
import org.apache.hadoop.mapred.JobClient;
import org.apache.hadoop.mapred.JobConf;
import org.apache.hadoop.mapred.MapReduceBase;
import org.apache.hadoop.mapred.Mapper;
import org.apache.hadoop.mapred.OutputCollector;
import org.apache.hadoop.mapred.Reducer;
import org.apache.hadoop.mapred.Reporter;
import org.apache.hadoop.mapred.TextInputFormat;
import org.apache.hadoop.mapred.TextOutputFormat;

public class location {
        public static class Map extends MapReduceBase implements   Mapper<LongWritable , Text, Text,
IntWritable>
          {

            public void map(LongWritable key, Text value,  OutputCollector<Text, IntWritable> output,
Reporter reporter) throws IOException {
              String line = value.toString();
              String [] count=line.split(",");
              if(count[4].equals("Poor Network")){
             output.collect(new Text(count[0]), new IntWritable(1));
              }
            }
          }

public static class Reduce extends MapReduceBase implements Reducer< Text, IntWritable, Text,
IntWritable > {


                     public void reduce( Text key, Iterator <IntWritable> values,
                     OutputCollector<Text, IntWritable> output, Reporter reporter) throws IOException {
                         int sum_result=0;
                       while (values.hasNext()) {
                                    sum_result+=values.next().get();
                         }
                       output.collect(key, new IntWritable(sum_result));
                     }
                   }
         public static void main(String args[])throws Exception {
            JobConf conf = new JobConf(location.class);

            conf.setJobName("max_count");
            conf.setOutputKeyClass(Text.class);
            conf.setOutputValueClass(IntWritable.class);
            conf.setMapperClass(Map.class);
            conf.setCombinerClass(Reduce.class);
            conf.setReducerClass(Reduce.class);
            conf.setInputFormat(TextInputFormat.class);
```

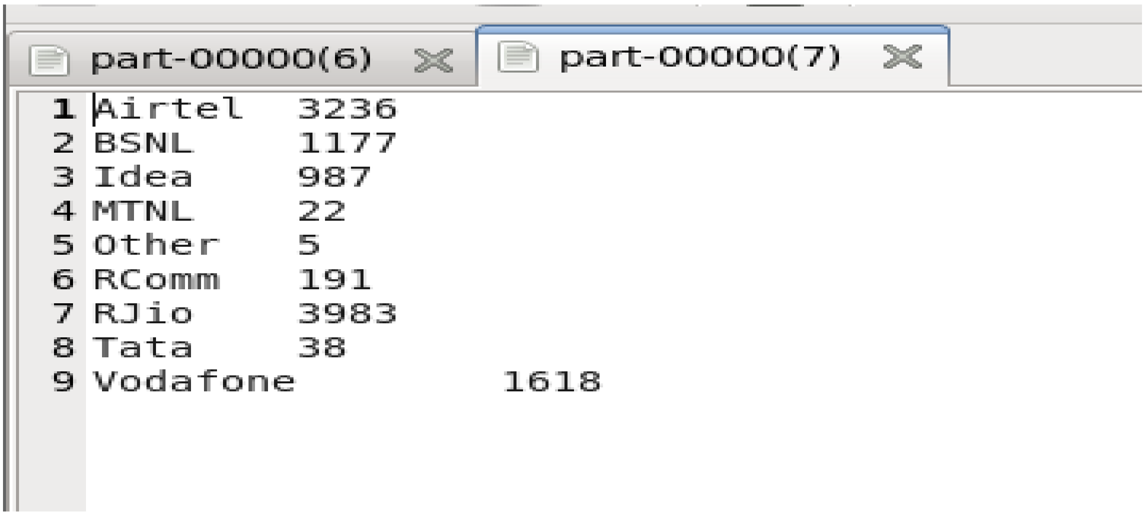
```
		conf.setOutputFormat(TextOutputFormat.class);

		FileInputFormat.setInputPaths(conf, new Path(args[0]));
		FileOutputFormat.setOutputPath(conf, new Path(args[1]));

		JobClient.runJob(conf);
	}


}
```

Execution :



```
		Map output bytes=114453
		Map output materialized bytes=214
		Input split bytes=210
		Combine input records=11257
		Combine output records=17
		Reduce input groups=9
		Reduce shuffle bytes=214
		Reduce input records=17
		Reduce output records=9
		Spilled Records=34
		Shuffled Maps =2
		Failed Shuffles=0
		Merged Map outputs=2
		GC time elapsed (ms)=907
		CPU time spent (ms)=4040
		Physical memory (bytes) snapshot=536006656
		Virtual memory (bytes) snapshot=8251064320
		Total committed heap usage (bytes)=392372224
	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=3792803
	File Output Format Counters
		Bytes Written=89
cloudera@quickstart Desktop]$ ~
```

Output:



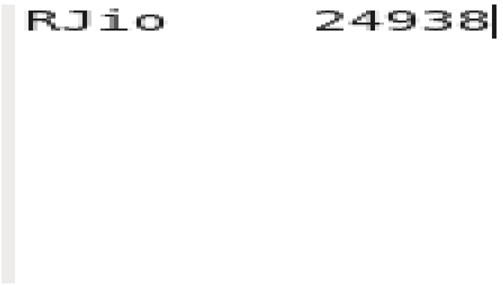
```
Airtel	3236
BSNL	1177
Idea	987
MTNL	22
Other	5
RComm	191
RJio	3983
Tata	38
Vodafone		1618
```

4) Which company has highest 4g locations. [Pig]

Program:

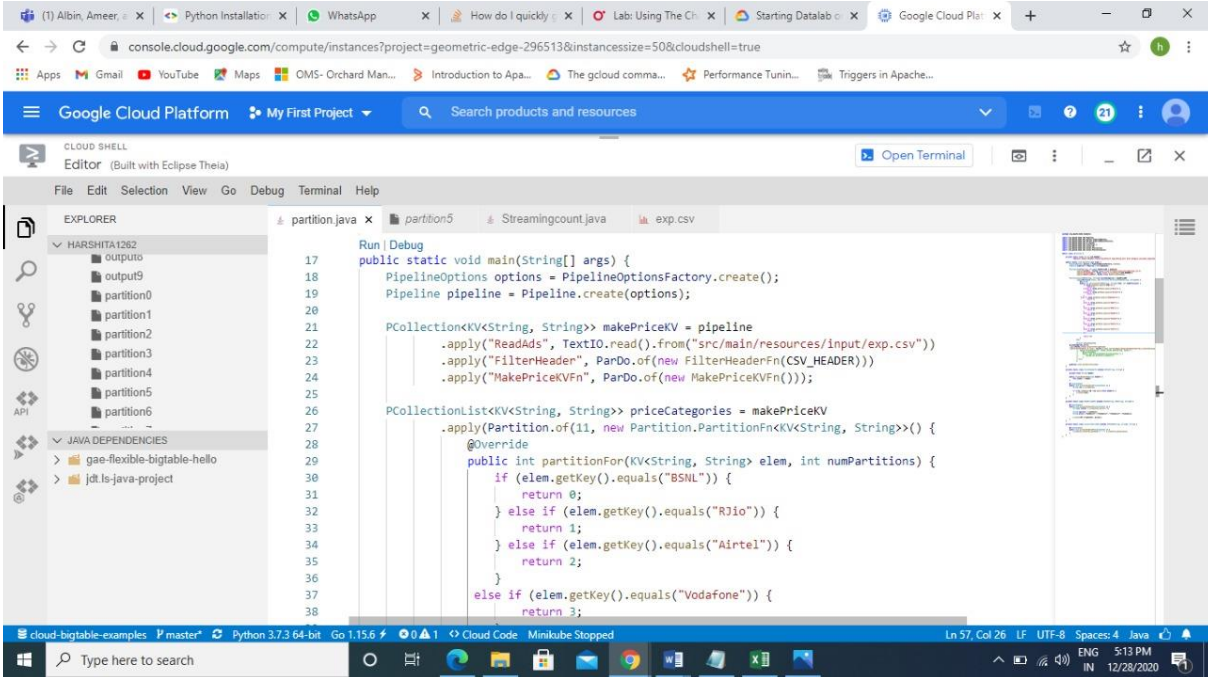
```
    file =load '/home/cloudera/Desktop/CQE.csv' using PigStorage(',') as
(Operator,Indoor_Outdoor_Travelling,NetworkType,Rating,CallDropCategory,Latitude,Longitude,StateName);
filterdata= FILTER file BY NetworkType == '4G';

result_group =group filterdata by Operator;

result_grouping= foreach result_group generate group,COUNT(filterdata.NetworkType)as
compliant;

final_result = ORDER result_grouping BY compliant DESC;

final = LIMIT final_result 1;
store final into 'secondprogram4561';

dump final;
```

output:

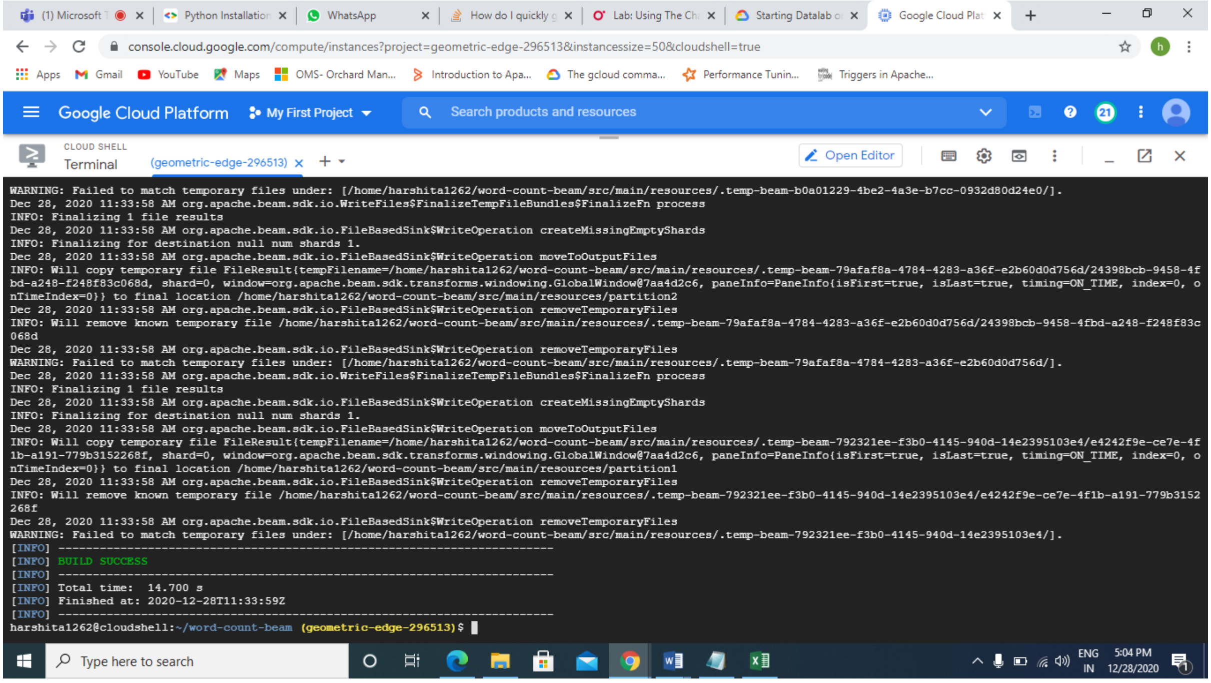


```
RJio      24938
```

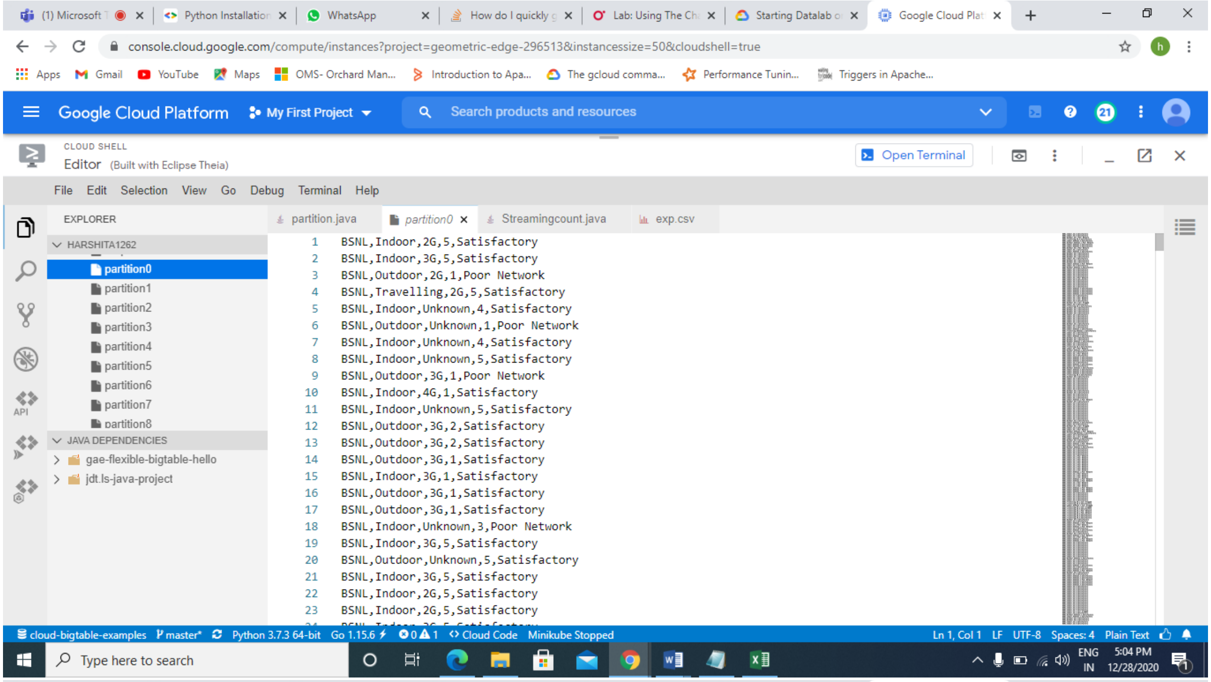
5.Partition data by Providers--->Apache beam partion



```java
public static void main(String[] args) {
    PipelineOptions options = PipelineOptionsFactory.create();
    Pipeline pipeline = Pipeline.create(options);

    PCollection<KV<String, String>> makePriceKV = pipeline
            .apply("ReadAds", TextIO.read().from("src/main/resources/input/exp.csv"))
            .apply("FilterHeader", ParDo.of(new FilterHeaderFn(CSV_HEADER)))
            .apply("MakePriceKVFn", ParDo.of(new MakePriceKVFn()));

    PCollectionList<KV<String, String>> priceCategories = makePriceKV
            .apply(Partition.of(11, new Partition.PartitionFn<KV<String, String>>() {
                @Override
                public int partitionFor(KV<String, String> elem, int numPartitions) {
                    if (elem.getKey().equals("BSNL")) {
                        return 0;
                    } else if (elem.getKey().equals("RJio")) {
                        return 1;
                    } else if (elem.getKey().equals("Airtel")) {
                        return 2;
                    }
                 else if (elem.getKey().equals("Vodafone")) {
                        return 3;
```

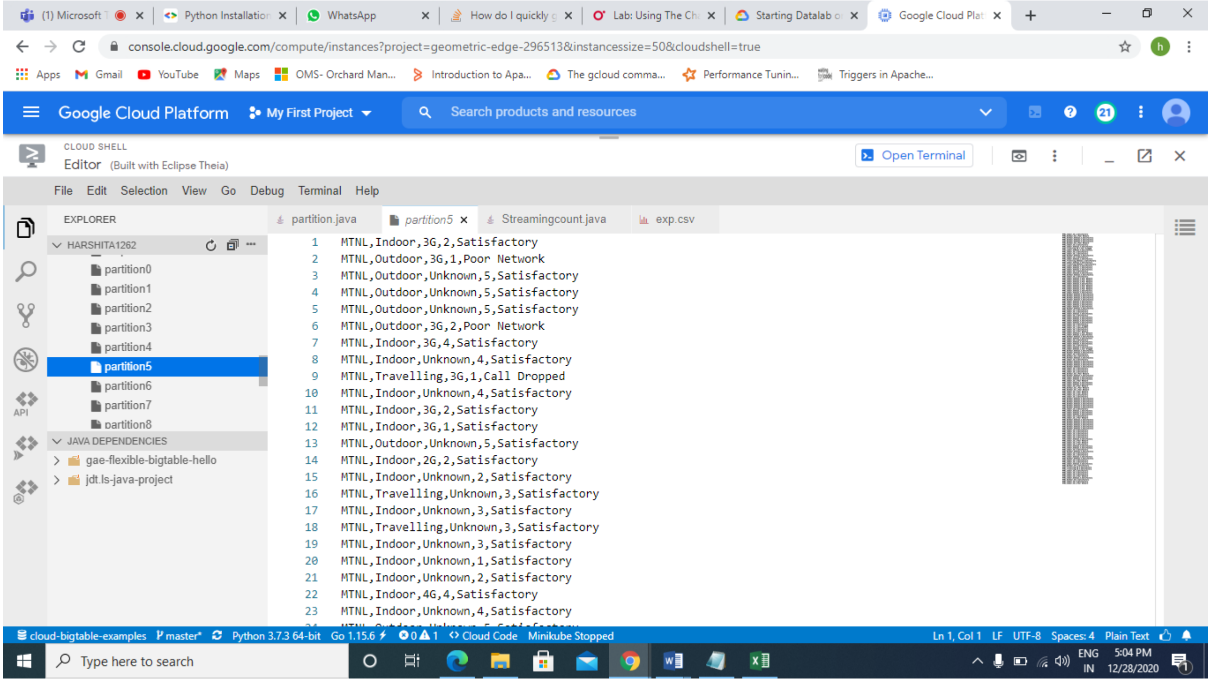

console.cloud.google.com/compute/instances?project=geometric-edge-296513&instancesize=50&cloudshell=true
Google Cloud Platform
My First Project
Search products and resources
CLOUD SHELL
Terminal
(geometric-edge-296513)
Open Editor
Type here to search
5:04 PM
12/28/2020

```
WARNING: Failed to match temporary files under: [/home/harshita1262/word-count-beam/src/main/resources/.temp-beam-b0a01229-4be2-4a3e-b7cc-0932d80d24e0/].
Dec 28, 2020 11:33:58 AM org.apache.beam.sdk.io.WriteFiles$FinalizeTempFileBundles$FinalizeFn process
INFO: Finalizing 1 file results
Dec 28, 2020 11:33:58 AM org.apache.beam.sdk.io.FileBasedSink$WriteOperation createMissingEmptyShards
INFO: Finalizing for destination null num shards 1.
Dec 28, 2020 11:33:58 AM org.apache.beam.sdk.io.FileBasedSink$WriteOperation moveToOutputFiles
INFO: Will copy temporary file FileResult{tempFilename=/home/harshita1262/word-count-beam/src/main/resources/.temp-beam-79afaf8a-4784-4283-a36f-e2b60d0d756d/24398bcb-9458-4f
bd-a248-f248f83c068d, shard=0, window=org.apache.beam.sdk.transforms.windowing.GlobalWindow@7aa4d2c6, paneInfo=PaneInfo{isFirst=true, isLast=true, timing=ON_TIME, index=0, o
nTimeIndex=0}} to final location /home/harshita1262/word-count-beam/src/main/resources/partition2
Dec 28, 2020 11:33:58 AM org.apache.beam.sdk.io.FileBasedSink$WriteOperation removeTemporaryFiles
INFO: Will remove known temporary file /home/harshita1262/word-count-beam/src/main/resources/.temp-beam-79afaf8a-4784-4283-a36f-e2b60d0d756d/24398bcb-9458-4fbd-a248-f248f83c
068d
Dec 28, 2020 11:33:58 AM org.apache.beam.sdk.io.FileBasedSink$WriteOperation removeTemporaryFiles
WARNING: Failed to match temporary files under: [/home/harshita1262/word-count-beam/src/main/resources/.temp-beam-79afaf8a-4784-4283-a36f-e2b60d0d756d/].
Dec 28, 2020 11:33:58 AM org.apache.beam.sdk.io.WriteFiles$FinalizeTempFileBundles$FinalizeFn process
INFO: Finalizing 1 file results
Dec 28, 2020 11:33:58 AM org.apache.beam.sdk.io.FileBasedSink$WriteOperation createMissingEmptyShards
INFO: Finalizing for destination null num shards 1.
Dec 28, 2020 11:33:58 AM org.apache.beam.sdk.io.FileBasedSink$WriteOperation moveToOutputFiles
INFO: Will copy temporary file FileResult{tempFilename=/home/harshita1262/word-count-beam/src/main/resources/.temp-beam-792321ee-f3b0-4145-940d-14e2395103e4/e4242f9e-ce7e-4f
1b-a191-779b3152268f, shard=0, window=org.apache.beam.sdk.transforms.windowing.GlobalWindow@7aa4d2c6, paneInfo=PaneInfo{isFirst=true, isLast=true, timing=ON_TIME, index=0, o
nTimeIndex=0}} to final location /home/harshita1262/word-count-beam/src/main/resources/partition1
Dec 28, 2020 11:33:58 AM org.apache.beam.sdk.io.FileBasedSink$WriteOperation removeTemporaryFiles
INFO: Will remove known temporary file /home/harshita1262/word-count-beam/src/main/resources/.temp-beam-792321ee-f3b0-4145-940d-14e2395103e4/e4242f9e-ce7e-4f1b-a191-779b3152
268f
Dec 28, 2020 11:33:58 AM org.apache.beam.sdk.io.FileBasedSink$WriteOperation removeTemporaryFiles
WARNING: Failed to match temporary files under: [/home/harshita1262/word-count-beam/src/main/resources/.temp-beam-792321ee-f3b0-4145-940d-14e2395103e4/].
[INFO] ------------------------------------------------------------------------
[INFO] BUILD SUCCESS
[INFO] ------------------------------------------------------------------------
[INFO] Total time:  14.700 s
[INFO] Finished at: 2020-12-28T11:33:59Z
[INFO] ------------------------------------------------------------------------
harshita1262@cloudshell:~/word-count-beam (geometric-edge-296513)$
```



```
BSNL,Indoor,2G,5,Satisfactory
BSNL,Indoor,3G,5,Satisfactory
BSNL,Outdoor,2G,1,Poor Network
BSNL,Travelling,2G,5,Satisfactory
BSNL,Indoor,Unknown,4,Satisfactory
BSNL,Outdoor,Unknown,1,Poor Network
BSNL,Indoor,Unknown,4,Satisfactory
BSNL,Indoor,Unknown,5,Satisfactory
BSNL,Outdoor,3G,1,Poor Network
BSNL,Indoor,4G,1,Satisfactory
BSNL,Indoor,Unknown,5,Satisfactory
BSNL,Outdoor,3G,2,Satisfactory
BSNL,Outdoor,3G,2,Satisfactory
BSNL,Outdoor,3G,1,Satisfactory
BSNL,Indoor,3G,1,Satisfactory
BSNL,Outdoor,3G,1,Satisfactory
BSNL,Outdoor,3G,1,Satisfactory
BSNL,Indoor,Unknown,3,Poor Network
BSNL,Indoor,3G,5,Satisfactory
BSNL,Outdoor,Unknown,5,Satisfactory
BSNL,Indoor,3G,5,Satisfactory
BSNL,Indoor,2G,5,Satisfactory
BSNL,Indoor,2G,5,Satisfactory
```

cloud-bigtable-examples master* Python 3.7.3 64-bit Go 1.15.6 0 1 Cloud Code Minikube Stopped
Ln 1, Col 1 LF UTF-8 Spaces: 4 Plain Text
Type here to search
5:04 PM 12/28/2020

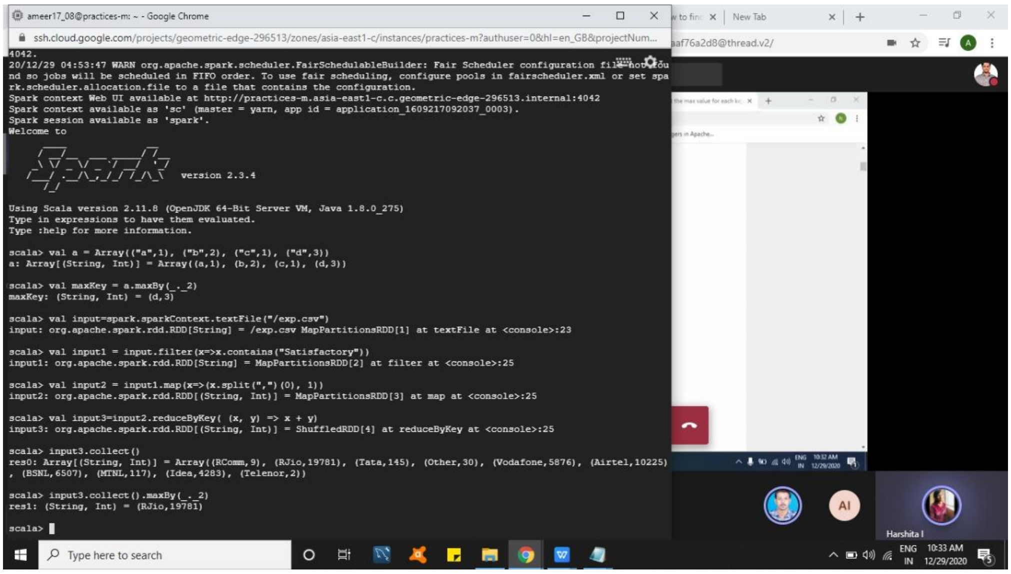


```
MTNL,Indoor,3G,2,Satisfactory
MTNL,Outdoor,3G,1,Poor Network
MTNL,Outdoor,Unknown,5,Satisfactory
MTNL,Outdoor,Unknown,5,Satisfactory
MTNL,Outdoor,Unknown,5,Satisfactory
MTNL,Outdoor,3G,2,Poor Network
MTNL,Indoor,3G,4,Satisfactory
MTNL,Indoor,Unknown,4,Satisfactory
MTNL,Travelling,3G,1,Call Dropped
MTNL,Indoor,Unknown,4,Satisfactory
MTNL,Indoor,3G,2,Satisfactory
MTNL,Indoor,3G,1,Satisfactory
MTNL,Outdoor,Unknown,5,Satisfactory
MTNL,Indoor,2G,2,Satisfactory
MTNL,Indoor,Unknown,2,Satisfactory
MTNL,Travelling,Unknown,3,Satisfactory
MTNL,Indoor,Unknown,3,Satisfactory
MTNL,Travelling,Unknown,3,Satisfactory
MTNL,Indoor,Unknown,3,Satisfactory
MTNL,Indoor,Unknown,1,Satisfactory
MTNL,Indoor,Unknown,2,Satisfactory
MTNL,Indoor,4G,4,Satisfactory
MTNL,Indoor,Unknown,4,Satisfactory
```

cloud-bigtable-examples master* Python 3.7.3 64-bit Go 1.15.6 0 1 Cloud Code Minikube Stopped
Ln 1, Col 1 LF UTF-8 Spaces: 4 Plain Text
Type here to search
5:04 PM 12/28/2020

6.List the network having most number of satisfactory calls.--->RDD

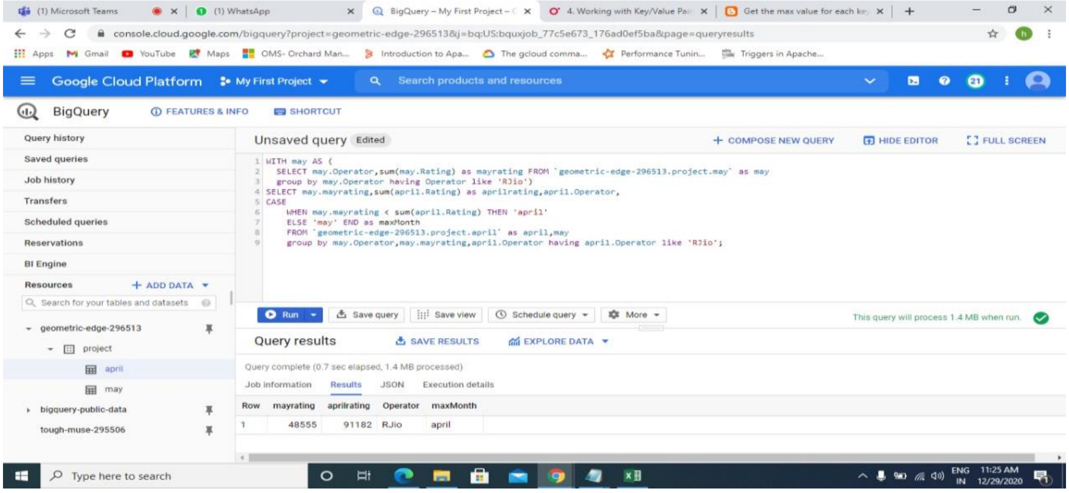
```
hdfs dfs -put /home/harshita1262/exp.csv /
val data1 =spark.sparkContext.textFile("/exp.csv")
val data3 =data1.filter(x=>x.contains("Satisfactory"))
val pairs = data3.map(x=>(x.split(",")(0), 1))
val groups=pairs.reduceByKey( (x, y) => x + y)
groups.collect().maxBy(_._2)
```



```
4042.
20/12/29 04:53:47 WARN org.apache.spark.scheduler.FairSchedulableBuilder: Fair Scheduler configuration file not fou
nd so jobs will be scheduled in FIFO order. To use fair scheduling, configure pools in fairscheduler.xml or set spa
rk.scheduler.allocation.file to a file that contains the configuration.
Spark context Web UI available at http://practices-m.asia-east1-c.c.geometric-edge-296513.internal:4042
Spark context available as 'sc' (master = yarn, app id = application_1609217092037_0003).
Spark session available as 'spark'.
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /___/ .__/\_,_/_/ /_/\_\   version 2.3.4
      /_/

Using Scala version 2.11.8 (OpenJDK 64-Bit Server VM, Java 1.8.0_275)
Type in expressions to have them evaluated.
Type :help for more information.

scala> val a = Array(("a",1), ("b",2), ("c",1), ("d",3))
a: Array[(String, Int)] = Array((a,1), (b,2), (c,1), (d,3))

scala> val maxKey = a.maxBy(_._2)
maxKey: (String, Int) = (d,3)

scala> val input=spark.sparkContext.textFile("/exp.csv")
input: org.apache.spark.rdd.RDD[String] = /exp.csv MapPartitionsRDD[1] at textFile at <console>:23

scala> val input1 = input.filter(x=>x.contains("Satisfactory"))
input1: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[2] at filter at <console>:25

scala> val input2 = input1.map(x=>(x.split(",")(0), 1))
input2: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[3] at map at <console>:25

scala> val input3=input2.reduceByKey( (x, y) => x + y)
input3: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[4] at reduceByKey at <console>:25

scala> input3.collect()
res0: Array[(String, Int)] = Array((RComm,9), (RJio,19781), (Tata,145), (Other,30), (Vodafone,5876), (Airtel,10225)
, (BSNL,6507), (MTNL,117), (Idea,4283), (Telenor,2))

scala> input3.collect().maxBy(_._2)
res1: (String, Int) = (RJio,19781)

scala> 
```

7) Compare the service provided by Jio in April and May based on User Rating. [BIGQUERY]

```
WITH may AS (
   SELECT may.Operator,sum(may.Rating) as mayrating FROM `geometric-edge-296513.project.may` as may
  group by may.Operator having Operator like 'RJio')
SELECT may.mayrating,sum(april.Rating) as aprilrating,april.Operator,
CASE
   WHEN may.mayrating < sum(april.Rating) THEN 'april'
   ELSE 'may' END as maxMonth
   FROM `geometric-edge-296513.project.april` as april,may
   group by may.Operator,may.mayrating,april.Operator having april.Operator like 'RJio';
```



Unsaved query Edited

```
WITH may AS (
  SELECT may.Operator,sum(may.Rating) as mayrating FROM `geometric-edge-296513.project.may` as may
  group by may.Operator having Operator like 'RJio')
SELECT may.mayrating,sum(april.Rating) as aprilrating,april.Operator,
CASE
    WHEN may.mayrating < sum(april.Rating) THEN 'april'
    ELSE 'may' END as maxMonth
    FROM `geometric-edge-296513.project.april` as april,may
    group by may.Operator,may.mayrating,april.Operator having april.Operator like 'RJio';
```

This query will process 1.4 MB when run.

Query results

Query complete (0.7 sec elapsed, 1.4 MB processed)

Row	mayrating	aprilrating	Operator	maxMonth
1	48555	91182	RJio	april

8 . Find out the most preffered network based on rating. [Hive]

Code:

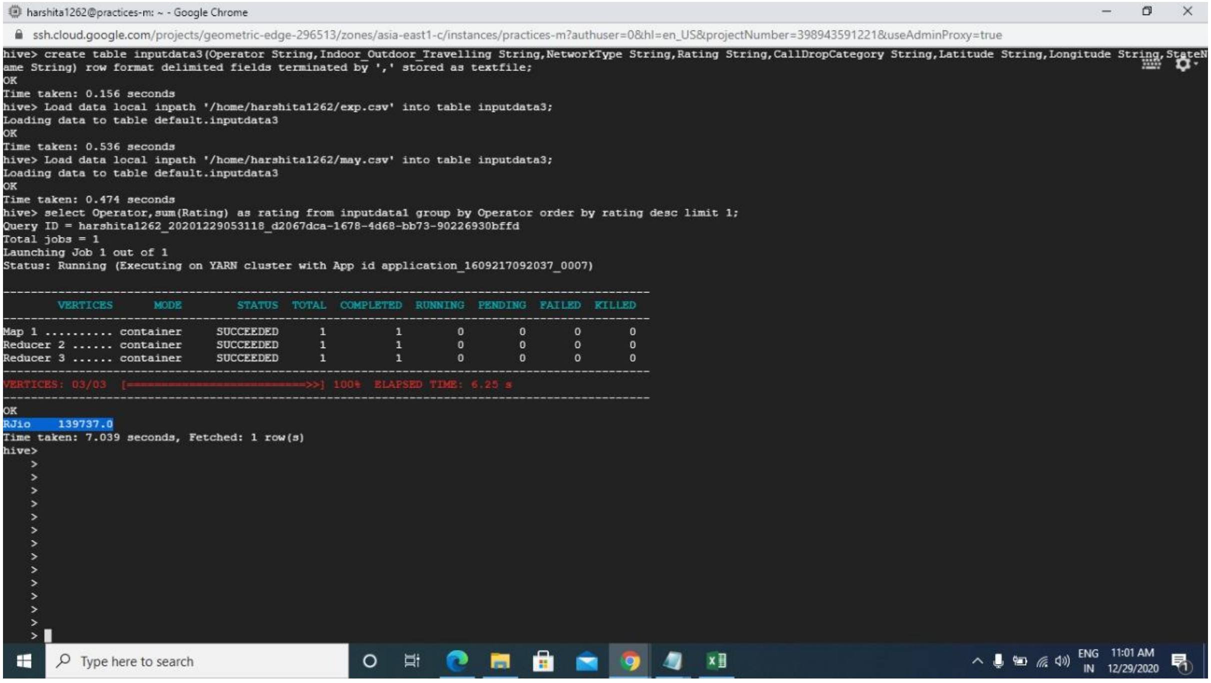
create table inputdata(Operator String,Indoor_Outdoor_Travelling String,NetworkType String,Rating String,CallDropCategory String,Latitude String,Longitude String,StateName String) row format delimited fields terminated by ',' stored as textfile

Load data local inpath '/home/harshita1262/exp.csv' into table inputdata1;

Load data local inpath '/home/harshita1262/may.csv' into table inputdata1

select Operator,sum(Rating) as rating from inputdata1 group by Operator order by rating desc limit 1;

output:


harshita1262@practices-m: ~ - Google Chrome
ssh.cloud.google.com/projects/geometric-edge-296513/zones/asia-east1-c/instances/practices-m?authuser=0&hl=en_US&projectNumber=398943591221&useAdminProxy=true
Type here to search
ENG
IN
11:01 AM
12/29/2020

```
hive> create table inputdata3(Operator String,Indoor_Outdoor_Travelling String,NetworkType String,Rating String,CallDropCategory String,Latitude String,Longitude String,StateN
ame String) row format delimited fields terminated by ',' stored as textfile;
OK
Time taken: 0.156 seconds
hive> Load data local inpath '/home/harshita1262/exp.csv' into table inputdata3;
Loading data to table default.inputdata3
OK
Time taken: 0.536 seconds
hive> Load data local inpath '/home/harshita1262/may.csv' into table inputdata3;
Loading data to table default.inputdata3
OK
Time taken: 0.474 seconds
hive> select Operator,sum(Rating) as rating from inputdata1 group by Operator order by rating desc limit 1;
Query ID = harshita1262_20201229053118_d2067dca-1678-4d68-bb73-90226930bffd
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1609217092037_0007)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      1          1        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
Reducer 3 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 6.25 s
----------------------------------------------------------------------------------------------
OK
RJio    139737.0
Time taken: 7.039 seconds, Fetched: 1 row(s)
hive>
    >
    >
    >
    >
    >
    >
    >
    >
    >
    >
    >
    >
    >
    >
```

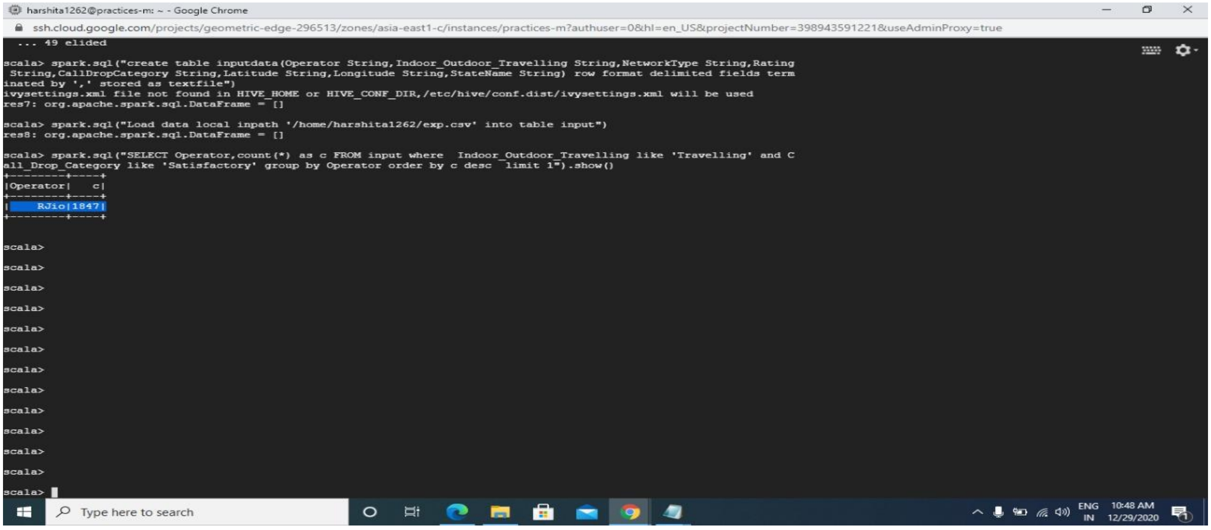
9.List the network having most number of satisfactory calls while the user was travelling.[Spark]

Code:

spark.sql("create table input(Operator string,Indoor_Outdoor_Travelling string,Network Type string,Rating Call Drop Category string,Latitude string,Longitude string,State Namestring)row format delimited fields terminated by ',' stored as textfile")

spark.sql("Load data local inpath 'exp.csv' into table input")

spark.sql("SELECT Operator,count(*) as c FROM input where Indoor_Outdoor_Travelling like 'Travelling' and Call_Drop_Category like 'Satisfactory' group by Operator order by c desc limit 1").show()

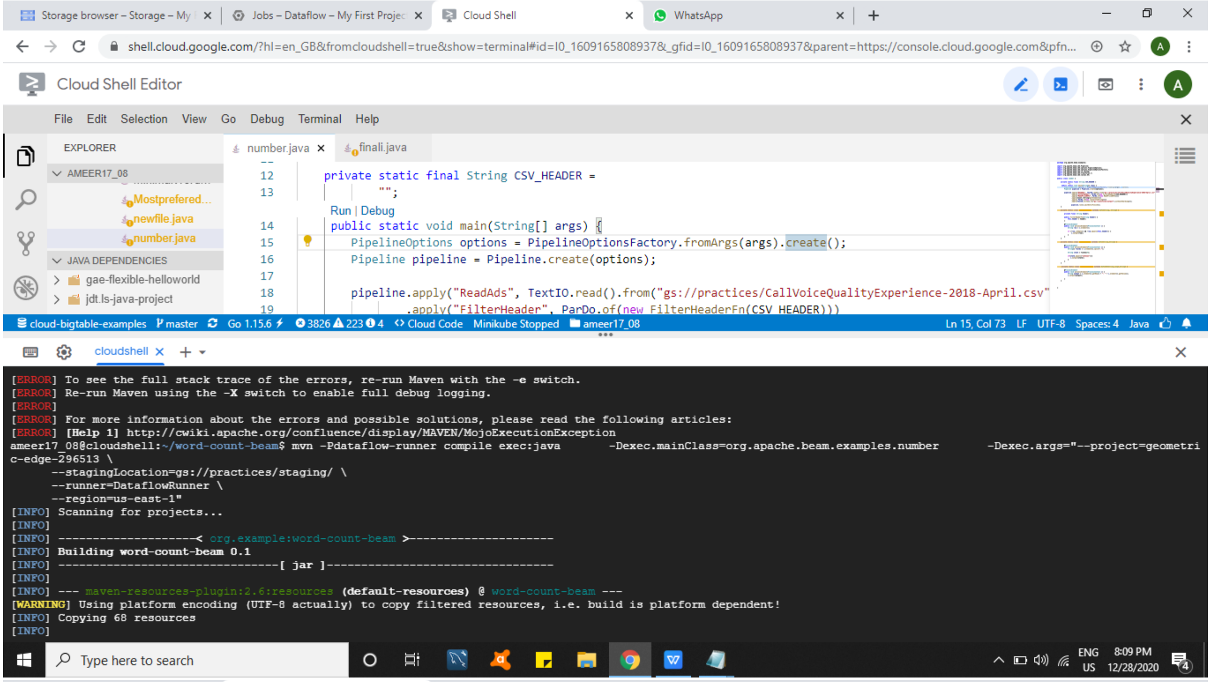


```
harshita1262@practices-m: ~ - Google Chrome
ssh.cloud.google.com/projects/geometric-edge-296513/zones/asia-east1-c/instances/practices-m?authuser=0&hl=en_US&projectNumber=398943591221&useAdminProxy=true
  ... 49 elided

scala> spark.sql("create table inputdata(Operator String,Indoor_Outdoor_Travelling String,NetworkType String,Rating
 String,CallDropCategory String,Latitude String,Longitude String,StateName String) row format delimited fields term
inated by ',' stored as textfile")
ivysettings.xml file not found in HIVE_HOME or HIVE_CONF_DIR,/etc/hive/conf.dist/ivysettings.xml will be used
res7: org.apache.spark.sql.DataFrame = []

scala> spark.sql("Load data local inpath '/home/harshita1262/exp.csv' into table input")
res8: org.apache.spark.sql.DataFrame = []

scala> spark.sql("SELECT Operator,count(*) as c FROM input where  Indoor_Outdoor_Travelling like 'Travelling' and C
all_Drop_Category like 'Satisfactory' group by Operator order by c desc  limit 1").show()
+--------+----+
|Operator|   c|
+--------+----+
|    RJio|1847|
+--------+----+


scala>

scala>

scala>

scala>

scala>

scala>

scala>

scala>

scala>

scala>

scala>

scala>

scala>
```

10.List the number of calls where the location of user could not be resolved. [Apache beam without Sql]

```
package org.apache.beam.examples;

import org.apache.beam.sdk.Pipeline;
import org.apache.beam.sdk.options.PipelineOptions;
import org.apache.beam.sdk.options.PipelineOptionsFactory;
import org.apache.beam.sdk.io.TextIO;
import org.apache.beam.sdk.transforms.*;
import org.apache.beam.sdk.values.KV;

public class number {

   private static final String CSV_HEADER =
         "Operator,Indoor_Outdoor_Travelling,Network Type,Rating,Call Drop
Category,Latitude,Longitude,State Name";
    public static void main(String[] args) {
      PipelineOptions options = PipelineOptionsFactory.fromArgs(args).create();
      Pipeline pipeline = Pipeline.create(options);

      pipeline.apply("ReadAds", TextIO.read().from("gs://practices/CallVoiceQualityExperience-
2018-April.csv"))
              .apply("FilterHeader", ParDo.of(new FilterHeaderFn(CSV_HEADER)))
              .apply("MakePriceKVFn", ParDo.of(new MakePriceKVFn()))
              .apply(Count.<String>perElement())
              .apply( ParDo.of(new converttoString()))
              .apply(TextIO.write().to("gs://practices/output").withoutSharding());

              pipeline.run().waitUntilFinish();
   }
```

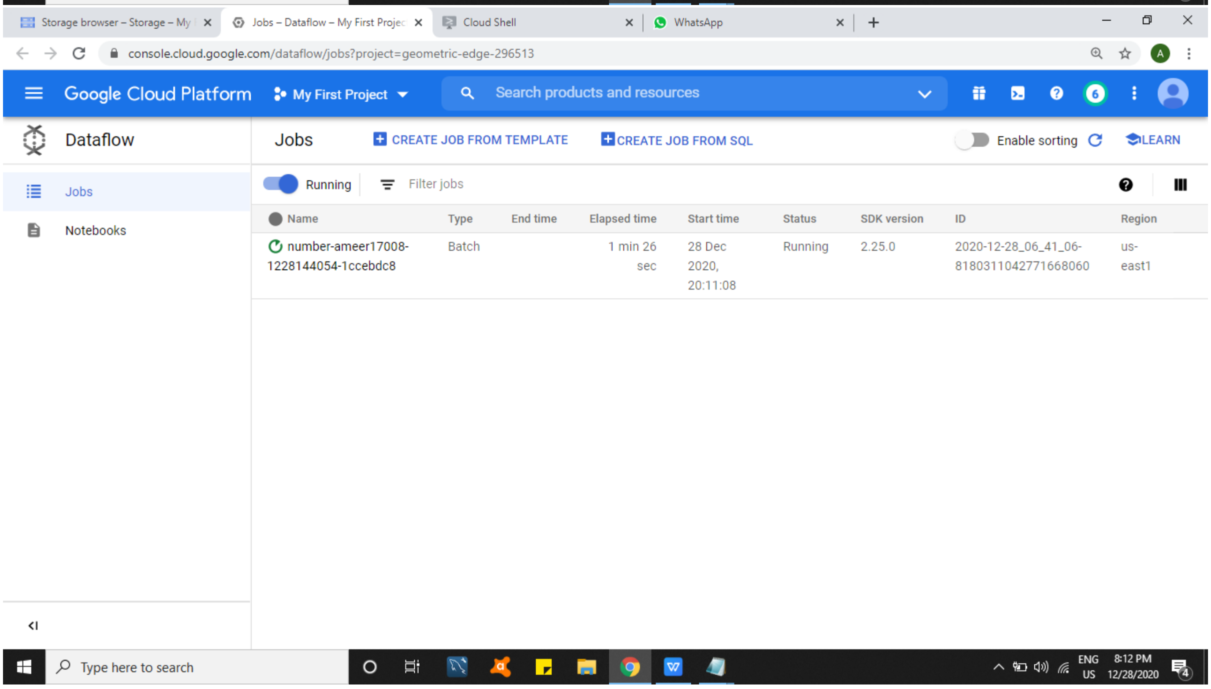
```
private static class FilterHeaderFn extends DoFn<String, String> {

    private final String header;

    public FilterHeaderFn(String header) {
        this.header = header;
    }

    @ProcessElement
    public void processElement(ProcessContext c) {
        String row = c.element();

        if (!row.isEmpty() && !row.equals(this.header)) {
            c.output(row);
        }
    }
}

private static class MakePriceKVFn extends DoFn<String,String> {

    @ProcessElement
    public void processElement(ProcessContext c) {
        String[] fields = c.element().split(",");

        String state = fields[7];

        if(state.equals("unknown")){
           c.output(state);
        }
    }
}

 private static class converttoString extends DoFn<KV<String,Long>,String> {

    @ProcessElement
    public void processElement(ProcessContext c) {
        String output = c.element().getKey() + ": " + c.element().getValue();
        c.output(output);

    }
}
```



```java
private static final String CSV_HEADER =
        "";
public static void main(String[] args) {
    PipelineOptions options = PipelineOptionsFactory.fromArgs(args).create();
    Pipeline pipeline = Pipeline.create(options);

    pipeline.apply("ReadAds", TextIO.read().from("gs://practices/CallVoiceQualityExperience-2018-April.csv"
            .apply("FilterHeader", ParDo.of(new FilterHeaderFn(CSV_HEADER)))
```

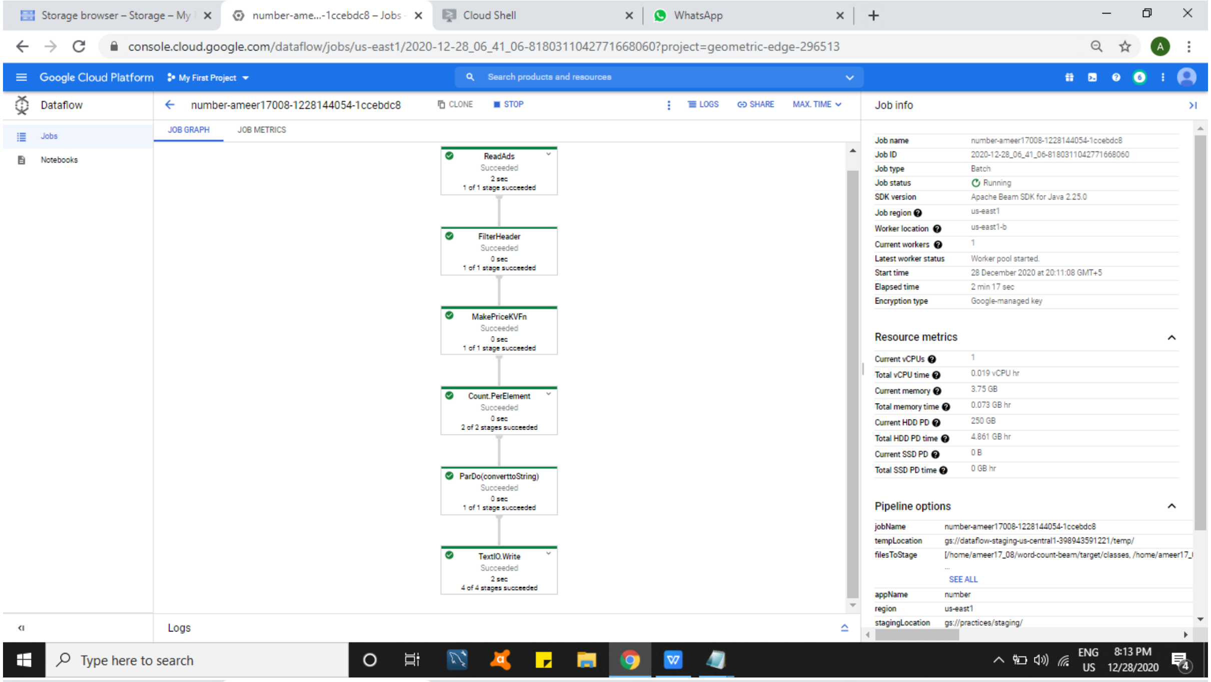
```
[ERROR] To see the full stack trace of the errors, re-run Maven with the -e switch.
[ERROR] Re-run Maven using the -X switch to enable full debug logging.
[ERROR]
[ERROR] For more information about the errors and possible solutions, please read the following articles:
[ERROR] [Help 1] http://cwiki.apache.org/confluence/display/MAVEN/MojoExecutionException
ameer17_08@cloudshell:~/word-count-beam$ mvn -Pdataflow-runner compile exec:java       -Dexec.mainClass=org.apache.beam.examples.number       -Dexec.args="--project=geometri
c-edge-296513 \
      --stagingLocation=gs://practices/staging/ \
      --runner=DataflowRunner \
      --region=us-east-1"
[INFO] Scanning for projects...
[INFO]
[INFO] -------------------< org.example:word-count-beam >--------------------
[INFO] Building word-count-beam 0.1
[INFO] --------------------------------[ jar ]---------------------------------
[INFO]
[INFO] --- maven-resources-plugin:2.6:resources (default-resources) @ word-count-beam ---
[WARNING] Using platform encoding (UTF-8 actually) to copy filtered resources, i.e. build is platform dependent!
[INFO] Copying 68 resources
[INFO]
```

Storage browser – Storage – My
Jobs – Dataflow – My First Projec
Cloud Shell
WhatsApp
shell.cloud.google.com/?hl=en_GB&fromcloudshell=true&show=terminal#id=I0_1609165808937&_gfid=I0_1609165808937&parent=https://console.cloud.google.com&pfn...
Cloud Shell Editor
File Edit Selection View Go Debug Terminal Help
EXPLORER
AMEER17_08
Mostprefered...
newfile.java
number.java
JAVA DEPENDENCIES
gae-flexible-helloworld
jdt.ls-java-project
number.java
finali.java
Run | Debug
cloud-bigtable-examples master Go 1.15.6 3826 223 4 Cloud Code Minikube Stopped ameer17_08
Ln 15, Col 73 LF UTF-8 Spaces: 4 Java
cloudshell
Type here to search
ENG US 8:09 PM 12/28/2020



Name	Type	End time	Elapsed time	Start time	Status	SDK version	ID	Region
number-ameer17008-1228144054-1ccebdc8	Batch		1 min 26 sec	28 Dec 2020, 20:11:08	Running	2.25.0	2020-12-28_06_41_06-8180311042771668060	us-east1

Storage browser – Storage – My
Jobs – Dataflow – My First Projec
Cloud Shell
WhatsApp
console.cloud.google.com/dataflow/jobs?project=geometric-edge-296513
Google Cloud Platform
My First Project
Search products and resources
Dataflow
Jobs
CREATE JOB FROM TEMPLATE
CREATE JOB FROM SQL
Enable sorting
LEARN
Jobs
Notebooks
Running
Filter jobs
Type here to search
ENG US 8:12 PM 12/28/2020



Storage browser – Storage – My
number-ame...-1ccebdc8 – Jobs
Cloud Shell
WhatsApp
console.cloud.google.com/dataflow/jobs/us-east1/2020-12-28_06_41_06-8180311042771668060?project=geometric-edge-296513
Google Cloud Platform
My First Project
Search products and resources
Dataflow
Jobs
Notebooks
number-ameer17008-1228144054-1ccebdc8
CLONE
STOP
LOGS
SHARE
MAX. TIME
JOB GRAPH
JOB METRICS
Logs
Type here to search
ENG
US
8:13 PM
12/28/2020

ReadAds
Succeeded
2 sec
1 of 1 stage succeeded

FilterHeader
Succeeded
0 sec
1 of 1 stage succeeded

MakePriceKVFn
Succeeded
0 sec
1 of 1 stage succeeded

Count.PerElement
Succeeded
0 sec
2 of 2 stages succeeded

ParDo(converttoString)
Succeeded
0 sec
1 of 1 stage succeeded

TextIO.Write
Succeeded
2 sec
4 of 4 stages succeeded

Job info

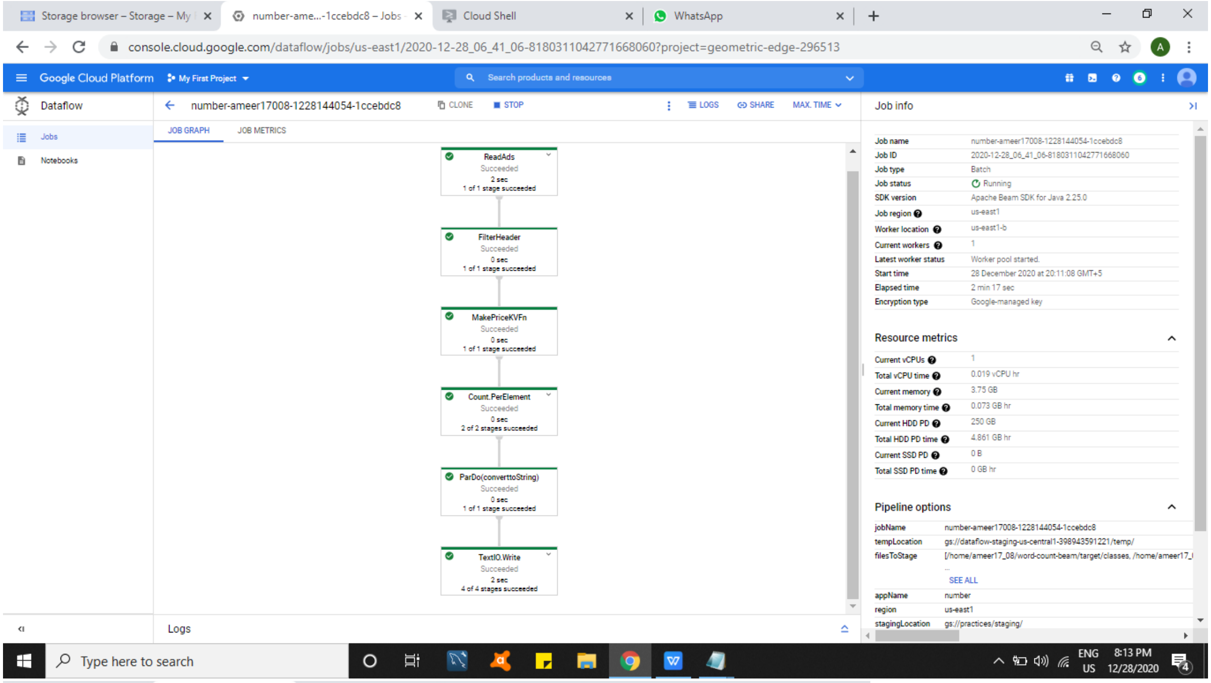
Job name	number-ameer17008-1228144054-1ccebdc8
Job ID	2020-12-28_06_41_06-8180311042771668060
Job type	Batch
Job status	Running
SDK version	Apache Beam SDK for Java 2.25.0
Job region	us-east1
Worker location	us-east1-b
Current workers	1
Latest worker status	Worker pool started.
Start time	28 December 2020 at 20:11:08 GMT+5
Elapsed time	2 min 17 sec
Encryption type	Google-managed key

Resource metrics

Current vCPUs	1
Total vCPU time	0.019 vCPU hr
Current memory	3.75 GB
Total memory time	0.073 GB hr
Current HDD PD	250 GB
Total HDD PD time	4.861 GB hr
Current SSD PD	0 B
Total SSD PD time	0 GB hr

Pipeline options

jobName	number-ameer17008-1228144054-1ccebdc8
tempLocation	gs://dataflow-staging-us-central1-398943591221/temp/
filesToStage	[/home/ameer17_08/word-count-beam/target/classes, /home/ameer17_ ... SEE ALL
appName	number
region	us-east1
stagingLocation	gs://practices/staging/



ReadAds
Succeeded
2 sec
1 of 1 stage succeeded

FilterHeader
Succeeded
0 sec
1 of 1 stage succeeded

MakePriceKVFn
Succeeded
0 sec
1 of 1 stage succeeded

Count.PerElement
Succeeded
0 sec
2 of 2 stages succeeded

ParDo(converttoString)
Succeeded
0 sec
1 of 1 stage succeeded

TextIO.Write
Succeeded
2 sec
4 of 4 stages succeeded

Job info

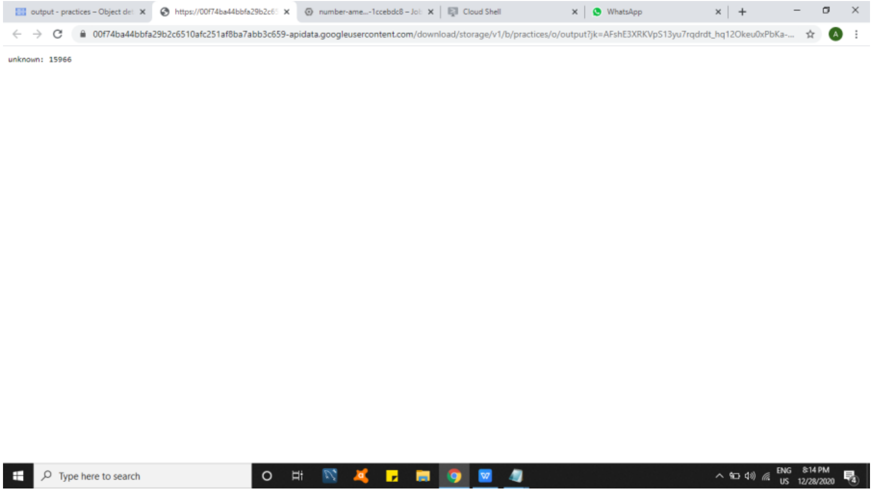
Job name number-ameer17008-1228144054-1ccebdc8
Job ID 2020-12-28_06_41_06-8180311042771668060
Job type Batch
Job status Running
SDK version Apache Beam SDK for Java 2.25.0
Job region us-east1
Worker location us-east1-b
Current workers 1
Latest worker status Worker pool started.
Start time 28 December 2020 at 20:11:08 GMT+5
Elapsed time 2 min 17 sec
Encryption type Google-managed key

Resource metrics

Current vCPUs 1
Total vCPU time 0.019 vCPU hr
Current memory 3.75 GB
Total memory time 0.073 GB hr
Current HDD PD 250 GB
Total HDD PD time 4.861 GB hr
Current SSD PD 0 B
Total SSD PD time 0 GB hr

Pipeline options

jobName number-ameer17008-1228144054-1ccebdc8
tempLocation gs://dataflow-staging-us-central1-398943591221/temp/
filesToStage [/home/ameer17_08/word-count-beam/target/classes, /home/ameer17_
SEE ALL
appName number
region us-east1
stagingLocation gs://practices/staging/



output - practices – Object det
https://00f74ba44bbfa29b2c65
number-ame...-1ccebdc8 – Job
Cloud Shell
WhatsApp
00f74ba44bbfa29b2c6510afc251af8ba7abb3c659-apidata.googleusercontent.com/download/storage/v1/b/practices/o/output?jk=AFshE3XRKVpS13yu7rqdrdt_hq12Okeu0xPbKa-...
Type here to search
ENG
US
8:14 PM
12/28/2020

```
unknown: 15966
```