

BAY AREA BIKE SHARE

The Bay Area Bike Share system allows users to rent bicycles for short journeys between stations throughout the city. Users can be annual members or short term (1 or 3 days).

There are 69 stations across 5 cities (San Jose, Redwood City, Mountain View, Palo Alto, San Francisco) in the Bike Share system.

Configured a multinode Hadoop cluster with NameNode, Secondary NameNode and two slaves to perform the map reduce jobs.

DataSet:

<http://www.bayareabikeshare.com/open-data>

FILE LIST

- 1) 201402_status_data.csv – approx. 17 million records of status data (bike and dock availability)**
- 2) 201402_station_data.csv – 69 records – station latitude, longitude, name, dockcount, installation date**
- 3) 201402_trip_data.csv – approx. 144,000 records of individual trips**

Files contain data from 8/29/13 (system launch) to 2/28/14.

1.Hive Analysis

1a. No of stations in every city: About 50 % of the stations are located in San Francisco.

```
Palo Alto      5
Redwood City   7
Mountain View  7
San Jose       15
San Francisco  35
Time taken: 49.145 seconds, Fetched: 5 row(s)
hive> █
```

1b. No of docks in every city: About 50 % of the docks are located in San Francisco.

```
Palo Alto      75
Redwood City   115
Mountain View  117
San Jose       249
San Francisco  665
Time taken: 49.689 seconds, Fetched: 5 row(s)
hive> █
```

1c. No of rides per city: About 90% of the total rides start in San Francisco.

```
Redwood City   793
Palo Alto      1706
Mountain View  2728
San Jose       8935
San Francisco  129853
```

2.Pig Analysis:

2a. Top 10 start stations: This displays the top ten start stations based

on the no of trips started at that particular start station.

```
(San Francisco Caltrain (Townsend at 4th),9838)
(Harry Bridges Plaza (Ferry Building),7343)
(Embarcadero at Sansome,6545)
(Market at Sansome,5922)
(Temporary Transbay Terminal (Howard at Beale),5113)
(Market at 4th,5030)
(2nd at Townsend,4987)
(San Francisco Caltrain 2 (330 Townsend),4976)
(Steuart at Market,4913)
(Townsend at 7th,4493)
grunt> █
```

2b. Top 10 destinations: This displays the top ten destinations based on the no of trips ending at that particular start station.

```
(San Francisco Caltrain (Townsend at 4th),11637)
(Embarcadero at Sansome,7590)
(Harry Bridges Plaza (Ferry Building),7475)
(Market at Sansome,6238)
(2nd at Townsend,5655)
(San Francisco Caltrain 2 (330 Townsend),5112)
(Market at 4th,5109)
(Steuart at Market,5080)
(Townsend at 7th,5073)
(2nd at South Park,4431)
grunt> █
```

2c. Top stations for trip starting at San Francisco Cal train station:

```
(Harry Bridges Plaza (Ferry Building),782)
(Embarcadero at Folsom,695)
(Temporary Transbay Terminal (Howard at Beale),644)
(Steuart at Market,609)
(Townsend at 7th,568)
(Yerba Buena Center of the Arts (3rd @ Howard),495)
(Market at Sansome,480)
(Howard at 2nd,427)
(Embarcadero at Bryant,394)
(Spear at Folsom,346)
grunt> █
```

2d. Busiest Day: This displays the most Busiest day in the 183 days based on the no trips.

```
(9/25/2013,1264)
grunt>
```

2e. Calmest Day: This displays the most calmest day in the 183 days based on the no of trips.

```
(2/9/2014,81)
grunt>
```

3. Map Reduce Analysis:

3a. This Analysis results in displaying the total number of customers and total number of subscribers among the users of the Bay Area Bike share.

```
ubuntu@ec2-52-10-114-38:~/hadoop/bin$ hadoop fs -cat /output/part-r-00000
Customer      30368
Subscriber    113647
ubuntu@ec2-52-10-114-38:~/hadoop/bin$
```

3b. This Analysis results in displaying the no of trips starting at every station.

2nd at Folsom	3776	
2nd at South Park	4458	
2nd at Townsend	4987	
5th at Howard	2635	
Adobe on Almaden	335	
Arena Green / SAP Center	339	
Beale at Market	3057	
Broadway St at Battery St	201	
Broadway at Main	45	
California Ave Caltrain Station	297	
Castro Street and El Camino Real	132	
Civic Center BART (7th at Market)	3074	
Clay at Battery	2419	
Commercial at Montgomery	2834	
Cowper at University	355	
Davis at Jackson	2300	
Embarcadero at Bryant	3497	
Embarcadero at Folsom	2931	
Embarcadero at Sansome	6545	
Embarcadero at Vallejo	2785	
Evelyn Park and Ride	218	
Franklin at Maple	99	
Golden Gate at Polk	1757	
Grant Avenue at Columbus Avenue	4004	
Harry Bridges Plaza (Ferry Building)	7343	
Howard at 2nd	3263	
Japantown	496	
MLK Library	528	
Market at 10th	3511	
Market at 4th	5030	
Market at Sansome	5922	
Mechanics Plaza (Market at Battery)	2868	
Mezes Park	3	
Mountain View Caltrain Station	1154	
Mountain View City Hall	630	
Palo Alto Caltrain Station	431	
Park at Olive	189	
Paseo de San Antonio	745	

```
Paseo de San Antonio      745
Post at Kearney 2503
Powell Street BART        3836
Powell at Post (Union Square) 2824
Redwood City Caltrain Station 378
Redwood City Medical Center 123
Redwood City Public Library 44
Rengstorff Avenue / California Street 248
SJSU - San Salvador at 9th 489
SJSU 4th at San Carlos 389
San Antonio Caltrain Station 238
San Antonio Shopping Center 108
San Francisco Caltrain (Townsend at 4th) 9838
San Francisco Caltrain 2 (330 Townsend) 4976
San Francisco City Hall 1257
San Jose City Hall 740
San Jose Civic Center 346
San Jose Diridon Caltrain Station 1972
San Jose Government Center 23
San Mateo County Center 101
San Pedro Square 715
San Salvador at 1st 597
Santa Clara at Almaden 855
South Van Ness at Market 3521
Spear at Folsom 3423
St James Park 366
Steuart at Market 4913
Temporary Transbay Terminal (Howard at Beale) 5113
Townsend at 7th 4493
University and Emerson 434
Washington at Kearney 1472
Yerba Buena Center of the Arts (3rd @ Howard) 2487
ubuntu@ec2-52-10-114-38:~/hadoop/bin$
```

3c. This analysis displays the no trips between every combination of start and end station using a Composite Key. This will further help to get the most Traveled routes.

```
ubuntu@ec2-52-10-114-38:~/hadoop/bin$ hadoop fs -cat /output12/part-r-00000
2nd at Folsom:2nd at Folsom 74
2nd at Folsom:2nd at South Park 238
2nd at Folsom:2nd at Townsend 219
2nd at Folsom:5th at Howard 93
2nd at Folsom:Beale at Market 48
2nd at Folsom:Broadway St at Battery St 9
2nd at Folsom:Civic Center BART (7th at Market) 70
2nd at Folsom:Clay at Battery 107
2nd at Folsom:Commercial at Montgomery 47
2nd at Folsom:Davis at Jackson 37
2nd at Folsom:Embarcadero at Bryant 60
2nd at Folsom:Embarcadero at Folsom 35
2nd at Folsom:Embarcadero at Sansome 64
2nd at Folsom:Embarcadero at Vallejo 20
2nd at Folsom:Golden Gate at Polk 6
2nd at Folsom:Grant Avenue at Columbus Avenue 90
2nd at Folsom:Harry Bridges Plaza (Ferry Building) 179
2nd at Folsom:Howard at 2nd 77
2nd at Folsom:Market at 10th 24
2nd at Folsom:Market at 4th 149
2nd at Folsom:Market at Sansome 537
2nd at Folsom:Mechanics Plaza (Market at Battery) 38
2nd at Folsom:Post at Kearney 94
2nd at Folsom:Powell Street BART 118
2nd at Folsom:Powell at Post (Union Square) 50
2nd at Folsom:San Francisco Caltrain (Townsend at 4th) 430
2nd at Folsom:San Francisco Caltrain 2 (330 Townsend) 181
2nd at Folsom:San Francisco City Hall 14
2nd at Folsom:South Van Ness at Market 22
2nd at Folsom:Spear at Folsom 241
2nd at Folsom:Steuart at Market 102
2nd at Folsom:Temporary Transbay Terminal (Howard at Beale) 110
2nd at Folsom:Townsend at 7th 97
2nd at Folsom:Washington at Kearney 37
2nd at Folsom:Yerba Buena Center of the Arts (3rd @ Howard) 59
2nd at South Park:2nd at Folsom 130
```

Grant Avenue at Columbus Avenue:San Francisco City Hall	16	
Grant Avenue at Columbus Avenue:South Van Ness at Market	72	
Grant Avenue at Columbus Avenue:Spear at Folsom	123	
Grant Avenue at Columbus Avenue:Steuart at Market	186	
Grant Avenue at Columbus Avenue:Temporary Transbay Terminal (Howard at Beale)	121	
Grant Avenue at Columbus Avenue:Townsend at 7th	49	
Grant Avenue at Columbus Avenue:Washington at Kearney	43	
Grant Avenue at Columbus Avenue:Yerba Buena Center of the Arts (3rd @ Howard)	56	
Harry Bridges Plaza (Ferry Building):2nd at Folsom	98	
Harry Bridges Plaza (Ferry Building):2nd at South Park	167	
Harry Bridges Plaza (Ferry Building):2nd at Townsend	710	
Harry Bridges Plaza (Ferry Building):5th at Howard	92	
Harry Bridges Plaza (Ferry Building):Beale at Market	55	
Harry Bridges Plaza (Ferry Building):Broadway St at Battery St	8	
Harry Bridges Plaza (Ferry Building):Civic Center BART (7th at Market)	52	
Harry Bridges Plaza (Ferry Building):Clay at Battery	86	
Harry Bridges Plaza (Ferry Building):Commercial at Montgomery	218	
Harry Bridges Plaza (Ferry Building):Davis at Jackson	122	
Harry Bridges Plaza (Ferry Building):Embarcadero at Bryant	293	
Harry Bridges Plaza (Ferry Building):Embarcadero at Folsom	54	
Harry Bridges Plaza (Ferry Building):Embarcadero at Sansome	1330	
Harry Bridges Plaza (Ferry Building):Embarcadero at Vallejo	332	
Harry Bridges Plaza (Ferry Building):Golden Gate at Polk	21	
Harry Bridges Plaza (Ferry Building):Grant Avenue at Columbus Avenue	153	
Harry Bridges Plaza (Ferry Building):Harry Bridges Plaza (Ferry Building)	550	
Harry Bridges Plaza (Ferry Building):Howard at 2nd	92	
Harry Bridges Plaza (Ferry Building):Market at 10th	46	
Harry Bridges Plaza (Ferry Building):Market at 4th	261	
Harry Bridges Plaza (Ferry Building):Market at Sansome	294	
Harry Bridges Plaza (Ferry Building):Mechanics Plaza (Market at Battery)	161	
Harry Bridges Plaza (Ferry Building):Post at Kearney	148	
Harry Bridges Plaza (Ferry Building):Powell Street BART	142	
Harry Bridges Plaza (Ferry Building):Powell at Post (Union Square)	152	
Harry Bridges Plaza (Ferry Building):San Francisco Caltrain (Townsend at 4th)	598	
Harry Bridges Plaza (Ferry Building):San Francisco Caltrain 2 (330 Townsend)	307	
Harry Bridges Plaza (Ferry Building):San Francisco City Hall	54	
Harry Bridges Plaza (Ferry Building):South Van Ness at Market	56	
Harry Bridges Plaza (Ferry Building):Spear at Folsom	148	
Harry Bridges Plaza (Ferry Building):Steuart at Market	22	
Harry Bridges Plaza (Ferry Building):Temporary Transbay Terminal (Howard at Beale)	136	
Harry Bridges Plaza (Ferry Building):Townsend at 7th	230	

San Francisco Caltrain 2 (330 Townsend):Post at Kearney	61
San Francisco Caltrain 2 (330 Townsend):Powell Street BART	521
San Francisco Caltrain 2 (330 Townsend):Powell at Post (Union Square)	46
San Francisco Caltrain 2 (330 Townsend):San Francisco Caltrain (Townsend at 4th)	37
San Francisco Caltrain 2 (330 Townsend):San Francisco Caltrain 2 (330 Townsend)	79
San Francisco Caltrain 2 (330 Townsend):San Francisco City Hall	24
San Francisco Caltrain 2 (330 Townsend):South Van Ness at Market	134
San Francisco Caltrain 2 (330 Townsend):Spear at Folsom	81
San Francisco Caltrain 2 (330 Townsend):Steuart at Market	168
San Francisco Caltrain 2 (330 Townsend):Temporary Transbay Terminal (Howard at Beale)	112
San Francisco Caltrain 2 (330 Townsend):Townsend at 7th	1116
San Francisco Caltrain 2 (330 Townsend):University and Emerson	2
San Francisco Caltrain 2 (330 Townsend):Washington at Kearney	46
San Francisco Caltrain 2 (330 Townsend):Verba Buena Center of the Arts (3rd @ Howard)	109
San Francisco City Hall:2nd at Folsom	7
San Francisco City Hall:2nd at South Park	26
San Francisco City Hall:2nd at Townsend	16
San Francisco City Hall:5th at Howard	18
San Francisco City Hall:Beale at Market	27
San Francisco City Hall:Civic Center BART (7th at Market)	31
San Francisco City Hall:Clay at Battery	5
San Francisco City Hall:Commercial at Montgomery	17
San Francisco City Hall:Davis at Jackson	8
San Francisco City Hall:Embarcadero at Bryant	4
San Francisco City Hall:Embarcadero at Folsom	1
San Francisco City Hall:Embarcadero at Sansome	30
San Francisco City Hall:Embarcadero at Vallejo	10
San Francisco City Hall:Golden Gate at Polk	21
San Francisco City Hall:Grant Avenue at Columbus Avenue	24
San Francisco City Hall:Harry Bridges Plaza (Ferry Building)	46
San Francisco City Hall:Howard at 2nd	6
San Francisco City Hall:Market at 10th	19
San Francisco City Hall:Market at 4th	119
San Francisco City Hall:Market at Sansome	56
San Francisco City Hall:Mechanics Plaza (Market at Battery)	16
San Francisco City Hall:Post at Kearney	74
San Francisco City Hall:Powell Street BART	99
San Francisco City Hall:Powell at Post (Union Square)	37
San Francisco City Hall:San Francisco Caltrain (Townsend at 4th)	131
San Francisco City Hall:San Francisco Caltrain 2 (330 Townsend)	49
San Francisco City Hall:San Francisco City Hall	96

Source Codes:

1a. #No of stations#

```
create EXTERNAL table stations(stationid int,name string,lat int,long  
int,dockcount int,landmark string,installation string)  
row format delimited  
fields terminated by ','  
location '/hive/stations';
```

```
load data inpath '/data/stations/201402_station_data.csv' into table  
stations;
```

```
select landmark, COUNT(*) as count from stations  
group by landmark  
order by count;
```

1b. #No of Docks#

```
create EXTERNAL table stations(stationid int,name string,lat int,long  
int,dockcount int,landmark string,installation string)  
row format delimited  
fields terminated by ','  
location '/hive/stations';
```

```
load data local inpath '/data/stations/201402_station_data.csv' into table  
stations;
```

```
select landmark, SUM(dockcount) as dockcount from stations  
group by landmark  
order by dockcount;
```

1c. #No of Trips#

```
create EXTERNAL table stations(stationid int,name string,lat int,long  
int,dockcount int,landmark string,installation string)  
row format delimited  
fields terminated by ','  
location '/hive/stations';
```

```
load data local inpath '/data/stations/201402_station_data.csv' into table  
stations;
```

```
create external table tripsnew (tripId int,duration int,startdate  
string,startstation string,startterminal int,  
enddate string,endstation string,endterminal int,bikeno  
int,subscriptiontype string,zipcode int)  
row format delimited  
fields terminated by ','  
location '/hive/tripsnew';
```

```
load data inpath '/data/trips/201402_trip_data.csv'into table tripsnew;
```

```
select stations.landmark,COUNT(tripsnew.tripId) as count  
from stations JOIN tripsnew  
on (stations.stationid=tripsnew.startterminal)  
group by stations.landmark  
order by count;
```

2a. # Top 10 start stations#

```
trip = LOAD
```

```
 '/data/year1/201402_babs_open_data/201402_trip_data.csv' using  
PigStorage(',') AS (tripId,duration,startdate,startstation,startterminal,  
enddate,endstation,endterminal,bikeno,subscriptiontype,zipcode);
```

```
tripone = GROUP trip BY startstation;
```

```
triptwo = FOREACH tripone GENERATE group AS startstation,  
COUNT(trip.startstation) AS count;
```

```
tripthree= ORDER triptwo BY count DESC;
```

```
tripfour= LIMIT tripthree 10;
```

2b. #Top 10 destinations#

```
trip = LOAD  
'/data/year1/201402_babs_open_data/201402_trip_data.csv' using  
PigStorage(',') AS (tripId,duration,startdate,startstation,startterminal,  
enddate,endstation,endterminal,bikeno,subscriptiontype,zipcode);
```

```
tripone = GROUP trip BY endstation;
```

```
triptwo = FOREACH tripone GENERATE group AS endstation,  
COUNT(trip.endstation) AS count;
```

```
tripthree= ORDER triptwo BY count DESC;
```

```
tripfour= LIMIT tripthree 10;
```

2c. #Top stations for trip starting at San Francisco Caltrain station#

```
stations = LOAD  
'/data/year1/201402_babs_open_data/201402_station_data.csv' using  
PigStorage(',') AS  
(stationid,name,lat,long,dockcount,landmark,installation);
```

```
trip = LOAD  
'/data/year1/201402_babs_open_data/201402_trip_data.csv' using
```

**PigStorage(',') AS (tripId,duration,startdate,startstation,startterminal,
enddate,endstation,endterminal,bikeno,subscriptiontype,zipcode);**

joined= JOIN stations by stationid,trip by endterminal;

report = FILTER joined BY (startterminal==70);

triptwo= GROUP report BY stations::name;

**tripthree= FOREACH triptwo GENERATE group as
stationname,COUNT(report.stationid) as count;**

tripfour = ORDER tripthree BY count DESC;

tripfive =LIMIT tripfour 10;

2d. #Busiest day#

**trip = LOAD '/data/201402_babs_open_data/201402_trip_data.csv' using
PigStorage(',') AS
(tripId:int,duration:int,startdate:chararray,startstation:chararray,startt
erminal:chararray,
enddate:chararray,endstation:chararray,endterminal:chararray,bikeno:
int,subscriptiontype:chararray,zipcode:int);**

**split_date =FOREACH trip {GENERATE
FLATTEN(STRSPLIT(startdate,'\\s',2)) AS (date,time)};**

date = FOREACH split_date GENERATE date AS date;

tripthree= GROUP date BY date;

**tripfour= FOREACH tripthree GENERATE group AS date,
COUNT(date) AS count;**

tripfive= ORDER tripfour BY count DESC;

tripsix= LIMIT tripfive 1;

2e. #CalmeSt day#

**trip = LOAD '/data/201402_babs_open_data/201402_trip_data.csv' using
PigStorage(',') AS
(tripId:int,duration:int,startdate:chararray,startstation:chararray,startt
erminal:chararray,
enddate:chararray,endstation:chararray,endterminal:chararray,bikeno:
int,subscriptiontype:chararray,zipcode:int);**

**split_date =FOREACH trip {GENERATE
FLATTEN(STRSPLIT(startdate,'\\s',2)) AS (date,time);};**

date = FOREACH split_date GENERATE date AS date;

tripthree= GROUP date BY date;

**tripfour= FOREACH tripthree GENERATE group AS date,
COUNT(date) AS count;**

tripseven= ORDER tripfour BY count;

tripeight= LIMIT tripseven 1;

Hive Partition Creation:

Hive provides the facility to store data in a table in partitions. The following code was implemented to divide the trips table into partitions based on the year. This helps fast retrieval of data.

```
CREATE EXTERNAL TABLE IF NOT EXISTS trips (tripId  
int,duration int,startdate string,startstation string,startterminal int,  
enddate string,endstation string,endterminal int,bikeno  
int,subscriptiontype string,zipcode int )  
PARTITIONED BY (year string)  
ROW FORMAT DELIMITED  
FIELDS TERMINATED BY ','  
LOCATION '/hive/trips';
```

```
LOAD DATA LOCAL INPATH '/data/trips/201402_trip_data.csv'  
INTO TABLE trips PARTITION (year='year1');  
LOAD DATA LOCAL INPATH '/data/trips/201408_trip_data.csv'  
INTO TABLE trips PARTITION (year='year2');
```

```
select * from trips where year ='year2';
```

#MapReduceThree.java#

```
/*
 * To change this license header, choose License Headers in Project
Properties.
 * To change this template file, choose Tools | Templates
 * and open the template in the editor.
 */
package mapreducethree;

import java.io.IOException;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.conf.Configured;
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.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
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.apache.hadoop.util.Tool;
import org.apache.hadoop.util.ToolRunner;

/**
 *
 * @author priyanka
 */
public class MapReduceThree extends Configured implements Tool{
```



```

/**
 * @param args the command line arguments
 */
public static class MapReduceThreeMapper extends
Mapper<LongWritable, Text, Text, IntWritable>{
    // private IntWritable trips = new IntWritable(1);
    // private Text startstation = new Text();

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

        String[] words = value.toString().split(",");
        if (words.length > 1)
        {
            IntWritable trips = new IntWritable(1);
            Text startstation = new Text();
            String sstation = words[3];
            startstation.set(sstation);
            context.write(startstation, trips);
        }

    }
}

public static class MapReduceThreeReducer extends Reducer <Text,
IntWritable, Text, IntWritable>{
    private IntWritable result=new IntWritable();

    @Override
    public void reduce(Text key, Iterable<IntWritable> values, Context
context)
        throws IOException, InterruptedException
    {
        int total = 0;

```

```

        for (IntWritable value : values)
        {
            total += value.get();
        }
        result.set(total);
        context.write(key, result);
    }

}

@Override
public int run(String[] args) throws Exception {

    Configuration configuration = getConf();
    Job job = Job.getInstance(configuration, "MapReduceThree");
    job.setJarByClass(MapReduceThree.class);
    FileInputFormat.addInputPath(job, new Path(args[0]));
    FileOutputFormat.setOutputPath(job, new Path(args[1]));
    job.setMapperClass(MapReduceThreeMapper.class);
    job.setCombinerClass(MapReduceThreeReducer.class);
    job.setReducerClass(MapReduceThreeReducer.class);
    job.setOutputKeyClass(Text.class);
    job.setOutputValueClass(IntWritable.class);
    System.exit(job.waitForCompletion(true) ? 0 : 1);
    return 0;
}

public static void main(String[] args) throws Exception{
    // TODO code application logic here
    int res = ToolRunner.run(new Configuration(), new
MapReduceThree(), args);
    System.exit(res);
}

}

```

#MapReduceFour.java#

```
/*
 * To change this license header, choose License Headers in Project
Properties.
 * To change this template file, choose Tools | Templates
 * and open the template in the editor.
 */
package mapreducefour;

import java.io.IOException;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.conf.Configured;
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.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Mapper.Context;
import org.apache.hadoop.mapreduce.Reducer;
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.apache.hadoop.util.Tool;
import org.apache.hadoop.util.ToolRunner;

/**
 *
 * @author priyanka
 */
public class MapReduceFour extends Configured implements Tool {
```

```
/**  
 * @param args the command line arguments  
 */
```

```
public static class MapReduceMapper extends  
Mapper<LongWritable, Text, Text, IntWritable>{  
    // private Text SubscriptionType = new Text();  
    // private IntWritable uno = new IntWritable(1);  
    // private Text SubscriptionType = new Text();
```

```
@Override
```

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

```
    // String line = value.toString();  
    // CSVReader R = new CSVReader(new StringReader(line));
```

```
    // String[] ParsedLine = R.readNext();  
    // R.close();
```

```
    // SubscriptionType.set(ParsedLine[10]);  
    // context.write(SubscriptionType, uno);
```

```
    //output.collect(SubscriptionType,);
```

```
String[] words = value.toString().split(",");  
    if (words.length > 1)  
    {  
        IntWritable count = new IntWritable(1);  
        Text SubscriptionType = new Text();  
        String SType = words[9];  
        SubscriptionType.set(SType);  
        context.write(SubscriptionType, count);  
    }
```

```

}
}
    public static class MapReduceReducer extends Reducer<Text,
IntWritable, Text, IntWritable> {
    private IntWritable result=new IntWritable();

    @Override
    public void reduce(Text key, Iterable<IntWritable> values, Context
context)
        throws IOException, InterruptedException
    {
        int total = 0;
        for (IntWritable value : values)
        {
            total += value.get();
        }
        result.set(total);
        context.write(key, result);
    }

}

```

```

    @Override
    public int run(String[] args) throws Exception
    {

```

```

        Configuration configuration = getConf();
        Job job = Job.getInstance(configuration, "MapReduceFour");
        job.setJarByClass(MapReduceFour.class);
        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));
        job.setMapperClass(MapReduceMapper.class);

```

```
    job.setCombinerClass(MapReduceReducer.class);  
    job.setReducerClass(MapReduceReducer.class);  
    job.setOutputKeyClass(Text.class);  
    job.setOutputValueClass(IntWritable.class);  
    System.exit(job.waitForCompletion(true) ? 0 : 1);  
    return 0;  
}
```

```
public static void main(String[] args) throws Exception {  
    // TODO code application logic here  
    int res = ToolRunner.run(new Configuration(), new  
MapReduceFour(), args);  
    System.exit(res);  
}  
  
}
```

#MapReduceFive.java#

```
/*
 * To change this license header, choose License Headers in Project
 * Properties.
 * To change this template file, choose Tools | Templates
 * and open the template in the editor.
 */
package mapreducefive;

import java.io.IOException;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.conf.Configured;
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.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.util.Tool;
import org.apache.hadoop.util.ToolRunner;

/**
 *
 * @author priyanka
 */
public class MapReduceFive extends Configured implements Tool {

    /**
     * @param args the command line arguments
     */
}
```

```

public static class MapReduceFiveMapper extends
Mapper<LongWritable, Text, CompositeGroupKey, IntWritable>{
    private IntWritable tripid = new IntWritable(1);
    // private Text startstation = new Text();
    private CompositeGroupKey comp= new CompositeGroupKey();
@Override
public void map(LongWritable key, Text value, Context context) throws
IOException,InterruptedException {

```

```

    String[] words = value.toString().split(",");
    if (words.length > 1)
    {
        // IntWritable tripid = new IntWritable(1);

        // tripid.set(Integer.parseInt(words[0]));
        comp.setStart(words[3]);
        comp.setEnd(words[6]);
        context.write(comp, tripid);
    }

```

```

}
}

```

```

    public static class MapReduceFiveReducer extends Reducer
<CompositeGroupKey, IntWritable, CompositeGroupKey,
IntWritable>{
    private IntWritable result=new IntWritable();
@Override
public void reduce(CompositeGroupKey key, Iterable<IntWritable>
values, Context context)
    throws IOException, InterruptedException
    {
        int total = 0;

```



```

        for (IntWritable value : values)
        {
            total += value.get();
        }
        result.set(total);
        context.write(key, result);
    }

}

@Override
public int run(String[] args) throws Exception {
    Configuration conf = new Configuration();
    Job job = Job.getInstance(conf, "MapReduceFive");
    job.setJarByClass(MapReduceFive.class);
    FileInputFormat.addInputPath(job, new Path(args[0]));
    FileOutputFormat.setOutputPath(job, new Path(args[1]));
    job.setMapperClass(MapReduceFiveMapper.class);
    job.setReducerClass(MapReduceFiveReducer.class);
    job.setMapOutputKeyClass(CompositeGroupKey.class);
    job.setMapOutputValueClass(IntWritable.class);
    job.setOutputKeyClass(CompositeGroupKey.class);
    job.setOutputValueClass(IntWritable.class);

    System.exit(job.waitForCompletion(true) ? 0 : 1);

    return 0;
}

public static void main(String[] args) throws Exception{
    // TODO code application logic here
    int res = ToolRunner.run(new Configuration(), new
    MapReduceFive(), args);
    System.exit(res);
}

```

```
}
```

```
/*
```

```
 * To change this license header, choose License Headers in Project Properties.
```

```
 * To change this template file, choose Tools | Templates
```

```
 * and open the template in the editor.
```

```
*/
```

```
package mapreducefive;
```

```
import java.io.DataInput;
```

```
import java.io.DataOutput;
```

```
import java.io.IOException;
```

```
import org.apache.hadoop.io.WritableComparable;
```

```
/**
```

```
 *
```

```
 * @author priyanka
```

```
*/
```

```
public class CompositeGroupKey implements  
WritableComparable<CompositeGroupKey> {
```

```
    public CompositeGroupKey(){
```

```
    }
```

```
    public String getStart() {
```

```
        return start;
```

```
    }
```

```
    public void setStart(String start) {
```

```
    this.start = start;
}
```

```
public String getEnd() {
    return end;
}
```

```
public void setEnd(String end) {
    this.end = end;
}
```

```
public CompositeGroupKey(String start,String end){
    this.start=start;
    this.end=end;
}
```

```
String start;
String end;
```

```
@Override
```

```
public int compareTo(CompositeGroupKey comp) {
    if (comp == null)
        return 0;
    int intstart = start.compareTo(comp.start);
    return intstart == 0 ? end.compareTo(comp.end) : intstart;
}
```

```
@Override
```

```
public String toString() {
    return start.toString() + ":" + end.toString();
}
```

```
@Override
```

```
public void write(DataOutput out) throws IOException {
    out.writeUTF(start);
    out.writeUTF(end);
}
```

```
@Override
```

```
public void readFields(DataInput in) throws IOException {  
    start = in.readUTF();  
    end = in.readUTF();  
}  
}
```