

TRAVEL WEBSITE COMPANY - RECOMMENDATIONS

Introduction:

Typically, travel website companies provide information and/or services related to travel for users. The information may include different destinations, hotels in a selected destination, flights, reviews, fare, etc. The websites act as a platform with various features to view the information and select destinations and hotels accordingly. The companies are usually a medium between the customer and the service providers (e.g., hotels, airlines). The companies may generate revenue from advertising, marketing agencies or from collaboration with the service providers based on their business model. As such, the companies would want to increase the number of users that frequent the website.

Users would prefer to use the website that provides unique information. In order to develop such a relation with its users, the company would need to improve user experience, provide accurate and relevant data to users. The information obtained from such usage is valuable for other service providers such as airliners, hotels etc., as well as for marketing and advertising purposes. Furthermore, this information would also serve as a historic data to further improve experiences. To enable this, collecting data and timely updating becomes an important characteristic for the company. Similarly, data analytics is becoming increasingly popular to predict user preferences and provide user specific data. Searching through large sets of data is a tedious task and such recommendations would save time for users.

Currently, the competitive advantage lies with the company that provides the best user experience which in turn is dependent on how accurately the user preferences are predicted. The idea is to use the data and predict which hotel the user is going to book and recommend it to the user.

For the final project, a sample data set from one of the travel website “expedia.com” is chosen. The expedia website provides a user interface that allows users to search for different destinations, hotels and corresponding ratings, reviews fare etc. The website also allows the users

to provide various inputs including destination, check in date, check out date, number of guests. Based on the received inputs, search results may be displayed to the users. The task lies in increasing the number of active users by providing recommendations and improving user experience by sorting the search results based on relevancy to the user.

Model 1: Recommendation engine:

Data driven recommendation engine that uses Map Reduce algorithm is developed. The engine uses historic data from the site as an input to provide user specific recommendations. The historic data extracted from the site has various fields including geographic information of the user, user country, user region, user city, check in date, check out date, number of guests, destination distance, destination country, hotel cluster id, package or not, number of bookings, no of clicks, marketing channel for various transactions of user with the site.

Map Reduce algorithm is used to analyze the input historic data to predict a destination, a hotel or a package for a user. Accordingly, the algorithm is configured to output destination recommendations, hotel recommendations, package recommendations to the user. Different use cases have been programmed with various input fields which will be described in detail below.

Input data:

| Column name | Description | Data type |
|---------------------------|---|-----------|
| date_time | Timestamp | string |
| site_name | ID of the Expedia point of sale (i.e. Expedia.com, Expedia.co.uk, Expedia.co.jp, ...) | int |
| posa_continent | ID of continent associated with site_name | int |
| user_location_country | The ID of the country the customer is located | int |
| user_location_region | The ID of the region the customer is located | int |
| user_location_city | The ID of the city the customer is located | int |
| orig_destination_distance | Physical distance between a hotel and a customer at the time of search. A null means the distance could not be calculated | double |
| user_id | ID of user | int |
| is_mobile | 1 when a user connected from a mobile device, 0 otherwise | tinyint |
| is_package | 1 if the click/booking was generated as a part of a package (i.e. combined with a flight), 0 otherwise | int |
| channel | ID of a marketing channel | int |
| srch_ci | Checkin date | string |
| srch_co | Checkout date | string |
| srch_adults_cnt | The number of adults specified in the hotel room | int |
| srch_children_cnt | The number of (extra occupancy) children specified in the hotel room | int |
| srch_rm_cnt | The number of hotel rooms specified in the search | int |
| srch_destination_id | ID of the destination where the hotel search was performed | int |
| srch_destination_type_id | Type of destination | int |
| hotel_continent | Hotel continent | int |
| hotel_country | Hotel country | int |
| hotel_market | Hotel market | int |
| is_booking | 1 if a booking, 0 if a click | tinyint |
| cnt | Numer of similar events in the context of the same user session | bigint |
| hotel_cluster | ID of a hotel cluster | int |

Use case 1: Current User Travel Recommendations

In this case, users with similar behavior patterns are determined. The similarity between the users is identified based on the chosen destination ID and the hotel cluster. Accordingly, packages and/or hotels are recommended to users based on these user similarities.

Use Case 1A: Travel Package Recommendation

When a user visits the site, based on the booking history of the user, other users with similar booking history are determined. Further, it is determined if these users have extra bookings compared to the bookings of the user that visited the site. These destinations and corresponding hotels may be recommended as a package to the user. The package may also include flights to the destination with additional data.

The use case is explained in more detail in the below example:

For example, user U1 has bookings in destination D1, hotel H1, and in destination D2 and in hotel D2. The algorithm is directed at finding other users (say U2, U3) who also have bookings in D1, H1, and D2, H2. These users U2, U3 may be considered to have similar booking pattern as user U1. Further, the program determines recommendations that can be made to user U1 based on the booking history of the users U2 and U3. If the users U2 and U3 have also visited other destinations apart from D1 and D2, these destinations as a package may be suggested to the user U1. Furthermore, package may include flight options to these destinations, length of stay, hotels etc.

U1: D1 - H1 : 4.0 ; D2 - H2 : 4.0

U2: D1 - H1 : 4.0 ; D2 - H2 : 4.0 ; D3 – H3 : 5.0

U3: D1 – H1 : 4.0 ; D2 - H2 : 4.0 ; D4 – H4 : 4.0

As can be seen, the users U2 have bookings in destinations D3 and U3 have bookings in destination D4 apart from D1 and D2. In such a cases, recommendation of packages including (D3, H3) and (D4, H4) are provided.

Moreover, for same destination recommendations, hotels may be ordered based on popularity of the hotels. The hotels are scored based on number of bookings and number of clicks. For instance in the above example, the user U3 has bookings also in destination D3 and hotel H4, the recommendations may be given based on the hotel score for the hotels H3 and H4. As discussed above, the recommendations of the hotels in the destination D3 may be provided in the decreasing order of the hotel score that is calculated based on the number of bookings and clicks.

The recommendation engine for Use case 1A is described in details in various steps:

Step1: Map reduce program is used to determine the scores of the unique destination and hotel for each user. In this case, scores for hotel are determined based on number of bookings and number of clicks by a user.

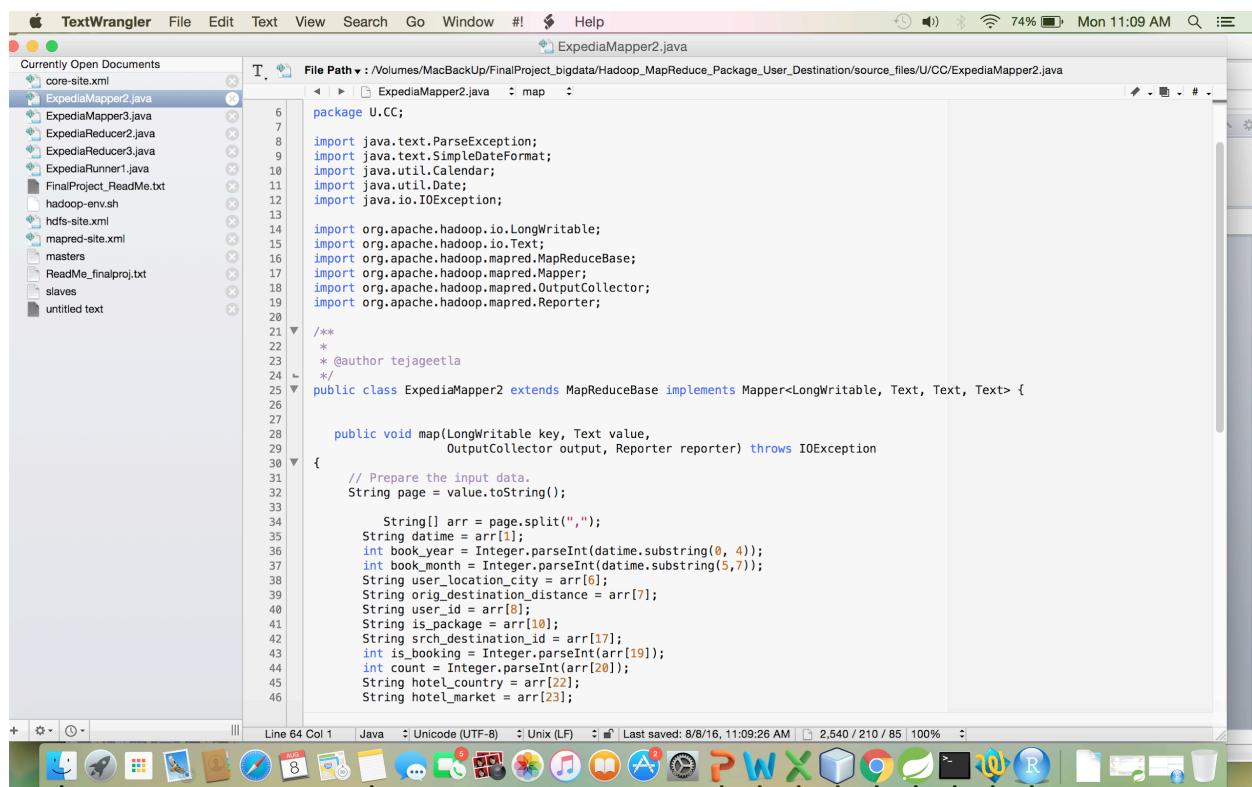
Hotel_score = 5*(No of bookings) OR 1*(No of clicks)

Step2: The output of step1 is served as input to mahout user-based recommender engine. User similarity co-occurrence matrix is used in finding the similar users using the scores calculated in step1. As a result a set of recommendations with destination and hotelId is generated for each user.

Step1: Map Reduce

ExpediaMapper2.java

1. Extract the destination id which is emitted as key in the output
2. Extract the hotel cluster id, is_booking, count and time stamp and add it to string builder.
3. Emit (destinationid, hotelclusterId isbooking count user_id timestamp)



The screenshot shows a Mac OS X desktop with a TextWrangler window open. The window title is "ExpediaMapper2.java". The code editor displays Java code for a MapReduce Mapper. The imports at the top include packages for java.util, java.io, org.apache.hadoop.io, org.apache.hadoop.mapred, and org.apache.hadoop.mapred.Mapper. The class definition starts with a package declaration "package U.CC;" and a Javadoc comment. The map method is defined with parameters LongWritable key, Text value, OutputCollector output, and Reporter reporter. Inside the map method, the code reads a line of input, splits it by commas, and extracts various fields such as book_year, book_month, user_location_city, orig_destination_distance, user_id, is_package, srch_destination_id, is_booking, count, hotel_country, and hotel_market. These values are then combined into a single string using a StringBuilder and emitted as the value.

```
package U.CC;

import java.text.ParseException;
import java.text.SimpleDateFormat;
import java.util.Calendar;
import java.util.Date;
import java.io.IOException;

import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.MapReduceBase;
import org.apache.hadoop.mapred.Mapper;
import org.apache.hadoop.mapred.OutputCollector;
import org.apache.hadoop.mapred.Reporter;

public class ExpediaMapper2 extends MapReduceBase implements Mapper<LongWritable, Text, Text, Text> {

    public void map(LongWritable key, Text value,
                    OutputCollector output, Reporter reporter) throws IOException {
        // Prepare the input data.
        String page = value.toString();

        String[] arr = page.split(",");
        String datime = arr[1];
        int book_year = Integer.parseInt(datime.substring(0, 4));
        int book_month = Integer.parseInt(datime.substring(5, 7));
        String user_location_city = arr[6];
        String orig_destination_distance = arr[7];
        String user_id = arr[8];
        String is_package = arr[10];
        String srch_destination_id = arr[12];
        int is_booking = Integer.parseInt(arr[19]);
        int count = Integer.parseInt(arr[20]);
        String hotel_country = arr[22];
        String hotel_market = arr[23];
    }
}
```

```

TextWrangler File Edit Text Search Go Window #! ⌘ Help
File Path : /Volumes/MacBackUp/FinalProject_bigdata/Hadoop_MapReduce_Package_User_Destination/source_files/U/CC/ExpediaMapper2.java
T ExpediaMapper2.java
Currently Open Documents
ExpediaMapper2.java
ExpediaMapper3.java
ExpediaReducer2.java
ExpediaReducer3.java
ExpediaRunner1.java
FinalProject_ReadMe.txt
hadoop-env.sh
hdfs-site.xml
mapred-site.xml
masters
ReadMe_finalproj.txt
slaves
untitled text

43     int is_booking = Integer.parseInt(arr[19]);
44     int count = Integer.parseInt(arr[20]);
45     String hotel_country = arr[22];
46     String hotel_market = arr[23];
47     String hotel_cluster = arr[24];
48     String timeStamp = "";
49     try
50     {
51         SimpleDateFormat format = new SimpleDateFormat("yyyy-MM-dd HH:mm:ss");
52         Date parseDate = format.parse(arr[1].substring(0,19));
53         // System.out.println(parseDate);
54         Calendar c = Calendar.getInstance();
55         c.setTime(parseDate);
56         long time = c.getTimeInMillis();
57         timeStamp = String.valueOf(time);
58         // System.out.println("ts"+time);
59     }
60     catch(ParseException pe)
61     {
62         pe.printStackTrace();
63     }
64
65
66     StringBuilder builder = new StringBuilder();
67     builder.append(hotel_cluster);
68     builder.append(" ");
69     builder.append(is_booking);
70     builder.append(" ");
71     builder.append(count);
72     builder.append(" ");
73     builder.append(user_id);
74     builder.append(" ");
75     builder.append(timeStamp);
76
77
78     output.collect(new Text(srch_destination_id), new Text(builder.toString()));
79 }
80
81
82
83

```

Line 64 Col 1 Java Unicode (UTF-8) Unix (LF) Last saved: 8/8/16, 11:09:26 AM 2,540 / 210 / 85 100%

ExpediaReducer2.java

1. Extract the user id from the values, which will be used as key.
2. Calculate the score for each destination. For booking multiply the number of bookings by 5 and for clicks multiply the number of clicks by 1.
3. Append the destination id with the hotel cluster id and the score to emit as value
4. Emit (user_id, destination id hotelclusterId score timestamp)

TextWrangler

File Edit Text Search Go Window #! ⚡ Help

ExpediaReducer2.java

File Path : /Volumes/MacBackUp/FinalProject_bigdata/Hadoop_MapReduce_Package_User_Destination/source_files/U/CC/ExpediaReducer2.java

Currently Open Documents

- core-site.xml
- ExpediaMapper2.java
- ExpediaMapper3.java
- ExpediaReducer2.java
- ExpediaReducer3.java
- ExpediaRunner1.java
- FinalProject_ReadMe.txt
- hadoop-env.sh
- hdfs-site.xml
- mapped-site.xml
- masters
- ReadMe_finalproj.txt
- slaves
- untitled text

```
20
21  * @author tejageetla
22  */
23  public class ExpediaReducer2 extends MapReduceBase implements Reducer<Text, Text, Text, Text>
24  {
25
26      public void reduce(Text key, Iterator<Text> values,
27                         OutputCollector<Text, Text> output, Reporter reporter) throws IOException {
28
29          reporter.setStatus(key.toString());
30
31          String newKey ;
32          // double count = 0;
33          String hotel_clusterId = "";
34          double total=0;
35          while(values.hasNext())
36          {
37              StringBuilder toWrite = new StringBuilder();
38              String page = ((Text)values.next()).toString();
39              String[] valTokens = page.split(" ");
40              int bkcnt = Integer.valueOf(valTokens[2]);
41              String isBooking = valTokens[1];
42              String userId = valTokens[3];
43              String timestamp = valTokens[4];
44              hotel_clusterId = valTokens[0];
45
46              newKey = key.toString() + " " + hotel_clusterId;
47              // toWrite.append(val);
48
49              // System.out.println("val"+newKey);
50
51              if(isBooking.equalsIgnoreCase("1"))
52              {
53                  total = bkcnt*5;
54              }
55              else if(isBooking.equalsIgnoreCase("0"))
56              {
57                  total = bkcnt*1;
58              }
59
60              DoubleWritable i = new DoubleWritable(total);
61
62              String num = (i).toString();
63              toWrite.append(newKey);
64              toWrite.append(" ");
65              // toWrite.append("#");
66              toWrite.append(num);
67              toWrite.append(" ");
68              toWrite.append(timestamp);
69              output.collect(new Text(userId), new Text(toWrite.toString()));
70
71
72              // System.out.println("toWrite before val" + toWrite.toString());
73
74              // System.out.println("toWrite after val" + toWrite.toString());
75              // System.out.println("Reducer >>> value "+toWrite );
76              // output.collect(key, new Text(toWrite.toString()));
77
78      }
79  }
```

Line 65 Col 11 Java Unicode (UTF-8) Unix (LF) Last saved: 8/5/16, 10:37:15 PM 2,409 / 186 / 81 100%

TextWrangler

File Edit Text Search Go Window #! ⚡ Help

ExpediaReducer2.java

File Path : /Volumes/MacBackUp/FinalProject_bigdata/Hadoop_MapReduce_Package_User_Destination/source_files/U/CC/ExpediaReducer2.java

Currently Open Documents

- core-site.xml
- ExpediaMapper2.java
- ExpediaMapper3.java
- ExpediaReducer2.java
- ExpediaReducer3.java
- ExpediaRunner1.java
- FinalProject_ReadMe.txt
- hadoop-env.sh
- hdfs-site.xml
- mapped-site.xml
- masters
- ReadMe_finalproj.txt
- slaves
- untitled text

```
39
40      String[] valTokens = page.split(" ");
41      int bkcnt = Integer.valueOf(valTokens[2]);
42      String isBooking = valTokens[1];
43      String userId = valTokens[3];
44      String timestamp = valTokens[4];
45      hotel_clusterId = valTokens[0];
46
47      newKey = key.toString() + " " + hotel_clusterId;
48      // toWrite.append(val);
49
50      // System.out.println("val"+newKey);
51
52      if(isBooking.equalsIgnoreCase("1"))
53      {
54          total = bkcnt*5;
55      }
56      else if(isBooking.equalsIgnoreCase("0"))
57      {
58          total = bkcnt*1;
59      }
59
60      DoubleWritable i = new DoubleWritable(total);
61      String num = (i).toString();
62      toWrite.append(newKey);
63      toWrite.append(" ");
64      // toWrite.append("#");
65      toWrite.append(num);
66      toWrite.append(" ");
67      toWrite.append(timestamp);
68      output.collect(new Text(userId), new Text(toWrite.toString()));
69
70
71      // System.out.println("toWrite before val" + toWrite.toString());
72
73      // System.out.println("toWrite after val" + toWrite.toString());
74      // System.out.println("Reducer >>> value "+toWrite );
75      // output.collect(key, new Text(toWrite.toString()));
77
78  }
```

Line 65 Col 11 Java Unicode (UTF-8) Unix (LF) Last saved: 8/5/16, 10:37:15 PM 2,409 / 186 / 81 100%

ExpediaMapper3.java

1. Extract the user id from the intermediate output and combine it with the destination id and hotel cluster id to emit as key
2. Combine the score and time stamp using string builder to emit as value
3. Emit((user_id,destinationId:hotelcluster_Id), (score timestamp))

The screenshot shows a Mac OS X desktop with the TextWrangler application open. The window title is "ExpediaMapper3.java". The code editor displays the Java code for the ExpediaMapper3 class. The code includes imports for java.io, org.apache.hadoop.io, org.apache.hadoop.mapreduce, and org.apache.hadoop.mapred. It defines a class ExpediaMapper3 that extends MapReduceBase and implements Mapper<LongWritable, Text, Text, Text>. The map method takes a LongWritable key, a Text value, an OutputCollector output, and a Reporter reporter. Inside the map method, the reporter.setStatus is called with the key.toString(). The value is converted to a String page. The page is split into an array arr of strings. The userDestHot is constructed by concatenating arr[0], arr[1], arr[2], and arr[3]. The toEmit is constructed by concatenating arr[3] and arr[4]. The output.collect method is called with a Text object for userDestHot and another for toEmit. The code editor interface includes a sidebar with other files like core-site.xml and ExpediaReducer2.java, and a status bar at the bottom.

```
2  * To change this license header, choose License Headers in Project Properties.
3  * To change this template file, choose Tools | Templates
4  * and open the template in the editor.
5  */
6 package U.CC;
7
8 import java.io.IOException;
9
10 import org.apache.hadoop.io.LongWritable;
11 import org.apache.hadoop.io.Text;
12 import org.apache.hadoop.mapred.MapReduceBase;
13 import org.apache.hadoop.mapred.Mapper;
14 import org.apache.hadoop.mapred.OutputCollector;
15 import org.apache.hadoop.mapred.Reporter;
16
17 /**
18 *
19 * @author tejageetla
20 */
21 public class ExpediaMapper3 extends MapReduceBase implements Mapper<LongWritable, Text, Text, Text> {
22
23
24     public void map(LongWritable key, Text value,
25                     OutputCollector output, Reporter reporter) throws IOException
26     {
27         reporter.setStatus(key.toString());
28         String page = value.toString();
29         String[] arr = page.split("\\s+");
30
31         String userDestHot = arr[0] + "," + arr[1] + ":" + arr[2];
32         // System.out.println("userDestHot>>>" + userDestHot);
33         String toEmit = arr[3] + " " + arr[4];
34
35         // System.out.println("toEmit>>>" + toEmit);
36
37         output.collect(new Text(userDestHot), new Text(toEmit));
38     }
39
40 }
41 }
```

ExpediaReducer3.java

1. Emit the key same as of mapper3.
2. Add the scores and emit as value.

The screenshot shows a Mac OS X desktop with a TextWrangler window open. The window title is "ExpediaReducer3.java". The code editor displays Java code for a reducer. The code imports org.apache.hadoop.mapred.Reporter and java.lang.StringBuilder. It defines a class ExpediaReducer3 that implements MapReduceBase and Reducer<Text, Text, Text, Text>. The reduce method takes a Text key and an Iterator<Text> values, and outputs to an OutputCollector<Text, Text>. The code iterates through the values, calculating a total score and timestamp. It then outputs a key-value pair where the key is the page URL and the value is the total score and timestamp.

```
import org.apache.hadoop.mapred.Reporter;
import java.lang.StringBuilder;

public class ExpediaReducer3 extends MapReduceBase implements Reducer<Text, Text, Text, Text> {

    public void reduce(Text key, Iterator<Text> values,
                      OutputCollector<Text, Text> output, Reporter reporter) throws IOException {
        reporter.setStatus(key.toString());
        long timestamp = 0;
        long timestamp0 = 0;
        double count = 0;
        String hotel_clusterId = "";
        double total=0;
        while(values.hasNext())
        {
            StringBuilder toWrite = new StringBuilder();
            String page = ((Text)values.next()).toString();
            String[] valTokens = page.split(" ");
            double score = Double.valueOf(valTokens[0]);
            // System.out.println("score : "+score);
            total+= score;
            timestamp = Long.parseLong(valTokens[1]);
            if(timestamp>timestamp0)
            {
                timestamp0 = timestamp;
            }
            toWrite = String.valueOf((int)total) ;
            //+ " " + String.valueOf(timestamp);
            output.collect(key, new Text(toWrite));
        }
    }
}
```

ExpediaRunner1.java:

1. Execute the above jobs to generate results.

TextWrangler

```
File Path : /Volumes/MacBackUp/FinalProject_bigdata/Hadoop_MapReduce_Package_User_Destination/source_files/U/CC/ExpediaRunner1.java
```

```
1  /*
2  * To change this license header, choose License Headers in Project Properties.
3  * To change this template file, choose Tools | Templates
4  * and open the template in the editor.
5  */
6 package U.CC;
7
8 import org.apache.hadoop.fs.Path;
9 import org.apache.hadoop.io.Text;
10 import org.apache.hadoop.mapred.FileInputFormat;
11 import org.apache.hadoop.mapred.FileOutputFormat;
12 import org.apache.hadoop.mapred.JobClient;
13 import org.apache.hadoop.mapred.JobConf;
14
15 /**
16 *
17 * @author tejageetla
18 */
19 public class ExpediaRunner1 {
20
21     // private static String inputPath = "hdfs://ec2-52-43-100-208.us-west-2.compute.amazonaws.com:8021/home/ubuntu/hadoop/train";
22     // private static String outputPath = "hdfs://ec2-52-43-100-208.us-west-2.compute.amazonaws.com:8021/home/ubuntu/hdfstmp/expedia";
23     // private static String temp = "hdfs://ec2-52-43-100-208.us-west-2.compute.amazonaws.com:8021/home/ubuntu/hdfstmp/expedia/temp";
24
25     // private static final String OUTPUT_PATH = "intermediate_output";
26
27     public static void main(String[] args) throws Exception {
28
29         String inputPath = "/Volumes/MacBackUp/trainfullData2.txt";
30         String outputPath = "/Volumes/MacBackUp/expedia/outputExpedia/";
31         String temp = "/Volumes/MacBackUp/expedia/tempExp/";
32
33         // String inputPath = "hdfs://ec2-52-43-100-208.us-west-2.compute.amazonaws.com:8021/home/ubuntu/hadoop/train4.txt";
34         // String outputPath = "hdfs://ec2-52-43-100-208.us-west-2.compute.amazonaws.com:8021/home/ubuntu/hdfstmp/expedia/outputtex/";
35         // String temp = "hdfs://ec2-52-43-100-208.us-west-2.compute.amazonaws.com:8021/home/ubuntu/hdfstmp/expedia/tempex/";
36
37         JobClient client = new JobClient();
38         JobConf conf = new JobConf(ExpediaRunner1.class);
39         conf.setJobName("ExpediaMapper1");
40         conf.setMapperClass(ExpediaMapper2.class);
41         conf.setMapOutputKeyClass(Text.class);
42         conf.setMapOutputValueClass(Text.class);
43         conf.setReducerClass(ExpediaReducer2.class);
44         FileInputFormat.setInputPaths(conf, new Path(inputPath));
45         FileOutputFormat.setOutputPath(conf, new Path(temp));
46
47         client.setConf(conf);
48         try {
49             JobClient.runJob(conf);
50         } catch (Exception e) {
51             e.printStackTrace();
52         }
53
54         // chaining
55
56         JobConf conf1 = new JobConf(ExpediaRunner1.class);
57         conf1.setJobName("ExpediaMapper2");
58         conf1.setMapperClass(ExpediaMapper3.class);
59         conf1.setMapOutputKeyClass(Text.class);
60         conf1.setMapOutputValueClass(Text.class);
61         conf1.setReducerClass(ExpediaReducer3.class);
62         FileInputFormat.setInputPaths(conf1, new Path(temp));
63         FileOutputFormat.setOutputPath(conf1, new Path(outputPath));
64
65         // FileInputFormat.setInputPaths(conf1, new Path(OUTPUT_PATH));
66         // FileOutputFormat.setOutputPath(conf1, new Path(args[1]));
67
68         client.setConf(conf1);
69         try {
70             JobClient.runJob(conf1);
71         } catch (Exception e) {
72             e.printStackTrace();
73         }
74     }
75 }
```

TextWrangler

```
File Path : /Volumes/MacBackUp/FinalProject_bigdata/Hadoop_MapReduce_Package_User_Destination/source_files/U/CC/ExpediaRunner1.java
```

```
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
```

```
Terminal Shell Edit View Window Help hadoop-0.20.2 - bash - 177x48
16/08/06 19:22:20 INFO mapred.LocalJobRunner: reduce > reduce
16/08/06 19:22:23 INFO mapred.LocalJobRunner: reduce > reduce
16/08/06 19:22:26 INFO mapred.LocalJobRunner: reduce > reduce
16/08/06 19:22:29 INFO mapred.LocalJobRunner: reduce > reduce
16/08/06 19:22:32 INFO mapred.LocalJobRunner: reduce > reduce
16/08/06 19:22:35 INFO mapred.LocalJobRunner: reduce > reduce
16/08/06 19:22:38 INFO mapred.LocalJobRunner: reduce > reduce
16/08/06 19:22:39 INFO mapred.TaskRunner: Task:attempt_local_0010_r_00000_0 is done. And is in the process of committing
16/08/06 19:22:39 INFO mapred.TaskRunner: Task attempt_local_0010_r_00000_0 is allowed to commit now
16/08/06 19:22:39 INFO output.FileOutputCommitter: Saved output of task 'attempt_local_0010_r_00000_0' to /Volumes/MacBackup/expedia/mahout_recommend
16/08/06 19:22:39 INFO mapred.LocalJobRunner: reduce > reduce
16/08/06 19:22:39 INFO mapred.TaskRunner: Task 'attempt_local_0010_r_00000_0' done.
16/08/06 19:22:39 INFO mapred.JobClient: map 100% reduce 100%
16/08/06 19:22:39 INFO mapred.JobClient: Job complete: job_local_0010
16/08/06 19:22:39 INFO mapred.JobClient: Counters: 12
16/08/06 19:22:39 INFO mapred.JobClient: Map-Reduce Framework
16/08/06 19:22:39 INFO mapred.JobClient: Combine output records=0
16/08/06 19:22:39 INFO mapred.JobClient: Spilled Records=31838319
16/08/06 19:22:39 INFO mapred.JobClient: Reduce input records=10612773
16/08/06 19:22:39 INFO mapred.JobClient: Reduce output records=166041
16/08/06 19:22:39 INFO mapred.JobClient: Map input records=30024
16/08/06 19:22:39 INFO mapred.JobClient: Map output records=10612773
16/08/06 19:22:39 INFO mapred.JobClient: Map output byts=9518956112
16/08/06 19:22:39 INFO mapred.JobClient: Reduce shuffle bytes=0
16/08/06 19:22:39 INFO mapred.JobClient: Combine input records=0
16/08/06 19:22:39 INFO mapred.JobClient: Reduce input groups=1198786
16/08/06 19:22:39 INFO mapred.JobClient: FileSystemCounters
16/08/06 19:22:39 INFO mapred.JobClient: FILE_BYTES_WRITE=210327848289
16/08/06 19:22:39 INFO mapred.JobClient: FILE_BYTES_READ=196147212385
Tejswaroops-MacBook-Pro:hadoop-0.20.2 tejageetla$ jar cvf exp_destfrmusrCity.jar -C exp_destfrmusrCity/
added manifest
adding: .DS_Store(in = 6148) (out= 465)(deflated 92%)
adding: exp_descriptors/(in = 0) (out= 0)(stored 0%)
adding: exp_descriptors/.DS_Store(in = 6148) (out= 306)(deflated 95%)
adding: exp_descriptors/ExpDestFromUserCity.class(in = 1971) (out= 976)(deflated 50%)
adding: exp_descriptors/ExpMapperDest.class(in = 2599) (out= 1168)(deflated 55%)
adding: exp_descriptors/ExpReducerDest1.class(in = 1545) (out= 680)(deflated 55%)
adding: exp_descriptors/ExpReducerDest.class(in = 5168) (out= 2175)(deflated 57%)
adding: exp_descriptors/ExpReducerDest.class(in = 5264) (out= 2222)(deflated 57%)
Tejswaroops-MacBook-Pro:hadoop-0.20.2 tejageetla$ bin/hadoop jar exp_destfrmusrCity.jar exp_destfrmusrCity.ExpDestFromUserCity
added manifest
adding: .DS_Store(in = 6148) (out= 178)(deflated 97%)
adding: exp_descriptors/(in = 0) (out= 0)(stored 0%)
adding: exp_descriptors/ExpDestFromUserCity.class(in = 1996) (out= 995)(deflated 50%)
adding: exp_descriptors/ExpMapperDest.class(in = 2680) (out= 1209)(deflated 54%)
adding: exp_descriptors/ExpReducerDest1.class(in = 1589) (out= 701)(deflated 55%)
adding: exp_descriptors/ExpReducerDest.class(in = 5264) (out= 2222)(deflated 57%)
Tejswaroops-MacBook-Pro:hadoop-0.20.2 tejageetla$
```

```
Terminal Shell Edit View Window Help hadoop-0.20.2 - java - 177x48
16/08/07 00:09:14 INFO mapred.MapTask: kvstart = 0; kvend = 262144; length = 327680
16/08/07 00:09:15 INFO mapred.MapTask: Starting flush of map output
16/08/07 00:09:15 INFO mapred.MapTask: Finished spill 0
16/08/07 00:09:15 INFO mapred.MapTask: Finished spill 1
16/08/07 00:09:15 INFO mapred.Merger: Merging 2 sorted segments
16/08/07 00:09:15 INFO mapred.Merger: Down to the last merge-pass, with 2 segments left of total size: 10973993 bytes
16/08/07 00:09:15 INFO mapred.TaskRunner: Task:attempt_local_0001_m_000001_0 is done. And is in the process of committing
16/08/07 00:09:15 INFO mapred.LocalJobRunner: file:/Volumes/MacBackup/trainfullData2.txt:33554432+33554432
16/08/07 00:09:15 INFO mapred.TaskRunner: Task 'attempt_local_0001_m_000001_0' done.
16/08/07 00:09:15 INFO mapred.MapTask: numReduceTasks: 1
16/08/07 00:09:15 INFO mapred.MapTask: io.sort.mb = 100
16/08/07 00:09:15 INFO mapred.MapTask: data buffer = 79691776/99614720
16/08/07 00:09:15 INFO mapred.MapTask: record buffer = 262144/327680
16/08/07 00:09:16 INFO mapred.MapTask: Spilling map output: record full = true
16/08/07 00:09:16 INFO mapred.MapTask: bufstart = 0; bufend = 9350320; bufvoid = 99614720
16/08/07 00:09:16 INFO mapred.MapTask: kvstart = 0; kvend = 262144; length = 327680
16/08/07 00:09:16 INFO mapred.MapTask: Starting flush of map output
16/08/07 00:09:16 INFO mapred.MapTask: Finished spill 0
16/08/07 00:09:16 INFO mapred.MapTask: Finished spill 1
16/08/07 00:09:16 INFO mapred.Merger: Merging 2 sorted segments
16/08/07 00:09:16 INFO mapred.Merger: Down to the last merge-pass, with 2 segments left of total size: 10944761 bytes
16/08/07 00:09:16 INFO mapred.TaskRunner: Task:attempt_local_0001_m_000002_0 is done. And is in the process of committing
16/08/07 00:09:16 INFO mapred.LocalJobRunner: file:/Volumes/MacBackup/trainfullData2.txt:67108864+33554432
16/08/07 00:09:16 INFO mapred.TaskRunner: Task 'attempt_local_0001_m_000002_0' done.
16/08/07 00:09:16 INFO mapred.MapTask: numReduceTasks: 1
16/08/07 00:09:16 INFO mapred.MapTask: io.sort.mb = 100
16/08/07 00:09:16 INFO mapred.MapTask: data buffer = 79691776/99614720
16/08/07 00:09:16 INFO mapred.MapTask: record buffer = 262144/327680
16/08/07 00:09:17 INFO mapred.MapTask: Spilling map output: record full = true
16/08/07 00:09:17 INFO mapred.MapTask: bufstart = 0; bufend = 9405910; bufvoid = 99614720
16/08/07 00:09:17 INFO mapred.MapTask: kvstart = 0; kvend = 262144; length = 327680
16/08/07 00:09:17 INFO mapred.MapTask: Starting flush of map output
16/08/07 00:09:17 INFO mapred.MapTask: Finished spill 0
16/08/07 00:09:17 INFO mapred.MapTask: Finished spill 1
16/08/07 00:09:17 INFO mapred.Merger: Merging 2 sorted segments
16/08/07 00:09:17 INFO mapred.Merger: Down to the last merge-pass, with 2 segments left of total size: 10924445 bytes
16/08/07 00:09:17 INFO mapred.TaskRunner: Task:attempt_local_0001_m_000003_0 is done. And is in the process of committing
16/08/07 00:09:17 INFO mapred.LocalJobRunner: file:/Volumes/MacBackup/trainfullData2.txt:100663296+33554432
16/08/07 00:09:17 INFO mapred.TaskRunner: Task 'attempt_local_0001_m_000003_0' done.
16/08/07 00:09:17 INFO mapred.MapTask: numReduceTasks: 1
16/08/07 00:09:17 INFO mapred.MapTask: io.sort.mb = 100
16/08/07 00:09:17 INFO mapred.MapTask: data buffer = 79691776/99614720
16/08/07 00:09:17 INFO mapred.MapTask: record buffer = 262144/327680
16/08/07 00:09:18 INFO mapred.MapTask: Spilling map output: record full = true
16/08/07 00:09:18 INFO mapred.MapTask: bufstart = 0; bufend = 9440168; bufvoid = 99614720
16/08/07 00:09:18 INFO mapred.MapTask: kvstart = 0; kvend = 262144; length = 327680
16/08/07 00:09:18 INFO mapred.MapTask: Starting flush of map output
```

A screenshot of a Mac OS X desktop environment. At the top is a Dock bar with various icons. Above the Dock is a menu bar with 'Terminal', 'Shell', 'Edit', 'View', 'Window', and 'Help'. A search bar is also present. The main window is a Terminal session titled 'hadoop-0.20.2 – bash – 163x40'. The terminal output shows logs from a Hadoop job, including map and reduce stages, task attempts, and file output. Below the terminal is a NetBeans IDE window showing a Java project structure with files like 'Test Packages', 'Libraries', 'Test Libraries', and 'TestFRApp'. A status bar at the bottom indicates the current time as 19:33/11:19 and battery level at 66%.

```

16/08/07 00:18:35 INFO mapred.LocalJobRunner: 66_174_40600_50 > reduce
16/08/07 00:18:36 INFO mapred.JobClient: map 100% reduce 81%
16/08/07 00:18:38 INFO mapred.LocalJobRunner: 66_220_3781_50 > reduce
16/08/07 00:18:39 INFO mapred.JobClient: map 100% reduce 84%
16/08/07 00:18:41 INFO mapred.LocalJobRunner: 66_311_48189_8 > reduce
16/08/07 00:18:42 INFO mapred.JobClient: map 100% reduce 87%
16/08/07 00:18:44 INFO mapred.LocalJobRunner: 66_346_31371_128 > reduce
16/08/07 00:18:45 INFO mapred.LocalJobRunner: 66_363_36597_50 > reduce
16/08/07 00:18:47 INFO mapred.LocalJobRunner: 66_462_27117_50 > reduce
16/08/07 00:18:48 INFO mapred.JobClient: map 100% reduce 92%
16/08/07 00:18:50 INFO mapred.LocalJobRunner: 66_462_27117_50 > reduce
16/08/07 00:18:51 INFO mapred.JobClient: map 100% reduce 95%
16/08/07 00:18:53 INFO mapred.LocalJobRunner: 69_994_22525_50 > reduce
16/08/07 00:18:54 INFO mapred.JobClient: map 100% reduce 98%
16/08/07 00:18:55 INFO mapred.TaskRunner: Task attempt_local_0001_r_000000_0 is done. And is in the process of committing
16/08/07 00:18:55 INFO mapred.LocalJobRunner: 69_994_22525_50 > reduce
16/08/07 00:18:55 INFO mapred.TaskRunner: Task attempt_local_0001_r_000000_0 is allowed to commit now
16/08/07 00:18:55 INFO mapred.FileOutputCommitter: Saved output of task 'attempt_local_0001_r_000000_0' to file:/Volumes/MacBackUp/expedia/output_HtlDestUsrCityRec
on
16/08/07 00:18:55 INFO mapred.LocalJobRunner: 9_29_3697_8 > reduce
16/08/07 00:18:55 INFO mapred.TaskRunner: Task attempt_local_0001_r_000000_0 done.
16/08/07 00:18:55 INFO mapred.JobClient: map 100% reduce 100%
16/08/07 00:18:56 INFO mapred.JobClient: Job completed job_local_0001
16/08/07 00:18:56 INFO mapred.JobClient: Counters: 13
16/08/07 00:18:56 INFO mapred.JobClient: Map-Reduce Framework
16/08/07 00:18:56 INFO mapred.JobClient: Combine output records=0
16/08/07 00:18:56 INFO mapred.JobClient: Spilled Records=160656493
16/08/07 00:18:56 INFO mapred.JobClient: Reduce input records=37670293
16/08/07 00:18:56 INFO mapred.JobClient: Reduce output records=602195
16/08/07 00:18:56 INFO mapred.JobClient: Map input records=37670293
16/08/07 00:18:56 INFO mapred.JobClient: Map output records=37670293
16/08/07 00:18:56 INFO mapred.JobClient: Map output bytes=1353063468
16/08/07 00:18:56 INFO mapred.JobClient: Reduce shuffle bytes=0
16/08/07 00:18:56 INFO mapred.JobClient: Combine input records=0
16/08/07 00:18:56 INFO mapred.JobClient: Map input bytes=425411958
16/08/07 00:18:56 INFO mapred.JobClient: Reduce input groups=602195
16/08/07 00:18:56 INFO mapred.JobClient: FileSystemCounters
16/08/07 00:18:56 INFO mapred.JobClient: FILE_BYTES_WRITTEN=195152991184
16/08/07 00:18:56 INFO mapred.JobClient: FILE_BYTES_READ=398893279594
Tejageetela-MacBook-Pro:hadoop-0.20.2 tejageetela$ 
```

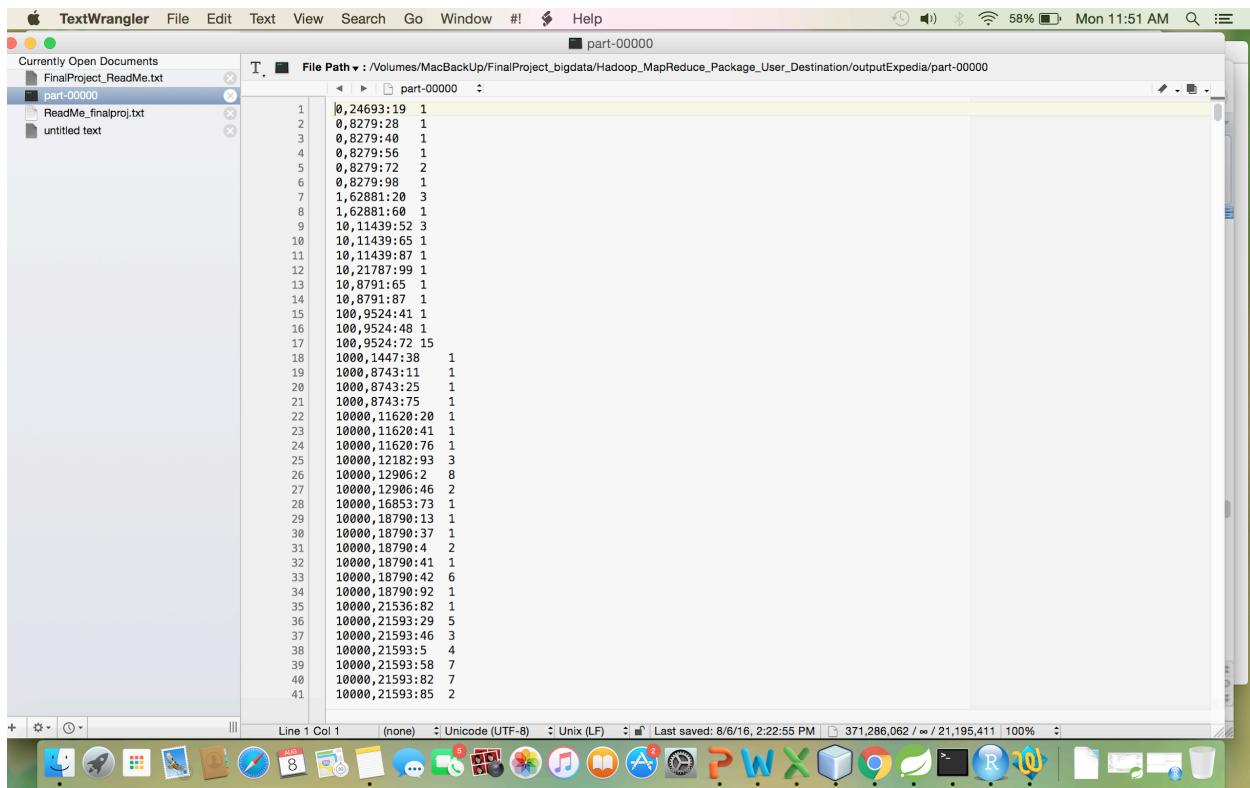
Output:

Intermediate Output:

A screenshot of a Mac OS X desktop environment. The main window is a TextWrangler editor showing a file named 'part-00000'. The file contains a large amount of tab-separated data. The columns include integers (e.g., 1, 2, 3, 4, 5, 6) and floating-point numbers (e.g., 11879, 12, 2.0, 1365140060000). The data appears to be a sequence of records, likely generated by a MapReduce job's map phase. The TextWrangler interface includes a toolbar, a menu bar with 'TextWrangler', 'File', 'Edit', 'Text', 'View', 'Search', 'Go', 'Window', '#!', 'Help', and a status bar at the bottom.

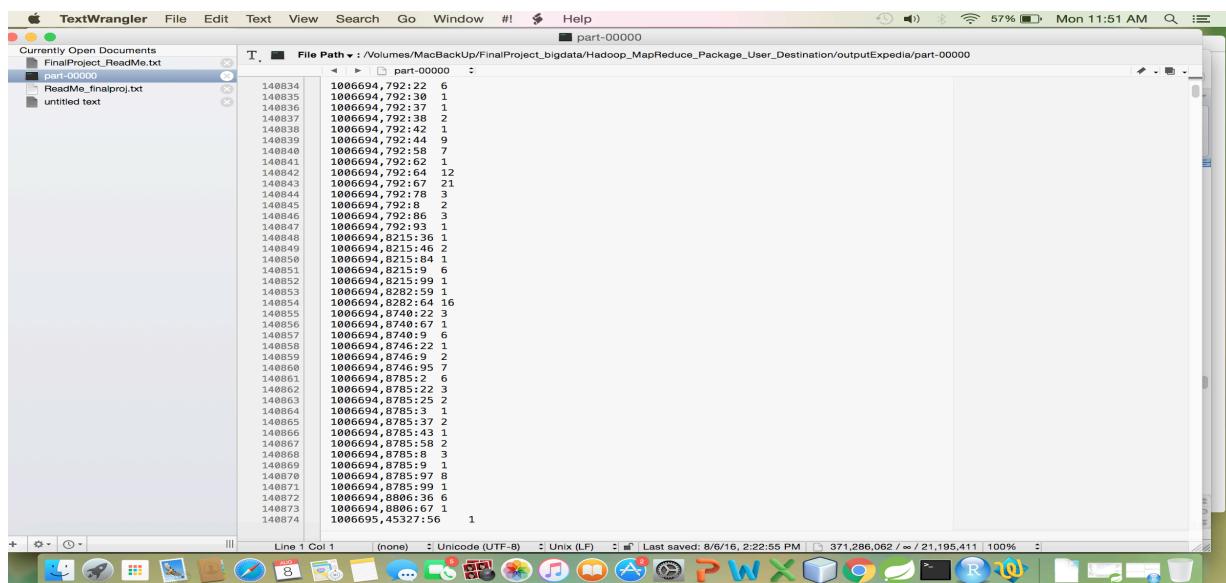
| 1 | 4413 | 11879 | 12 | 2.0 | 1365140060000 |
|----|------|-------|----|-----|-----------------|
| 2 | 4413 | 11879 | 12 | 5.0 | 136513118940000 |
| 3 | 4413 | 11879 | 12 | 1.0 | 13651311433000 |
| 4 | 6458 | 11361 | 32 | 1.0 | 14025986660000 |
| 5 | 6458 | 11361 | 99 | 1.0 | 14025987010000 |
| 6 | 6458 | 11361 | 99 | 1.0 | 14025987010000 |
| 7 | 5778 | 11439 | 65 | 7.0 | 13894111330000 |
| 8 | 5778 | 11439 | 52 | 1.0 | 13894110710000 |
| 9 | 5778 | 11439 | 65 | 6.0 | 13894110710000 |
| 10 | 4539 | 11439 | 65 | 1.0 | 13987018000000 |
| 11 | 4539 | 11439 | 65 | 3.0 | 13987000060000 |
| 12 | 6458 | 1152 | 8 | 1.0 | 14048433000000 |
| 13 | 6458 | 1152 | 66 | 1.0 | 14048405137000 |
| 14 | 5677 | 11835 | 54 | 1.0 | 14168738800000 |
| 15 | 3313 | 11835 | 54 | 1.0 | 14168738800000 |
| 16 | 3313 | 11835 | 54 | 1.0 | 14068604940000 |
| 17 | 3313 | 11835 | 54 | 1.0 | 14132280550000 |
| 18 | 3313 | 11835 | 54 | 1.0 | 14132280550000 |
| 19 | 3313 | 11835 | 54 | 1.0 | 14132282450000 |
| 20 | 3313 | 11835 | 76 | 1. | 14132282800000 |
| 21 | 3313 | 11835 | 17 | 1.0 | 14132283050000 |
| 22 | 3313 | 11835 | 17 | 1.0 | 14132283370000 |
| 23 | 3313 | 11835 | 54 | 1.0 | 14136713310000 |
| 24 | 3313 | 11835 | 54 | 4.0 | 14136713310000 |
| 25 | 3313 | 11835 | 54 | 2.0 | 14136714100000 |
| 26 | 3313 | 11835 | 54 | 2.0 | 14136714590000 |
| 27 | 3313 | 11835 | 54 | 2.0 | 14136714590000 |
| 28 | 3313 | 11835 | 17 | 2.0 | 14191983980000 |
| 29 | 3313 | 11835 | 54 | 2.0 | 14191969140000 |
| 30 | 3313 | 11835 | 54 | 3.0 | 14191967740000 |
| 31 | 6304 | 11938 | 80 | 1.0 | 14072727960000 |
| 32 | 6304 | 11938 | 80 | 4.0 | 14072727960000 |
| 33 | 6458 | 11938 | 80 | 1.0 | 14072538330000 |
| 34 | 6458 | 11938 | 80 | 2.0 | 14073201410000 |
| 35 | 6458 | 11938 | 80 | 5.0 | 14073224500000 |
| 36 | 3925 | 12804 | 32 | 5.0 | 13994872650000 |
| 37 | 3925 | 12804 | 32 | 3.0 | 13994870830000 |
| 38 | 3925 | 12804 | 32 | 1.0 | 13887848660000 |
| 39 | 1482 | 12809 | 95 | 1.0 | 13887848660000 |
| 40 | 1482 | 12809 | 95 | 1.0 | 13852874840000 |
| 41 | 1482 | 12809 | 95 | 1.0 | 13852874840000 |

Final Output:



The screenshot shows a Mac OS X desktop with two TextWrangler windows open. Both windows display the same file, "part-00000", which contains a large dataset of user-destination pairs. The data consists of two columns separated by a colon. The first column is a user ID (e.g., 1000, 1000, 1000, ..., 10000) and the second column is a destination ID (e.g., 19, 28, 40, ..., 85). The file is 371,286,062 bytes large and contains 21,195,411 lines.

| User ID | Destination ID |
|----------------|----------------|
| 1000,1447:38 | 1 |
| 1000,8743:11 | 1 |
| 1000,8743:25 | 1 |
| 1000,8743:75 | 1 |
| 10000,11620:20 | 1 |
| 10000,11620:41 | 1 |
| 10000,11620:76 | 1 |
| 10000,12182:93 | 3 |
| 10000,12906:2 | 8 |
| 10000,12906:46 | 2 |
| 10000,16853:73 | 1 |
| 10000,18790:13 | 1 |
| 10000,18790:37 | 1 |
| 10000,18790:4 | 2 |
| 10000,18790:41 | 1 |
| 10000,18790:42 | 6 |
| 10000,18790:92 | 1 |
| 10000,21536:62 | 1 |
| 10000,21593:29 | 5 |
| 10000,21593:46 | 3 |
| 10000,21593:5 | 4 |
| 10000,21593:58 | 7 |
| 10000,21593:82 | 7 |
| 10000,21593:85 | 2 |



The screenshot shows a Mac OS X desktop with two TextWrangler windows open. Both windows display the same file, "part-00000", containing a large dataset of user-destination pairs. The data consists of two columns separated by a colon. The first column is a user ID (e.g., 10000, 10000, 10000, ..., 100000) and the second column is a destination ID (e.g., 792, 792, 792, ..., 8746). The file is 371,286,062 bytes large and contains 21,195,411 lines.

| User ID | Destination ID |
|----------------|----------------|
| 10000,9215:2 | 6 |
| 10000,9215:30 | 1 |
| 10000,9215:37 | 1 |
| 10000,9215:38 | 2 |
| 10000,9215:42 | 1 |
| 10000,9215:44 | 9 |
| 10000,9215:58 | 7 |
| 10000,9215:62 | 1 |
| 10000,9215:64 | 12 |
| 10000,9215:67 | 21 |
| 10000,9215:78 | 3 |
| 10000,9215:8 | 2 |
| 10000,9215:86 | 3 |
| 10000,9215:93 | 1 |
| 10000,9215:96 | 1 |
| 10000,9215:96 | 1 |
| 10000,9215:46 | 2 |
| 10000,9215:46 | 2 |
| 10000,9215:84 | 1 |
| 10000,9215:51 | 6 |
| 10000,9215:99 | 1 |
| 10000,9222:59 | 1 |
| 10000,9222:64 | 16 |
| 10000,9240:22 | 3 |
| 10000,9240:67 | 1 |
| 10000,9240:67 | 1 |
| 10000,9240:6 | 6 |
| 10000,9240:22 | 1 |
| 10000,9240:9 | 2 |
| 10000,9240:95 | 7 |
| 10000,9260:6 | 6 |
| 10000,9785:22 | 3 |
| 10000,9785:25 | 2 |
| 10000,9785:1 | 1 |
| 10000,9785:37 | 2 |
| 10000,9785:43 | 1 |
| 10000,9785:58 | 2 |
| 10000,9785:8 | 3 |
| 10000,9785:1 | 1 |
| 10000,9785:97 | 8 |
| 10000,9785:99 | 1 |
| 10000,8806:36 | 6 |
| 10000,8806:67 | 1 |
| 10000,45327:56 | 1 |

The screenshot shows a Mac OS X desktop with a TextWrangler window open. The window title is "part-00000". The file path is "/Volumes/MacBackUp/FinalProject_bigrdata/Hadoop_MapReduce_Package_User_Destination/outputExpedia/part-00000". The text content of the file is a log of user similarity values, with each line representing a user ID followed by a colon-separated list of other user IDs and their similarity scores. The log starts with user 21195371 and ends with user 21195411.

```

21195371 999994,8745:21 1
21195372 999994,8745:41 1
21195373 999994,8745:47 1
21195374 999994,8745:91 11
21195375 999995,12000:13 1
21195376 999995,12000:32 4
21195377 999995,12000:39 13
21195378 999995,12000:76 4
21195379 999995,13221:23 1
21195380 999995,13221:39 8
21195381 999995,13237:10 1
21195382 999995,13237:14 1
21195383 999995,13237:39 7
21195384 999995,13237:50 8
21195385 999995,44691:32 1
21195386 999995,44691:39 14
21195387 999996,1901:17 1
21195388 999996,1901:47 1
21195389 999996,1901:55 1
21195390 999996,1901:68 1
21195391 999997,8278:51 1
21195392 999997,8278:95 15
21195393 999997,8278:98 3
21195394 999998,63924:50 3
21195395 999999,8255:46 9
21195396 999999,8255:64 7
21195397 999999,8255:69 2
21195398 999999,8255:97 2
21195399 999999,8278:69 3
21195400 999999,8278:92 1
21195401 999999,8788:11 1
21195402 999999,8788:25 1
21195403 999999,8788:29 2
21195404 999999,8788:37 1
21195405 999999,8788:59 18
21195406 999999,8788:64 14
21195407 999999,8788:68 1
21195408 999999,8788:8 5
21195409 999999,8788:86 1
21195410 999999,8788:97 1
21195411

```

Step2: Mahout User-Based Recommender

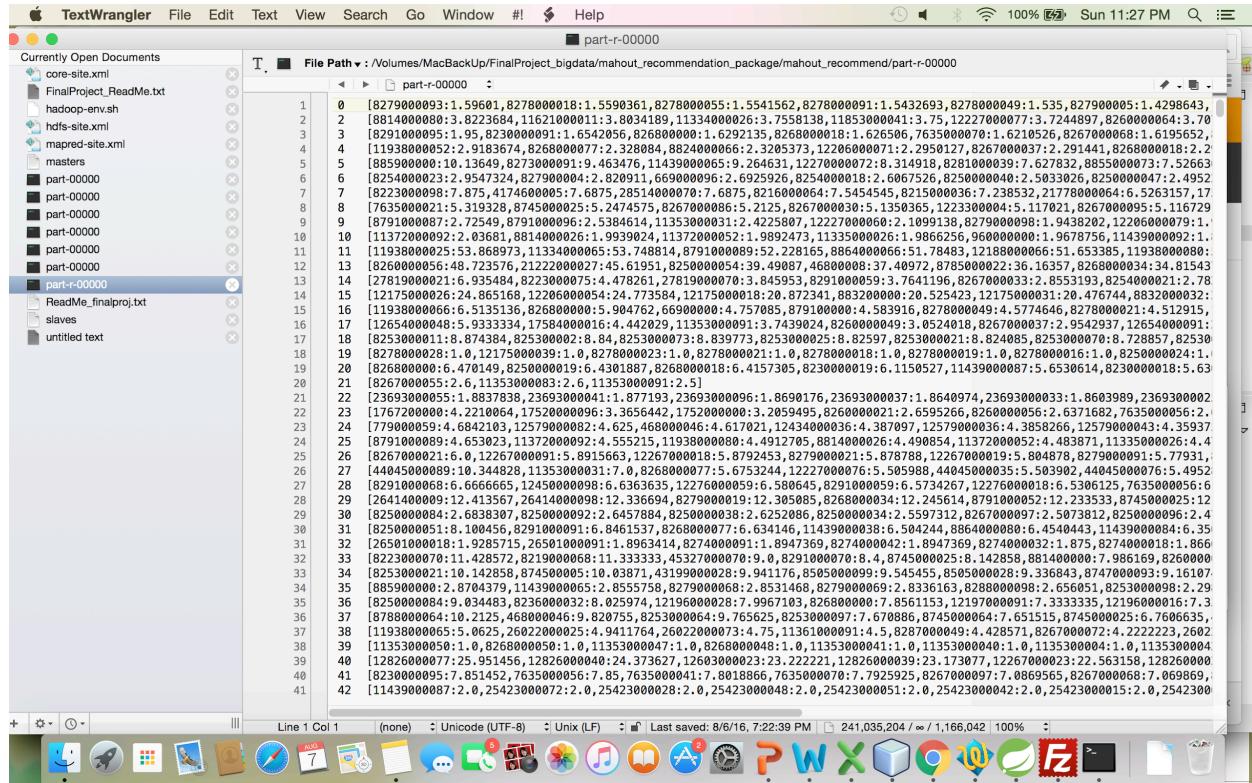
The final output of step1 is served as input to mahout recommendation. Mahout recommends the destination with hotel package using user similarity co-occurrence matrix.

```

bin/hadoop jar mahout-core-0.7-job.jar
org.apache.mahout.cf.taste.hadoop.item.RecommenderJob -s
SIMILARITY_COOCURRENCE --input
/Volumes/MacBackUp/expedia/mahFullInput.txt --output
/Volumes/MacBackUp/expedia/mahout_recommend

```

Output:



The screenshot shows a Mac OS X desktop with the TextWrangler application open. The window title is "part-r-00000". The file path is "/Volumes/MacBackUp/FinalProject_b/gdata/mahout_recommendation_package/mahout_recommend/part-r-00000". The text area contains a large list of recommendations, each consisting of a user ID, item ID, and a score. The list is sorted by score in descending order. The items listed include various file paths and IDs such as "/Volumes/MacBackUp/FinalProject_b/gdata/mahout_recommend/part-r-00000", "8279000093:1, 59661, 8278000018:1, 5590361, 8278000051:1, 5541562, 8278000091:1, 5432693, 8278000049:1, 535, 827900005:1, 4298643, 8821400080:3, 8223684, 11621000011:3, 8034189, 11334000026:3, 7558138, 11853000041:3, 75, 1227000077:3, 7424987, 8260000064:3, 70, 1291000095:1, 95, 8230000091:1, 6542056, 8268000001:6, 6292135, 8268000018:1, 626506, 76350000070:1, 6210526, 8267000068:1, 6195652, 11038000052:2, 0183674, 8268000077:2, 323884, 8824000061:2, 3205373, 1226000071:2, 2950127, 8267000037:2, 201441, 8268000018:2, 2, 1105900002:10, 13649, 8278000091:9, 463476, 11439000065:9, 264633, 1227000072:8, 314918, 8281000039:7, 627832, 8855000073:7, 52663, 8254000023:6, 1820911, 6690000096:2, 6925926, 8254000018:2, 6067526, 6250000040:2, 503026, 8250000047:2, 4952, 8223000098:7, 875, 4174600085:7, 6875, 2851400070:7, 5454545, 8215000036:7, 238532, 2177800064:6, 5263157, 17, 8763500002:5, 219328, 8745000025:5, 2474575, 8267000086:5, 2125, 8267000086:5, 1350365, 1223300004:5, 117021, 8267000095:5, 116729, 8791000087:2, 72549, 8791000086:2, 5384614, 11353000031:2, 4225807, 1227000058:2, 1099138, 8279000098:1, 9438202, 12265000079:1, 11372000092:1, 03681, 8814000026:1, 9939024, 11335000026:1, 9892473, 11335000026:1, 9866256, 960000000:1, 9678756, 11439000092:1, 11193800025:52, 868973, 11334000065:52, 748814, 8791000089:52, 228165, 8864000066:51, 653383, 11938000080: 11, 12871900021:6, 935484, 8223000075:4, 478261, 27819000070:3, 845593, 8291000059:3, 7641198, 8267000033:2, 8553193, 8254000021:2, 78, 12175000026:24, 8651568, 12206000054:24, 773584, 12175000018:26, 872341, 8832000002:26, 525423, 12175000031:26, 476744, 8832000032:1, 11938000066:6, 5135136, 8268000005:5, 904762, 669000008:4, 577085, 8791000004:4, 583916, 8278000049:4, 5774646, 8278000021:4, 512915, 12654000048:5, 933334, 17584000026:4, 4424018, 8267000086:5, 2125, 8267000086:5, 1350365, 1223300004:5, 117021, 8267000087:2, 9542937, 12654000091:1, 18253000011:8, 874384, 825300002:8, 8.839773, 8253000025:8, 82597, 8253000021:8, 824085, 8253000070:8, 728857, 82530, 18278000028:1, 0, 12175000031:1, 0, 8278000023:1, 0, 8278000019:1, 0, 8278000016:1, 0, 8250000024:1, 1826800006:6, 470149, 8250000019:6, 4301887, 8268000018:6, 4157305, 8230000019:6, 1150527, 11439000087:5, 6530614, 8230000018:20, 218267000055:2, 6, 11353000091:2, 51, 23693000055:1, 8837838, 23693000041:1, 877193, 23693000096:1, 8690176, 23693000037:1, 8640974, 23693000033:1, 8603989, 2369300002, 1767200004:2, 2210064, 1752000006:3, 3656442, 1752000003:3, 2059495, 826000002:2, 659526, 7635000056:2, 6371682, 7635000056:2, 279000059:4, 6842103, 12579000082:4, 625, 468000046:4, 617021, 12434000036:4, 387097, 12579000036:4, 3858266, 12579000043:4, 35937, 8791000089:4, 653023, 11372000092:4, 555215, 11938000080:4, 4, 4912705, 8267000049:3, 4982788, 11372000052:4, 4, 483871, 11335000026:4, 4, 2867000021:6, 8, 12267000091:5, 8915663, 12267000018:5, 8782453, 8279000021:5, 878788, 12267000019:5, 804878, 8279000091:5, 77931, 244845000089:18, 344828, 11353000031:17, 0, 826800007:5, 17, 6573244, 12227000076:5, 505988, 4445000035:5, 503902, 440450000876:5, 4952, 2891000068:6, 12450000098:6, 6, 636363, 12276000056:6, 292641400009:12, 413567, 26414000098:12, 336694, 8279000019:12, 305085, 8791000052:12, 233533, 8745000025:12, 28925000084:2, 6838703, 8250000092:2, 6457894, 8250000038:2, 6252086, 8250000034:2, 5597312, 8267000097:2, 5073812, 8250000096:2, 4, 31825000051:8, 100456, 8291000091:6, 8461537, 8268000077:6, 634146, 11439000038:6, 504244, 8864000080:6, 4540443, 11439000084:6, 35312, 26501000018:1, 9285715, 26501000091:1, 8963414, 8274000091:1, 8947369, 8274000042:1, 8947369, 8274000032:1, 875, 8274000018:1, 866, 338223000070:11, 428572, 8219000068:11, 333333, 45327000078:9, 0, 8291000078:8, 4, 8745000025:8, 142858, 881400000:7, 986159, 8260000034, 348253000021:10, 142858, 8745000005:10, 03871, 43199000028:9, 941176, 8505000099:9, 545455, 8505000028:9, 3, 336843, 8747000093:9, 16107, 35885900000:2, 8704379, 11439000065:2, 8555758, 8279000068:2, 8511461, 8279000069:2, 8330163, 8288000098:2, 656581, 8253000098:2, 29356, 8250000084:9, 034483, 8236000032:8, 025974, 12196000028:7, 9961103, 8268000007:7, 8561153, 12196000016:7, 3, 3787800064:10, 2125, 468000046:9, 8287055, 8253000064:9, 765625, 8253000097:7, 670886, 8745000064:7, 651515, 8745000025:6, 7606635, 38311926000065:5, 0625, 26022000025:4, 9411764, 26022000073:4, 75, 11361000091:4, 5, 8278000049:4, 428571, 8267000072:4, 222223, 2602, 39111353000050:1, 0, 8268000050:1, 0, 11353000047:1, 0, 8268000048:1, 0, 11353000041:1, 0, 1135300004:1, 0, 1135300004, 40112826000077:25, 951456, 12826000040:24, 373627, 12603000023:23, 222221, 12826000039:23, 173077, 12267000023:22, 563158, 128260000, 41823000095:7, 851452, 7635000056:7, 85, 7635000041:7, 8018866, 7635000070:7, 7925925, 8267000097:7, 0869565, 8267000068:7, 069869, 4211439000087:2, 0, 25423000072:2, 0, 25423000028:2, 0, 25423000048:2, 0, 25423000051:2, 0, 25423000042:2, 0, 25423000015:2, 0, 254230000

Use Case 1B: Hotel Recommendation

In another aspect, if the user has chosen a destination, recommendations for hotels in that destination may be provided. For such application, the program determines if the similar users identified as per the above mentioned parameters have bookings in the chosen destination. The hotels chosen by these similar users may be recommended to the user in decreasing order of the hotel score. Further, remaining hotels in the chosen destination may be provided in the decreasing order of number of bookings.

However, if there are no similar users with the booking in the chosen destination, all the hotels in the chosen destination may be recommended in the decreasing order of the number of bookings per hotel.

From the output below we can recommend the hotel when a destination is chosen.

```

TextWrangler File Edit Text View Search Go Window #! ⌘ Help
File Path : /Volumes/MacBackUp/FinalProject/_bigdata/mahout_recommendation_package/mahout_recommend/part-r-00000
part-r-00000

Currently Open Documents
core-site.xml
FinalProject_ReadMe.txt
hadoop-env.sh
hdfs-site.xml
mapred-site.xml
masters
part-00000
ReadMe_finalproj.txt
slaves
untitled text

T part-r-00000

1 0 [8279000093:1, 59681, 8278000018:1, 1.5590361, 8278000055:1, 1.5541562, 8278000049:1, 1.535, 827800005:1, 4.229643,
2 [8279000093:1, 8034189, 11334000026:3, 7.558138, 12227000077:3, 7.7244897, 8260000064:3, 7.70
3 [8291000095:1, 95, 8230000091:1, 6542056, 826800000:1, 6292135, 8268000018:1, 626506, 8267000068:1, 6195652,
4 [11938000052:2, 9183674, 826800007:2, 328804, 8824000066:2, 3205373, 12260000071:2, 295127, 8267000037:2, 291441, 8268000018:2, 2.
5 [885900000:1, 136462, 8273000091:9, 463474, 11439000065:9, 264631, 12270000072:8, 314918, 8281000039:7, 6.272832, 8855000073:7, 5.2663
6 [885400000:2, 9547324, 8279000091:9, 463474, 11439000065:9, 264631, 12270000072:8, 314918, 8281000039:7, 6.272832, 8855000073:7, 5.2663
7 [8223000098:7, 875, 4174680005:7, 6875, 28514000070:7, 6875, 8216000064:7, 5454545, 8215000036:7, 238532, 21778000064:6, 5.263157, 17
8 [7635000092:1, 5, 319328, 8745000025:5, 2474575, 8267000086:5, 2125, 8267000030:5, 1350365, 1223300004:5, 117021, 8267000095:5, 116729
9 [8791000087:2, 725495, 4791000096:2, 5384614, 11353000031:2, 2235807, 12227000060:2, 1.199158, 8279000098:1, 9438202, 12206000079:1,
10 [11372000092:2, 03681, 8814000026:1, 9939024, 11372000052:1, 9894473, 11335000026:1, 9866256, 965000000:1, 9678756, 11439000092:1,
11 [11938000025:3, 868973, 11334000065:3, 748814, 8791000089:5, 228165, 8864000066:5, 51, 784843, 12188000065:51, 653385, 11938000080:
12 [8260000056:48, 723576, 21222000027:45, 61951, 8250000054:39, 49087, 468000000:37, 40972, 8785000022:36, 16357, 8268000034:34, 81543
13 [8260000056:48, 723576, 21222000027:45, 61951, 8250000054:39, 49087, 468000000:37, 40972, 8785000022:36, 16357, 8268000034:34, 81543
14 [282540000023:2, 9547324, 8279000075:4, 478261, 27819000070:3, 845953, 8291000059:3, 764194, 8267000033:2, 8553193, 8254000021:2, 78
15 [12175000026:24, 865168, 12206000054:24, 773584, 12175000018:26, 872341, 883200000:24, 525423, 12175000031:24, 476744, 8832000032:
16 [11938000056:6, 5135158, 8268000056:6, 904763, 669000004:6, 4.75085, 8267000052:6, 5.25085, 8278000021:4, 5.12915,
17 [12654000048:5, 933334, 17584000016:4, 442029, 11353000091:3, 7439024, 8260000049:3, 0.0524018, 8267000037:2, 9542937, 12654000091:
18 [8253000011:8, 8278000023:8, 8278000021:8, 8278000019:8, 839773, 82530000021:8, 824085, 8253000070:8, 728857, 82530
19 [8278000028:1, 0, 12175000039:1, 0, 8278000023:1, 0, 8278000021:1, 0, 8278000019:1, 0, 8278000016:1, 0, 8250000024:1,
20 [8268000019:6, 470148, 8250000019:6, 4301887, 8268000018:6, 4157143, 8230000019:6, 1150527, 11439000087:5, 6530614, 8230000018:5, 63
21 [8267000055:2, 6, 11353000083:2, 6, 11353000091:2, 5]
22 [23693000059:1, 8837838, 23693000041:1, 877193, 23693000096:1, 869107, 23693000037:1, 8604974, 23693000033:1, 8603989, 2369300002
23 [1167200000:4, 2210064, 17520000096:4, 2210064, 17520000096:4, 2.3059495, 8260000021:6, 6595266, 8260000056:2, 6371682, 7635000056:2,
24 [7790000059:4, 6842103, 12579000082:4, 625, 468000046:4, 617021, 12434000036:4, 387097, 12579000036:4, 3.358266, 12579000043:4, 35937
25 [8791000089:4, 653023, 11372000092:4, 5.552215, 11938000080:4, 4.912705, 8814000026:4, 4.90854, 11372000052:4, 4.483871, 11335000026:4, 4.4
26 [8267000021:6, 0, 12267000091:5, 8915663, 12267000019:5, 804878, 8279000021:5, 878788, 8279000091:5, 77931,
27 [44045000089:10, 344828, 11353000031:7, 0, 8268000077:5, 6753244, 12227000076:5, 5.059898, 44045000035:5, 5.0392, 44045000076:5, 4.952
28 [2825000068:6, 666666, 12450000098:6, 6363636, 12275000018:6, 5734267, 12275000018:6, 5306125, 7635000056:6
29 [2641400009:12, 413567, 26414000098:12, 336694, 8279000019:12, 336694, 8268000034:12, 245614, 8791000052:12, 233533, 8745000025:12
30 [3825000084:2, 6838307, 8250000092:2, 6457884, 8250000038:2, 5.5597312, 8252000034:2, 5.5597312, 8252000096:2, 4.
31 [8250000051:1, 100456, 8291000091:6, 8461537, 8268000080:6, 6.334146, 11439000038:6, 5.04244, 8864000080:6, 4.4540443, 11439000084:6, 35
32 [26501000018:1, 9.285715, 26501000091:1, 8963414, 8274000091:1, 8947369, 8274000032:1, 875, 8274000018:1, 866
33 [8223000070:11, 428572, 8219000068:11, 333333, 45327000070:7, 9, 8291000070:8, 4, 8278000019:8, 4, 9.861569, 881400000:8, 124854
34 [3250000021:10, 142858, 874500005:10, 83871, 4319900028:9, 941176, 8505000099:9, 5.45455, 8505000028:9, 3.36843, 8747000093:9, 16107
35 [895000009:2, 8704379, 11439000065:2, 855758, 8270000069:2, 8531468, 8279000069:2, 8336163, 8288000098:2, 656051, 8053000098:2, 39
36 [8250000084:9, 834483, 8236000032:8, 825974, 1219600028:7, 9967103, 82680000:7, 8561153, 12197000091:7, 3.333333, 12196000016:7, 3
37 [8786000064:10, 2125, 468000046:9, 820755, 8253000064:9, 876525, 8253000097:7, 670886, 8745000064:7, 651515, 8745000025:6, 7606635,
38 [11938000065:5, 0.625, 2602200025:4, 9.411764, 2602200073:4, 75, 11361000091:4, 5, 8287000049:4, 4.428571, 8267000072:4, 2.222223, 2602
39 [11353000050:1, 0, 8268000050:1, 0, 11353000047:1, 0, 8268000048:1, 0, 11353000041:1, 0, 11353000040:1, 0, 1135300004:1, 0, 1135300004
40 [12826000077:25, 951456, 12826000040:24, 373627, 12603000023:23, 222221, 12826000039:23, 173077, 12267000023:22, 563158, 12826000
41 [8230000095:7, 851456, 7635000056:7, 8018864, 7635000070:7, 8725925, 8267000087:7, 869565, 8267000068:7, 0.69869,
42 [11439000087:2, 0, 25423000072:2, 0, 25423000028:2, 0, 25423000048:2, 0, 25423000051:2, 0, 25423000042:2, 0, 25423000015:2, 0, 254230000

```

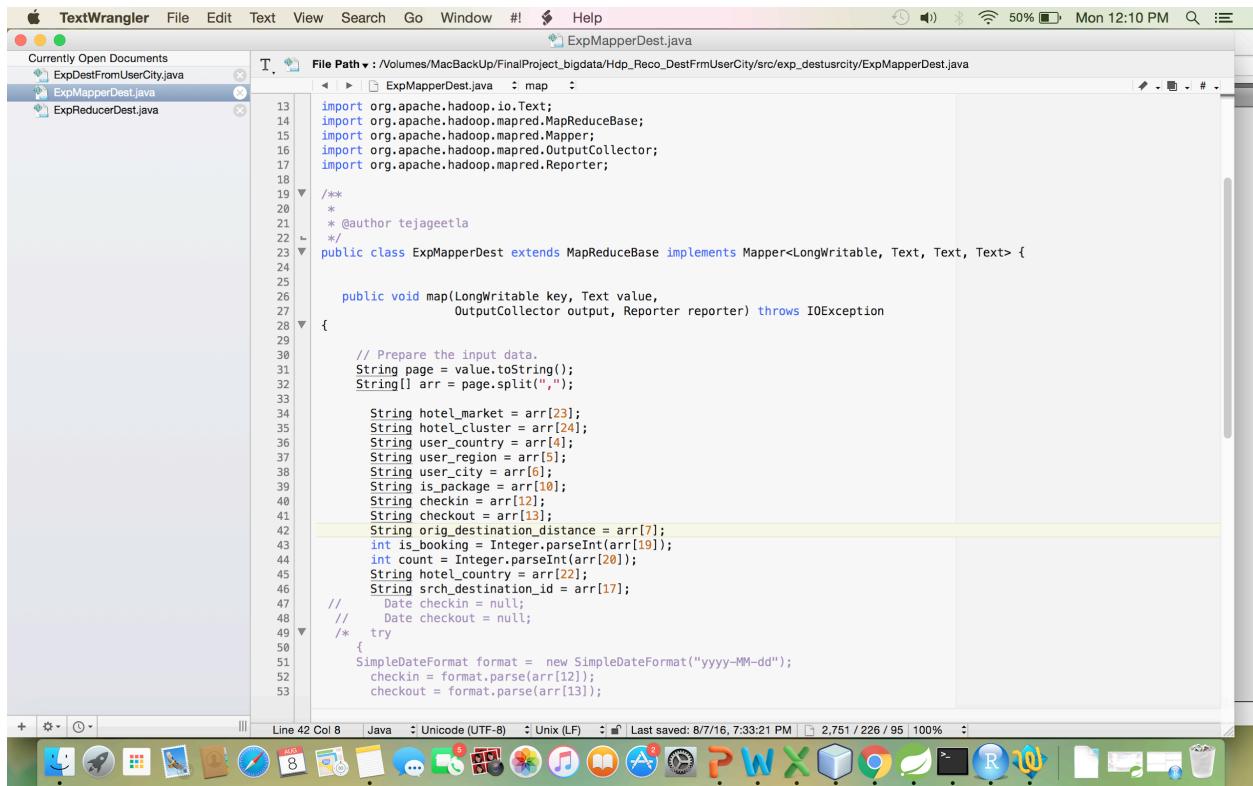
Use case 2: New Customers travel package recommendation

In this use case, the hotel recommendations are provided when the user information is not available or the user is not logged in yet. Firstly, the field user_country, user_region, user_city is mapped with the destination id and length of stay to determine which hotel IDs are booked for this combinations. The length of stay of users is determined using fields, check-in date and check out date. Further, the total number of hotel bookings is calculated separately for package and no package. The hotels are sorted based on the number of bookings.

When a user is logged into the website, inputs such as, user_city, destination id and length of stay are determined and hotel recommendations are provided based on similar previous bookings.

ExMapperDest.java

In this class we extract the user country, region, city, and destination id. Combine it by adding underscore and emit as key. The value emitted has details such as hotel_market, orig_destination_distance, is_booking, count, is_package, checkin, checkout, hotel_cluster.



The screenshot shows the TextWrangler application window on a Mac OS X desktop. The title bar reads "TextWrangler" and the file path is "/Volumes/MacBackUp/FinalProject_bigdata/Hdp_Reco_DestFrmUserCity/src/exp_destusrcty/ExpMapperDest.java". The code editor displays Java code for a MapReduce base class:

```
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.MapReduceBase;
import org.apache.hadoop.mapred.Mapper;
import org.apache.hadoop.mapred.OutputCollector;
import org.apache.hadoop.mapred.Reporter;

public class ExpMapperDest extends MapReduceBase implements Mapper<LongWritable, Text, Text, Text> {

    public void map(LongWritable key, Text value,
                    OutputCollector output, Reporter reporter) throws IOException {
        String page = value.toString();
        String[] arr = page.split("\n");

        String hotel_market = arr[23];
        String hotel_cluster = arr[24];
        String user_country = arr[4];
        String user_region = arr[5];
        String user_city = arr[6];
        String is_package = arr[10];
        String checkin = arr[12];
        String checkout = arr[13];
        String orig_destination_distance = arr[7];
        int is_booking = Integer.parseInt(arr[19]);
        int count = Integer.parseInt(arr[20]);
        String hotel_country = arr[22];
        String srch_destination_id = arr[17];
        Date checkin = null;
        Date checkout = null;
        try {
            SimpleDateFormat format = new SimpleDateFormat("yyyy-MM-dd");
            checkin = format.parse(arr[12]);
            checkout = format.parse(arr[13]);
        }
    }
}
```

The code uses imports for Text, MapReduceBase, Mapper, OutputCollector, and Reporter from the Apache Hadoop library. It defines a class ExpMapperDest that extends MapReduceBase and implements the Mapper interface. The map method processes a LongWritable key and a Text value, splitting the value into an array of strings. It then extracts specific fields like hotel_market, hotel_cluster, user_country, user_region, user_city, is_package, checkin, checkout, orig_destination_distance, is_booking, count, hotel_country, and srch_destination_id. It also attempts to parse the checkin and checkout dates using a SimpleDateFormat. The code editor shows syntax highlighting for Java keywords and comments, and the status bar at the bottom indicates the line (Line 42), column (Col 8), encoding (Unicode (UTF-8)), and file statistics (2,751 / 226 / 95).

TextWrangler File Edit Text Search Go Window #! Help

File Path : /Volumes/MacBackUp/FinalProject_bigdata/Hdp_Reco_DestFrmUserCity/src/exp_destusrcty/ExpMapperDest.java

```

55
56
57
58
59     int days =0;
60     try
61     {
62         if(checkout!=null && checkin!=null)
63         {
64             days = (int)((checkout.getTime() - checkin.getTime()) / (1000 * 60 * 60 * 24));
65         }
66     catch(ParseException pe)
67     {
68
69     } */
70
71     StringBuilder builder = new StringBuilder();
72     builder.append(orig_destination_distance);
73     builder.append(" ");
74     builder.append(hotel_market);
75     builder.append(" ");
76     builder.append(isBooking);
77     builder.append(" ");
78     builder.append(count);
79     builder.append(" ");
80     builder.append(is_package);
81     builder.append(" ");
82     builder.append(checkin);
83     builder.append(" ");
84     builder.append(checkout);
85     builder.append(" ");
86     builder.append(hotel_cluster);
87
88     String emitkey = user_country+ "_" + user_region + "_" + user_city +"_" + srch_destination_id ;
89
90     output.collect(new Text(emitkey), new Text(builder.toString()));
91
92 }
93
94
95
96
97
98
99
99

```

Line 42 Col 8 Java Unicode (UTF-8) Unix (LF) Last saved: 8/7/16, 7:33:21 PM 2,751 / 226 / 95 100% P W X R Q

ExpReducerDest.java

TextWrangler File Edit Text Search Go Window #! Help

File Path : /Volumes/MacBackUp/FinalProject_bigdata/Hdp_Reco_DestFrmUserCity/src/exp_destusrcty/ExpReducerDest.java

```

30 /**
31  * public class ExpReducerDest extends MapReduceBase implements Reducer<Text, Text, Text>
32 {
33
34     public void reduce(Text key, Iterator<Text> values,
35                         OutputCollector<Text, Text> output, Reporter reporter) throws IOException {
36
37         reporter.setStatus(key.toString());
38
39         String newKey ;
40         // double count = 0;
41         String hotel_clusterId = "";
42         HashMap<String, Integer> hotelmap = new HashMap<String, Integer>();
43         HashMap<String, Integer> hotelmapkg = new HashMap<String, Integer>();
44         double total=0;
45         StringBuilder sb = new StringBuilder();
46         StringBuilder sbpkg = new StringBuilder();
47         while(values.hasNext())
48     {
49             StringBuilder toWrite = new StringBuilder();
50             String page = ((Text)values.next()).toString();
51             String[] valTokens = page.split(" ");
52             int bkcnt = Integer.valueOf(valTokens[2]);
53             String distance = valTokens[3];
54             String hotel_mrk = valTokens[4];
55             String isBooking = valTokens[5];
56             int count = Integer.parseInt(valTokens[6]);
57             String destId = valTokens[7];
58             String isPckg = valTokens[8];
59             String checkin = valTokens[9];
60             String checkout = valTokens[10];
61             hotel_clusterId = valTokens[11];
62             int days =0;
63             try
64             {
65                 SimpleDateFormat format = new SimpleDateFormat("yyyy-MM-dd");
66                 Date d1 = format.parse(checkin);
67                 Date d2 = format.parse(checkout);
68                 if(checkout==null && checkin!=null)
69             {
70                 days = (int)((d2.getTime() - d1.getTime()) / (1000 * 60 * 60 * 24));
71
72                 hotelmap.put(destId, days);
73                 hotelmap.put(isPckg, count);
74                 hotelmap.put(isBooking, count);
75                 hotelmap.put(distance, count);
76                 hotelmap.put(hotel_mrk, count);
77                 hotelmap.put(bkcnt, count);
78                 hotelmap.put(checkin, count);
79                 hotelmap.put(checkout, count);
80                 hotelmap.put(hotel_clusterId, count);
81                 hotelmap.put(isBooking, count);
82                 hotelmap.put(isPckg, count);
83                 hotelmap.put(distance, count);
84                 hotelmap.put(hotel_mrk, count);
85                 hotelmap.put(bkcnt, count);
86                 hotelmap.put(checkin, count);
87                 hotelmap.put(checkout, count);
88                 hotelmap.put(hotel_clusterId, count);
89
90                 output.collect(new Text(newKey), new Text(hotelmap.toString()));
91
92             }
93
94         }
95
96
97
98
99
99

```

Line 31 Col 14 Java Unicode (UTF-8) Unix (LF) Last saved: 8/7/16, 9:07:39 PM 14 / 1 / 0 100% P W X R Q

TextWrangler File Edit Text Search Go Window #! ⚡ Help

File Path : /Volumes/MacBackUp/FinalProject_bigdata/Hdp_Reco_DestFrmUserCity/src/exp_destusrCity/ExpReducerDest.java

```
68
69     if(checkout!=null && checkin!=null)
70     {
71         days = (int)((d2.getTime() - d1.getTime()) / (1000 * 60 * 60 * 24));
72     }
73     catch(ParseException pe)
74     {
75     }
76     catch(NullPointerException npe)
77     {
78     }
79
80
81     String newHtlClstrId = hotel_clusterId + ":" + days;
82
83     // newKey = key.toString() + " " + hotel_clusterId;
84     // toWrite.append(val);
85
86     // System.out.println("val"+newKey);
87
88     if(isBooking.equalsIgnoreCase("1") && isPckg.equalsIgnoreCase("0"))
89     {
90         boolean check = checkIfAlreadyExist(hotelmap,newHtlClstrId);
91         if(check)
92         {
93             int temp=hotelmap.get(newHtlClstrId);
94             temp+= count;
95             hotelmap.put(newHtlClstrId, temp);
96         }
97         else
98         {
99             hotelmap.put(newHtlClstrId, count);
100        }
101
102
103
104     if(isBooking.equalsIgnoreCase("1") && isPckg.equalsIgnoreCase("1"))
105     {
106         boolean check = checkIfAlreadyExist(hotelmappkg,newHtlClstrId);
107         if(check)
```

TextWrangler File Edit Text Search Go Window #! ⚡ Help

File Path : /Volumes/MacBackUp/FinalProject_bigdata/Hdp_Reco_DestFrmUserCity/src/exp_destusrCity/ExpReducerDest.java

```
110     int temp=hotelmappkg.get(newHtlClstrId);
111     temp+= count;
112     hotelmappkg.put(newHtlClstrId, temp);
113
114
115
116
117
118
119
120
121
122
123     Set<Map.Entry<String, Integer>> hotelIdSet = hotelmap.entrySet();
124
125     ArrayList<Map.Entry<String, Integer>> hotelClusterlist = new ArrayList<Map.Entry<String, Integer>>(hotelIdSet);
126     Collections.sort( hotelClusterlist, new Comparator<Map.Entry<String, Integer>>()
127     {
128         public int compare( Map.Entry<String, Integer> o1, Map.Entry<String, Integer> o2 )
129         {
130             return (o2.getValue()).compareTo( o1.getValue() );
131         }
132     });
133     for(Map.Entry<String, Integer> entry:hotelClusterlist){
134         sb.append(entry.getKey());
135         sb.append(" ");
136         sb.append(entry.getValue());
137         sb.append(",");
138     }
139
140     Set<Map.Entry<String, Integer>> hotelIdSetPkg = hotelmappkg.entrySet();
141
142     ArrayList<Map.Entry<String, Integer>> hotelClusterlistpkg = new ArrayList<Map.Entry<String, Integer>>(hotelIdSetPkg);
143     Collections.sort( hotelClusterlistpkg, new Comparator<Map.Entry<String, Integer>>()
144     {
145         public int compare( Map.Entry<String, Integer> o1, Map.Entry<String, Integer> o2 )
146         {
147             return (o2.getValue()).compareTo( o1.getValue() );
148         }
149     });
150     for(Map.Entry<String, Integer> entry:hotelClusterlistpkg){
```

A screenshot of the TextWrangler application window on a Mac OS X desktop. The title bar reads "TextWrangler" and the active file is "ExpReducerDest.java". The code editor displays Java code for a reducer class. The code includes logic for comparing values, appending entries to a StringBuilder, and emitting key-value pairs. It also contains comments and some commented-out code. The status bar at the bottom shows "Line 31 Col 14 Java Unicode (UTF-8) Unix (LF) Last saved: 8/7/16, 9:07:39 PM 14 / 1 / 0 100%".

```
146     {
147         if( o2.getValue().compareTo( o1.getValue() ) < 0 )
148             return true;
149     }
150     for(Map.Entry<String, Integer> entry:hotelClusterlistpkg){
151         sbpkg.append(entry.getKey());
152         sbpkg.append(" ");
153         sbpkg.append(entry.getValue());
154         sbpkg.append(",");
155     }
156     if(sbpkg.toString() !=null && !sbpkg.toString().isEmpty())
157     {
158         String emitkey = "package_" +key.toString();
159         output.collect(new Text(emitkey), new Text(sbpkg.toString()));
160     }
161     if(sb.toString() !=null && !sb.toString().isEmpty())
162     {
163         output.collect(key, new Text(sb.toString()));
164     }
165 //     System.out.println("toWrite before val" + toWrite.toString());
166 //     System.out.println("REducere >>> value " + toWrite );
167 //     output.collect(key, new Text(toWrite.toString()));
168 }
169
170 public boolean checkIfAlreadyExist(HashMap<String, Integer> htlMap, String htId)
171 {
172     if(htlMap.containsKey(htId))
173     {
174         return true;
175     }
176     else
177     {
178         return false;
179     }
180 }
181
182 }
```

ExpDestFromUserCity.java

In this class we execute the job to generate results

A screenshot of the TextWrangler application window on a Mac OS X desktop. The title bar reads "TextWrangler" and the active file is "ExpDestFromUserCity.java". The code editor displays Java code for a main class. The code imports various Hadoop classes and defines a main method that sets up a JobClient, configures a JobConf, and runs the job. The status bar at the bottom shows "Line 19 Col 14 Java Unicode (UTF-8) Unix (LF) Last saved: 8/7/16, 9:08:15 PM 19 / 1 / 0 100%".

```
11 /**
12 * @author tejageetla
13 */
14 public class ExpDestFromUserCity {
15
16     /**
17      * @author tejageetla
18     */
19     public static void main(String[] args) throws Exception {
20         String inputPath = "/Volumes/MacBackUp/trainfullData2.txt";
21         String outputPath = "/Volumes/MacBackUp/expedia/Dest_Lngth_UsrCityReco/";
22
23         JobClient client = new JobClient();
24         JobConf conf = new JobConf(ExpDestFromUserCity.class);
25         conf.setJobName("ExpediaDestHotelRec");
26         conf.setMapperClass(ExpMapperDest.class);
27         conf.setMapOutputKeyClass(Text.class);
28         conf.setMapOutputValueClass(IntWritable.class);
29         conf.setReducerClass(ExpReducerDest.class);
30         conf.setInputPaths(conf, new Path(inputPath));
31         conf.setOutputPath(conf, new Path(outputPath));
32
33         client.setConf(conf);
34         try {
35             client.runJob(conf);
36         } catch (Exception e) {
37             e.printStackTrace();
38         }
39     }
40 }
```

Terminal Shell Edit View Window Help

RStudio

hadoop-0.20.2 - bash - 175x47

```

16/08/07 19:36:39 INFO mapred.LocalJobRunner: 66_348_47997_8254 > reduce
16/08/07 19:36:42 INFO mapred.LocalJobRunner: 66_348_53377_8791 > reduce
16/08/07 19:36:43 INFO mapred.JobClient: map 100% reduce 91%
16/08/07 19:36:45 INFO mapred.LocalJobRunner: 66_356_22202_8857 > reduce
16/08/07 19:36:46 INFO mapred.JobClient: map 100% reduce 92%
16/08/07 19:36:48 INFO mapred.LocalJobRunner: 66_403_14752_17859 > reduce
16/08/07 19:36:49 INFO mapred.JobClient: map 100% reduce 93%
16/08/07 19:36:51 INFO mapred.LocalJobRunner: 66_402_28624_19991 > reduce
16/08/07 19:36:54 INFO mapred.LocalJobRunner: 66_442_12589_12589 > reduce
16/08/07 19:36:55 INFO mapred.JobClient: map 100% reduce 94%
16/08/07 19:36:57 INFO mapred.LocalJobRunner: 66_440_52070_0354 > reduce
16/08/07 19:36:58 INFO mapred.JobClient: map 100% reduce 95%
16/08/07 19:37:00 INFO mapred.LocalJobRunner: 66_462_49283_12603 > reduce
16/08/07 19:37:01 INFO mapred.JobClient: map 100% reduce 96%
16/08/07 19:37:03 INFO mapred.LocalJobRunner: 69_1085_48189_22402 > reduce
16/08/07 19:37:06 INFO mapred.LocalJobRunner: 69_696_46027_46301 > reduce
16/08/07 19:37:07 INFO mapred.JobClient: map 100% reduce 97%
16/08/07 19:37:08 INFO mapred.LocalJobRunner: 69_923_27034_2930 > reduce
16/08/07 19:37:10 INFO mapred.JobClient: map 100% reduce 98%
16/08/07 19:37:12 INFO mapred.LocalJobRunner: 77_824_15015_34958 > reduce
16/08/07 19:37:13 INFO mapred.JobClient: map 100% reduce 99%
16/08/07 19:37:14 INFO mapred.LocalJobRunner: 80_41_22664_42222 > reduce
16/08/07 19:37:16 INFO mapred.TaskRunner: Task@attempt_local_0001_r_000000_0 is done. And is in the process of committing
16/08/07 19:37:16 INFO mapred.TaskRunner: Task attempt_local_0001_r_000000_0 is allowed to commit now
16/08/07 19:37:16 INFO mapred.FileOutputCommitter: Saved output of task 'attempt_local_0001_r_000000_0' to file:/Volumes/MacBackupUp/expedia/Dest_lngth_UsrCityReco
16/08/07 19:37:17 INFO mapred.LocalJobRunner: 9_29_3697_8832 > reduce
Ta 16/08/07 19:37:17 INFO mapred.TaskRunner: Task 'attempt_local_0001_r_000000_0' done.
16/08/07 19:37:17 INFO mapred.JobClient: map 100% reduce 100%
16/08/07 19:37:17 INFO mapred.JobClient: Job complete: job_local_0001
16/08/07 19:37:17 INFO mapred.JobClient: Counters: 13
16/08/07 19:37:17 INFO mapred.JobClient: Map-Reduce Framework
16/08/07 19:37:17 INFO mapred.JobClient: Combine output records=0
16/08/07 19:37:17 INFO mapred.JobClient: Spilled Records=160653298
16/08/07 19:37:17 INFO mapred.JobClient: Reduce input records=37670293
16/08/07 19:37:17 INFO mapred.JobClient: Reduce output records=1428592
16/08/07 19:37:17 INFO mapred.JobClient: Map input records=37670293
16/08/07 19:37:17 INFO mapred.JobClient: Map output records=37670293
16/08/07 19:37:17 INFO mapred.JobClient: Map output bytes=247014018
16/08/07 19:37:17 INFO mapred.JobClient: Reduce shuffle bytes=0
16/08/07 19:37:17 INFO mapred.JobClient: Combine input records=0
16/08/07 19:37:17 INFO mapred.JobClient: Map input bytes=4425411958
16/08/07 19:37:17 INFO mapred.JobClient: Reduce input groups=3810666
16/08/07 19:37:17 INFO mapred.JobClient: FileSystemsCounters
16/08/07 19:37:17 INFO mapred.JobClient: FILE_BYTES_WRITTEN=347810114329
16/08/07 19:37:17 INFO mapred.JobClient: FILE_BYTES_READ=477025361971
Tejswaroops-MacBook-Pro:hadoop-0.20.2 tejageeta@

```

Final Output:

TextWrangler File Edit Text View Search Go Window #! ⌘ Help

part-00000

File Path : /Volumes/MacBackupUp/expedia/Dest_lngth_UsrCityReco/part-00000

```

1 0_10_26750_1763 36:2 1,
2 package _0_10_26750_6305 2:5 1,
3 0_147_10568_104 36:1 1,82:2 1,
4 0_147_10568_11075 11:6 1,11:5 1,
5 0_147_10568_11357 51:13 1,51:14 1,16:4 1,51:20 1,51:21 1,
6 0_147_10568_1140 11:1 1,
7 0_147_10568_11591 48:13 1,
8 0_147_10568_11683 21:1 1,
9 0_147_10568_11815 78:5 4,46:5 1,81:3 1,
10 0_147_10568_11816 22:2 1,
11 0_147_10568_11835 77:2 1,
12 0_147_10568_11843 38:2 1,
13 0_147_10568_11850 98:1 1,
14 0_147_10568_11928 18:4 1,
15 package _0_147_10568_11980 33:1 1,
16 0_147_10568_12008 16:1 1,76:2 1,
17 0_147_10568_12153 42:2 1,
18 0_147_10568_12186 42:1 1,13:1 1,
19 0_147_10568_12189 47:1 1,59:1 1,19:1 1,18:1 1,6:1 1,
20 0_147_10568_12193 10:1 1,
21 0_147_10568_12206 11:1 1,
22 0_147_10568_12206 1:1 1,
23 0_147_10568_12209 33:2 2,83:1 1,
24 0_147_10568_12225 59:1 1,
25 0_147_10568_12233 70:1 1,
26 0_147_10568_12235 91:1 1,
27 0_147_10568_12236 77:2 1,48:2 1,50:2 1,
28 0_147_10568_12236 81:1 1,
29 0_147_10568_12254 81:6 1,
30 0_147_10568_12267 47:1 1,
31 0_147_10568_12304 42:1 1,
32 0_147_10568_12320 48:1 1,13:1 1,
33 0_147_10568_12356 46:3 1,
34 0_147_10568_12359 20:2 1,38:8 1,20:5 1,
35 0_147_10568_12359 36:1 2,81:1 1,
36 0_147_10568_12379 59:1 1,
37 0_147_10568_12432 47:2 1,
38 0_147_10568_12448 76:1 2,
39 0_147_10568_12451 57:1 1,57:4 1,60:1 1,
40 0_147_10568_12453 95:1 1,16:1 1,13:1 1,42:3 1,
41 0_147_10568_12462 57:1 1,

```

TextWrangler File Edit Text Search Go Window #! ⚡ Help

part-00000

Currently Open Documents

- ExpDestFromUserCity.java
- ExpMapperDest.java
- ExpReducerDest.java
- FinalProject_ReadMe.txt
- part-00000

File Path : /Volumes/MacBackup/expedia/Dest_Ingth_UsrCityReco/part-00000

```

383156 48_424_5957_13211 21:4 1,
383157 48_424_5957_13316 28:4 1,
383158 48_424_5957_13395 56:2 1,
383159 48_424_5957_13409 56:2 1,70:2 1,
383160 48_424_5957_13441 69:8 1,
383161 48_424_5957_13558 46:1 1,64:1 1,
383162 48_424_5957_13605 21:2 1,
383163 48_424_5957_13666 67:1 1,
383164 48_424_5957_1375 85:1 1,
383165 48_424_5957_13887 82:1 1,
383166 package_48_424_5957_13925 58:7 1,
383167 48_424_5957_13925 44:3 1,53 1,83:3 1,
383168 48_424_5957_13928 29:1 1,
383169 48_424_5957_14 38:7 1,
383170 48_424_5957_14038 10:2 1,22:1 1,8:3 1,78:1 1,78:3 1,61:2 1,62:2 1,22:3 1,10:3 1,
383171 48_424_5957_14041 58:1 1,37:1 1,28:1 1,
383172 48_424_5957_1407 38:5 2,
383173 package_48_424_5957_14127 71:5 1,
383174 48_424_5957_14139 9:4 1,9:7 1,41:3 1,
383175 48_424_5957_14207 25:4 1,
383176 48_424_5957_14384 56:3 1,21:3 1,
383177 48_424_5957_1476 78:1 1,
383178 48_424_5957_14782 93:5 1,
383179 48_424_5957_14796 26:3 1,33:2 1,
383180 48_424_5957_14797 48:1 1,
383181 48_424_5957_148 59:1 1,
383182 package_48_424_5957_14904 29:4 1,
383183 48_424_5957_14808 94:6 1,
383184 48_424_5957_14809 9:1 2,59:1 1,
383185 48_424_5957_14818 28:2 1,
383186 48_424_5957_14845 48:2 1,
383187 48_424_5957_14846 68:1 1,18:2 1,
383188 48_424_5957_14875 9:1 1,55:2 1,21:2 1,56:1 1,55:3 1,95:1 1,21:3 1,
383189 48_424_5957_14882 82:1 1,
383190 48_424_5957_14896 62:2 1,
383191 48_424_5957_1493 39:2 1,
383192 48_424_5957_14985 17:3 1,91:5 1,86:3 1,4:3 1,41:2 1,5:3 1,
383193 48_424_5957_15008 78:1 1,
383194 48_424_5957_15093 19:1 1,
383195 48_424_5957_1517 62:1 1,
383196 48_424_5957_15193 99:1 1,

```

Line 1 Col 1 (none) Unicode (UTF-8) Unix (LF) Last saved: 8/7/16, 7:37:16 PM 46,723,642 / 1,428,593 100%

TextWrangler File Edit Text Search Go Window #! ⚡ Help

part-00000

Currently Open Documents

- ExpDestFromUserCity.java
- ExpMapperDest.java
- ExpReducerDest.java
- FinalProject_ReadMe.txt
- part-00000

File Path : /Volumes/MacBackup/expedia/Dest_Ingth_UsrCityReco/part-00000

```

1428553 package_99_766_43752_8287 95:2 1,
1428554 99_766_43752_8351 25:1 2,75:7 1,25:5 1,
1428555 99_766_43752_8601 44:7 1,
1428556 99_766_43752_8625 21:2 1,
1428557 package_99_766_43752_8741 2:3 1,2:4 1,
1428558 99_766_43752_8745 11:7 1,
1428559 99_766_43752_8745 64:5 1,41:5 1,68:4 1,48:4 1,29:1 1,76:11 1,8:13 1,
1428560 99_766_43752_8746 22:3 2,37:1 1,9:4 1,78:21 1,60:3 1,
1428561 99_766_43752_878 20:1 2,20:3 1,
1428562 99_766_43752_8784 22:2 1,
1428563 99_766_43752_8785 86:3 1,
1428564 99_766_43752_8788 64:4 1,95:10 1,8:5 1,59:2 1,
1428565 package_99_766_43752_8791 65:5 1,65:7 1,
1428566 99_766_43752_8794 11:5 2,29:3 1,
1428567 package_99_766_43752_8797 29:4 1,
1428568 99_766_43752_8798 77:2 1,
1428569 99_766_43752_8799 97:1 1,
1428570 99_766_43752_8803 36:11 1,
1428571 99_766_43752_8804 77:14 1,
1428572 99_766_43752_8806 9:2 1,8:3 1,
1428573 99_766_43752_8815 82:1 2,61:16 1,30:14 1,25:8 1,25:7 1,25:1 1,64:2 1,
1428574 99_766_43752_8817 11:2 1,
1428575 99_766_43752_8818 61:1 1,
1428576 package_99_766_43752_8820 6:8 1,
1428577 package_99_766_43752_8823 8:2 1,
1428578 99_766_43752_8855 26:7 1,
1428579 99_766_43752_8864 65:6 1,
1428580 99_766_43752_972 38:1 1,
1428581 9_29_3697_11816 38:3 1,64:3 1,
1428582 9_29_3697_12851 66:1 1,64:2 1,
1428583 9_29_3697_12206 45:1 1,
1428584 9_29_3697_21266 82:2 1,
1428585 9_29_3697_22616 21:1 1,
1428586 9_29_3697_24392 6:1 1,
1428587 9_29_3697_24897 75:3 1,
1428588 9_29_3697_8213 16:1 1,
1428589 9_29_3697_8252 91:3 1,
1428590 9_29_3697_8261 16:2 1,
1428591 package_9_29_3697_8287 51:18 1,
1428592 9_29_3697_8832 10:1 1,
1428593

```

Line 1 Col 1 (none) Unicode (UTF-8) Unix (LF) Last saved: 8/7/16, 7:37:16 PM 46,723,642 / 1,428,593 100%

Apple TextWrangler File Edit Text View Search Go Window #! ⌘ Help part-00000

Currently Open Documents

- ExpDestFromUserCity.java
- ExpMapperDest.java
- ExpReducerDest.java
- FinalProject_ReadMe.txt
- part-00000

File Path : /Volumes/MacBackUp/expedia/Dest_Ingh_UsrCityReco/part-00000

```
part-00000
1428409 99_766_43752_50177 82:2 1,
1428500 99_766_43752_50997 85:1 1,
1428501 99_766_43752_52883 19:2 1,
1428502 99_766_43752_531 82:3 1,
1428503 package_99_766_43752_53533 70:5 1,
1428504 package_99_766_43752_53681 43:7 1,
1428505 99_766_43752_53681 29:6 1,29:11 1,
1428506 99_766_43752_53681 38:4 1,
1428507 99_766_43752_53681 36:2 1,
1428508 99_766_43752_53681 93:2 1,
1428509 99_766_43752_53924 85:3 2,
1428510 99_766_43752_57122 20:3 1,
1428511 99_766_43752_597 2:1 1,
1428512 99_766_43752_6051 21:2 1,
1428513 99_766_43752_60514 84:4 1,
1428514 99_766_43752_60718 46:2 1,
1428515 99_766_43752_61097 28:2 1,28:1 1,
1428516 99_766_43752_7509 41:2 1,
1428517 99_766_43752_7635 56:1 1,70:1 1,95:2 1,
1428518 package_99_766_43752_779 64:1 1,
1428519 99_766_43752_779 59:1 3,46:3 2,75:1 2,64:1 2,20:2 1,46:2 1,82:9 1,59:2 1,46:5 1,46:6 1,29:2 1,29:1 1,81:1 1,43:1 1,8:1 1,
1428520 99_766_43752_8090 82:1 4,44:3 2,46:2 2,82:2 2,82:3 2,64:3 2,46:3 1,59:2 1,44:1 1,
1428521 99_766_43752_8213 9:3 1,
1428522 package_99_766_43752_8220 59:5 1,68:6 1,68:8 1,
1428523 99_766_43752_8220 58:1 1,46:3 1,58:3 1,68:9 1,64:3 1,
1428524 99_766_43752_8222 82:2 1,
1428525 99_766_43752_8224 82:2 1,
1428526 99_766_43752_8235 48:2 1,
1428527 package_99_766_43752_8241 36:14 1,
1428528 package_99_766_43752_8243 72:4 1,
1428529 99_766_43752_8243 49:13 2,19:2 2,49:1 1,59:3 1,72:2 1,6:2 1,
1428530 99_766_43752_8245 46:4 1,5:1 1,
1428531 package_99_766_43752_8250 1:4 1,24:4 1,1:5 1,
1428532 99_766_43752_8250 45:3 1,
1428533 package_99_766_43752_8252 47:26 1,46:12 1,
1428534 99_766_43752_8252 46:2 4,46:3 4,46:1 2,2:2 2,58:1 1,48:1 1,46:4 1,46:6 1,58:5 1,57:7 1,82:2 1,2:1 1,64:21 1,46:15 1,82:18
1428535 package_99_766_43752_8253 9:6 1,
1428536 99_766_43752_8253 64:4 1,5:8 1,97:5 1,22:4 1,64:2 1,
1428537 package_99_766_43752_8254 98:6 1,
1428538 99_766_43752_8254 98:2 1,
1428539 package_99_766_43752_8255 9:2 1,
```

Line 1 Col 1 (none) Unicode (UTF-8) Unix (LF) Last saved: 8/7/16, 7:37:16 PM 46,723,642 / 1,428,593 100% +

