

Home Work - 5

CS 6350.002 - Big Data Management and Analytics

Student Name: Gowtham Uma Maheswari Jaganathan
Net Id : gxu150030

Deliverables:

- Map Reduce
 - Final.jar
- Scala
 - simple-project_2.10-1.0

Instructions:

MAP REDUCE - Execution instructions

```
hadoop jar final.jar hadoop.algorithms.joins.yelpData.Q1_Top10_Review_And_Business.Q1 UJG/OP1  
UJG/yelpdatafall/business/business.csv UJG/yelpdatafall/review/review.csv
```

```
hadoop jar final.jar hadoop.algorithms.joins.yelpData.Q2_User_Ave_Rating.User_Ave_Rating_Driver  
UJG/OP1 UJG/yelpdatafall/review/review.csv UJG/yelpdatafall/user/user.csv "Pgm J."
```

```
hadoop jar final.jar  
hadoop.algorithms.joins.yelpData.Q3_User_Rating_Stanford.User_Rating_Stanford_Driver UJG/OP1  
UJG/yelpdatafall/review/review.csv UJG/yelpdatafall/business/business.csv
```

```
hadoop jar final.jar hadoop.algorithms.joins.yelpData.Q4_Top_10_User_Review.Top_Most_Reviewers  
UJG/OP1 UJG/yelpdatafall/review/review.csv UJG/yelpdatafall/user/user.csv
```

```
hadoop jar final.jar  
hadoop.algorithms.joins.yelpData.Q5_Business_Count_TX.Top_Rated_Texas_Resurants UJG/OP1  
UJG/yelpdatafall/business/business.csv UJG/yelpdatafall/review/review.csv
```

Note: The paths to the files are to be located in HDFS.

SCALA - Execution instructions

```
spark-submit --master local --class org.apache.spark.examples.streaming.Q1 simple-project_2.10-1.0.jar /home/hduser/UJG/Assignment_5/yelpData/review.csv  
/home/hduser/UJG/Assignment_5/yelpData/business.csv
```

```
spark-submit --master local --class org.apache.spark.examples.streaming.Q2 simple-project_2.10-1.0.jar "Lisa W." /home/hduser/UJG/Assignment_5/yelpData/user.csv  
/home/hduser/UJG/Assignment_5/yelpData/review.csv
```

```
spark-submit --master local --class org.apache.spark.examples.streaming.Q3 simple-project_2.10-1.0.jar /home/hduser/UJG/Assignment_5/yelpData/business.csv  
/home/hduser/UJG/Assignment_5/yelpData/review.csv
```

```
spark-submit --master local --class org.apache.spark.examples.streaming.Q4 simple-project_2.10-1.0.jar /home/hduser/UJG/Assignment_5/yelpData/user.csv  
/home/hduser/UJG/Assignment_5/yelpData/review.csv
```

```
spark-submit --master local --class org.apache.spark.examples.streaming.Q5 simple-project_2.10-1.0.jar /home/hduser/UJG/Assignment_5/yelpData/business.csv  
/home/hduser/UJG/Assignment_5/yelpData/review.csv
```

Note: The paths to the files are to be located in 6360 server and not in HDFS.