

Home Work - 6

CS 6350.002 - Big Data Management and Analytics

Student Name: Gowtham Uma Maheswari Jaganathan
Net Id : gxu150030

Deliverables:

- Pig
 - Q1_Top10
 - Q2_Name
 - Q3_User_Star
 - Q4_Most_Review
 - Q5_Standofrd
 - findTopCommonWord
- Scala
 - SimpleApp
 - simple-project_2.10-1.0

Instructions:

PIG - Execution instructions

```
pig -x local -f Q1_Top10.pig -param business='/home/hduser/UJG/Assignment_5/yelpData/business.csv'  
-param review='/home/hduser/UJG/Assignment_5/yelpData/review.csv' -param  
user='/home/hduser/UJG/Assignment_5/yelpData/user.csv'
```

```
pig -x local -f Q2_Name.pig -param business='/home/hduser/UJG/Assignment_5/yelpData/business.csv'  
-param review='/home/hduser/UJG/Assignment_5/yelpData/review.csv' -param  
user='/home/hduser/UJG/Assignment_5/yelpData/user.csv' -param id="Matt H"
```

```
pig -x local -f Q3_User_Star.pig -param  
business='/home/hduser/UJG/Assignment_5/yelpData/business.csv' -param  
review='/home/hduser/UJG/Assignment_5/yelpData/review.csv' -param  
user='/home/hduser/UJG/Assignment_5/yelpData/user.csv'
```

```
pig -x local -f Q4_Most_Review.pig -param  
business='/home/hduser/UJG/Assignment_5/yelpData/business.csv' -param  
review='/home/hduser/UJG/Assignment_5/yelpData/review.csv' -param  
user='/home/hduser/UJG/Assignment_5/yelpData/user.csv'
```

```
pig -x local -f Q5_Standofrd.pig -param  
business='/home/hduser/UJG/Assignment_5/yelpData/business.csv' -param  
review='/home/hduser/UJG/Assignment_5/yelpData/review.csv' -param  
user='/home/hduser/UJG/Assignment_5/yelpData/user.csv'
```

PART 2 :

```
pig -x local -f findTopCommonWord.pig -param len=5
```

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

SCALA - Execution instructions

```
spark-submit --master local --class org.apache.spark.examples.streaming.Q6 simple-project_2.10-  
1.0.jar /home/hduser/UJG/Assignment_6/Shakespeare /home/hduser/UJG/Assignment_6/Bible/book
```

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