

Name: Arun Kumar N

Student ID : 2017CBDE030

Spark Assignment

Spark Submit Run Commands:

1. Single Row LookUp:

```
spark-submit --class com.spark.Assignment1.SparkTask1 --master yarn --deploy-mode  
client --name testproject --conf "spark.app.id=testproject" testproject-0.0.1-SNAPSHOT-  
jar-with-dependencies.jar  
/user/root/spark_assignment/input_dataset/yellow_tripdata_*  
/user/root/spark_assignment/output/single_row_lookup_SparkRDD
```

2. Filter By

```
spark-submit --class com.spark.Assignment2.SparkTask2 --master yarn --deploy-mode  
client --name testproject2 --conf "spark.app.id=testproject2" testproject-0.0.1-  
SNAPSHOT-jar-with-dependencies.jar  
/user/root/spark_assignment/input_dataset/yellow_tripdata_*  
/user/root/spark_assignment/output/FilterBy
```

3. Group By

```
spark-submit --class com.spark.Assignment3.SparkTask3 --master yarn --deploy-mode  
client --name testproject3 --conf "spark.app.id=testproject3" testproject-0.0.1-  
SNAPSHOT-jar-with-dependencies.jar  
/user/root/spark_assignment/input_dataset/yellow_tripdata_*  
/user/root/spark_assignment/output/GroupByRDD
```