Spark

1)lookUp

cp /home/ec2-user/sparkAssignment/lookUpRDD.jar /root/spark_Assignment

(go to /root/spark_Assignment folder and run)hadoop fs -put lookUpRDD.jar /user/root/spark_assignment

```
 spark-submit --class spark.question1.lookUpRDD \

--master yarn --deploy-mode client \

--name aws_spark --conf "spark.app.id=spark_aws_run" \

lookUpRDD.jar /user/root/spark_assignment/input_dataset/yellow_tripdata*
/user/root/spark_assignment/output/single_row_lookup_SparkRDD
```

2)filteredByRateId:

cp /home/ec2-user/sparkAssignment/FilterByRateCodeIdRDD.jar /root/spark_Assignment

(go to /root/spark_Assignment folder and run)hadoop fs -put FilterByRateCodeIdRDD.jar /user/root/spark_assignment

```
spark-submit --class spark.question2.FilterByRateCodeIdRDD \

--master yarn --deploy-mode client \
```

--name aws_spark --conf "spark.app.id=spark_aws_run" \

FilterByRateCodeIdRDD.jar
/user/root/spark_assignment/input_dataset/yellow_tripdata*
/user/root/spark_assignment/output/Filter_By_RateCodeId_SparkRDD

/user/root/spark_assignment/winUtils/bin/winutils.exe

3)**groupedRecords:**

cp /home/ec2-user/sparkAssignment/groupedRecordsRDD.jar
/root/spark_Assignment

(go to /root/spark_Assignment folder and run)hadoop fs -put
groupedRecordsRDD.jar /user/root/spark_assignment

spark-submit --class spark.question3.groupedRecordsRDD \

--master yarn --deploy-mode client \

--name aws_spark --conf "spark.app.id=spark_aws_run" \

groupedRecordsRDD.jar
/user/root/spark_assignment/input_dataset/yellow_tripdata*
/user/root/spark_assignment/output/grouped_Records_SparkRDD

Pig

- cp /home/ec2-user/pigAssignment/* /root/pig_Assignment
- pig lookUp.pig
- pig FilterByRateCodeId.pig
- pig groupedRecords.pig