Readme

1. Folder **pig_scripts** contain pig scripts for all 3 cases.

2. Folder **sparkjars** contain spark jar files for all 3 cases.

3. File **output.zip** contain output files from execution of pig scripts and spark code on ec2.

4. File **sparkassignment.zip** contain eclipse project for spark java code.

5. File **Code_pig_and_spark_with_explanation.pdf** contain code for all pig scripts and spark java programs along with explanations.

6. File **Execution_logs_pig_and_spark_with_commands.pdf** contain execution logs for all pig scripts and spark java programs along with commands to execute them.

7. File **Output_of_pig_and_spark.pdf** contain output of all pig scripts and spark java programs.

8. File **Run_time_comparison_and_conclusion.pdf** contain 2 tables showing run time comparison, record output comparison and conclusion.

9. Input path for all pig scripts and spark java programs in hdfs:

   /user/ec2-user/spark_assignment/input

10. Output path for various pig scripts on hdfs:

    /user/ec2-user/spark_assignment/pig_output1
    /user/ec2-user/spark_assignment/pig_output2
    /user/ec2-user/spark_assignment/pig_output3

11. Output path for various spark java programs on hdfs:

    /user/ec2-user/spark_assignment/spark_output1
    /user/ec2-user/spark_assignment/spark_output2
    /user/ec2-user/spark_assignment/spark_output3