**STREAM DATA INGESTION AND PROCESSING WORKSHOP**

**TEAM MEMBERS**

1. JAYARAJA S.K. – [jayaraja19041@cse.ssn.edu.in](mailto:jayaraja19041@cse.ssn.edu.in)
2. PRAVINKRISHNAN K – [pravinkrishnan19082@cse.ssn.edu.in](mailto:pravinkrishnan19082@cse.ssn.edu.in)
3. SAMYUKTHA G – [samyuktha19095@cse.ssn.edu.in](mailto:samyuktha19095@cse.ssn.edu.in)

**PROGRAMMING LANGUAGES USED**

1. JAVA
2. SQL

**TECHNOLOGIES USED**

1. SPARK
2. KAFKA

**GITHUB REPOSITORY LINK FOR THE PROJECT**

<https://github.com/Jayaraja-SK/Stream-Data-Ingestion-and-Processing-Workshop>

**USE CASE**

A Car Rental company which rents their electric cars of different variants to people and runs business in more than 100 countries

**FIELDS**

1. carId
2. customerId
3. carType – Randomly chosen from the following:
4. SUV
5. Sedan
6. Hatchback
7. Cabriolet
8. Maybach
9. Van
10. Speed
11. BootLoad
12. noOfPassengers – No. of passengers travelling
13. batteryChargeLeft – Charge left in the car
14. Time – The time at which the data is recorded
15. Location – In terms of latitude and longitude

**FILES**

1. CarInfo – Contains the fields described above and their respective getter-setter methods
2. Functions – Contains the static functions to generate random data and is imported by all main functions
3. CarTracking – Main java function to generate random data and print the same to the console
4. CarTrackingToFile – Main java function to generate random data and save the same to different .*txt* files under a directory called as CarDataJson
5. CarTrackingSingleBroker – Main java function to generate random data by producer and send the same to a single consumer/broker
6. CarTrackingMultiBroker – Main java function to generate random data by producer and send the same to multiple consumer’s/broker’s
7. CarSparkStreaming – Main java function that gets data from Kafka Consumer’s and sends the same to Spark UI
8. CarSparkStreamingSQL – Main java function that gets data from Kafka Consumer’s and sends the same to Spark UI where the data is stored in a directory called CarDataSpark and could be later queried in Spark-Shell

**OUTPUT**

1. **PRINTING RANDOM DATA TO CONSOLE**



1. **WORKING OF SINGLE BROKER SYSTEM**

<https://drive.google.com/file/d/1AN-kGj_8DUqMLKAILGhD3Fkx6nM8lKeJ/view?usp=sharing>

1. **WORKING OF MULTI BROKER SYSTEM**

<https://drive.google.com/file/d/19liIcBu3SWtBtx_7U0HbwbJMFLbYGI17/view?usp=sharing>

1. **WORKING OF STREAMING DATA TO SPARK AND LATER QUERIED USING SQL**

<https://drive.google.com/file/d/1M-qdHXUsanUqodqGWkVGCZ6vAKzHA_9L/view?usp=sharing>