

**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI**

**WORK INTEGRATED LEARNING PROGRAMMES**

## Stream Processing and Analytics

**Assignment 2**

**Assignment Group Number: 27**

**Contribution Table:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sl. No.** | **Name (as appears in Canvas)** | **ID NO** | **Contribution (%)** |
| 1 | Jagadeesh Panthati | 2019hc04101 | 100% |
| 2 | G V A SATISH KUMAR VEMPATI | 2019hc04401 | 100% |
| 3 | Srutakirti Mangaraj | 2019hc04106 | 100% |

**Software Used:**

Kafka - kafka\_2.12-2.8.0

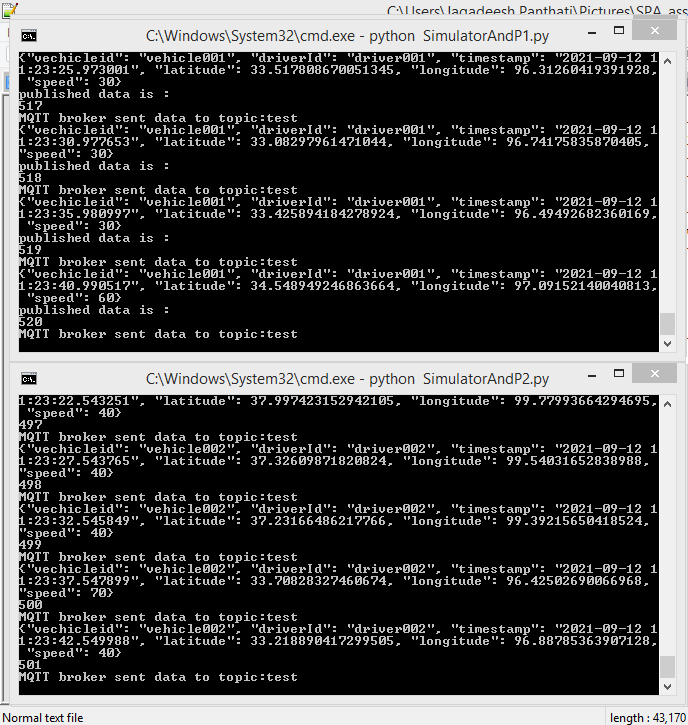
Spark - spark-2.4.8-bin-hadoop2.7

python3.7

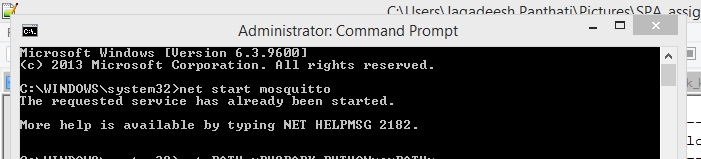
**Simulator program for the truck data movement over the period of time:**

SimulatorAndP1.py

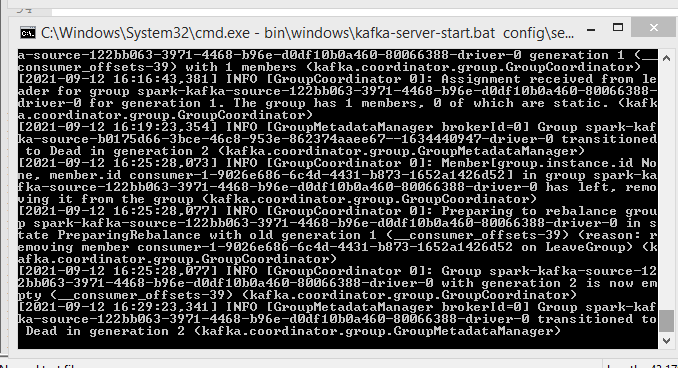
SimulatorAndP2.py

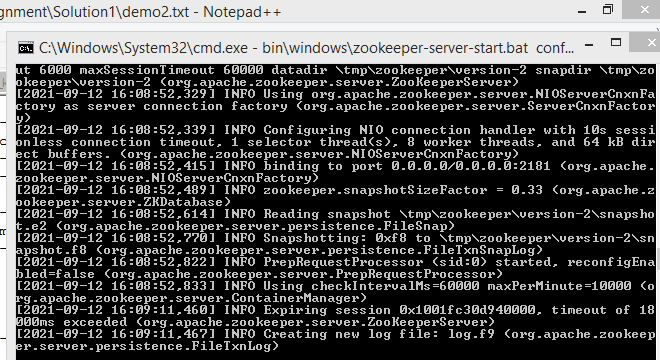


**Data Transfer program moving the data from the truck to central server like Mosquito broker through MQTT protocol:**

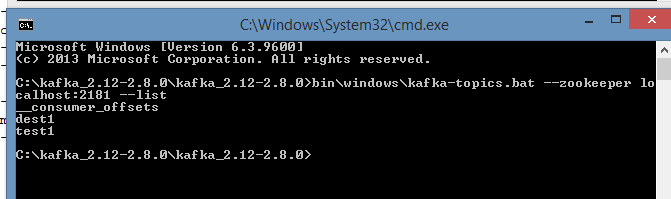
****

**Start** 🡪 **Kafka, Zookeper -- Data Flow tier:**

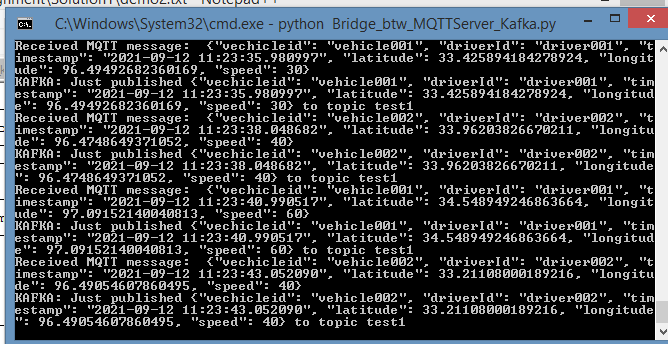
****

****

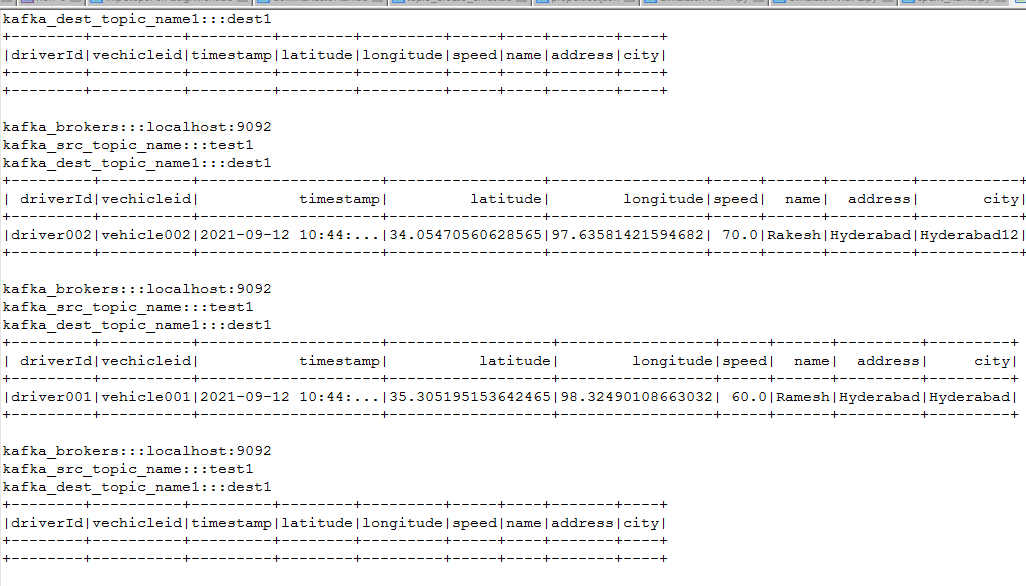
**Create Kafka Topic :**

****

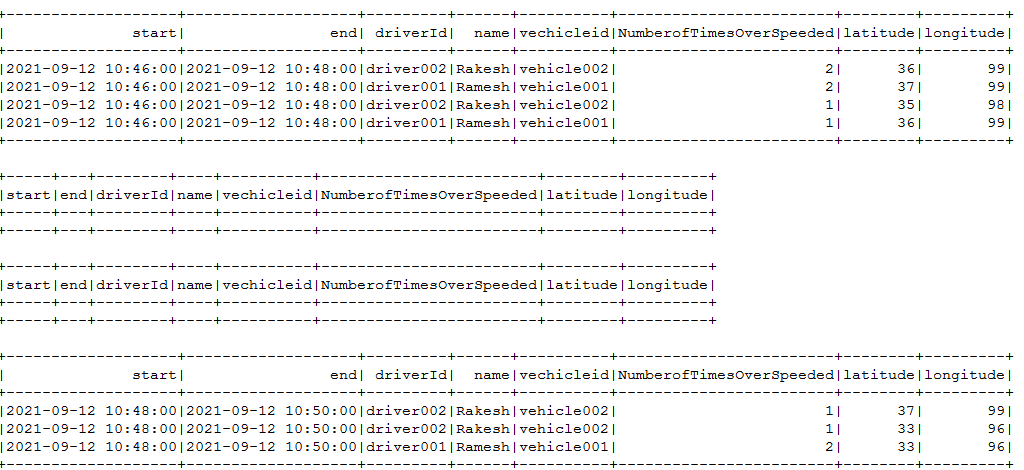
**Data transfer program from Mosquito broker to Kafka Topic :**

****

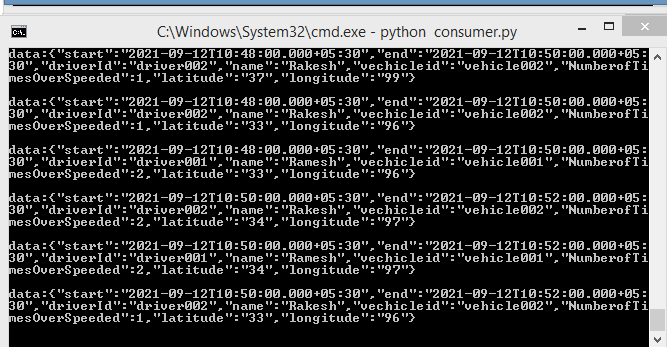
**Processing of data to identify the drivers exceeding the speed limits:**

****

**Program to keep statistics about over speeding cases over the period of time, for different routes, for different trucks:**

****

**A simple interface for showing over speeding statistics to the end consumers:**

****

**Commands to Run:**

**while running spark set the paths:**

set PATH=%PYSPARK\_PYTHON%;%PATH%

set PATH=%SPARK\_HOME%\bin;%PATH%

set PATH=%HADOOP\_HOME%\bin;%PATH%

set PATH=%JAVA\_HOME%\bin;%PATH%

set PYSPARK\_DRIVER\_PYTHON=python

set PATH=C:\apache-maven-3.8.2\bin;%PATH%

set TMP=C:/tmp99

**MQTT start [cmd with admin permissions] -- data ingestion tier :**

net start mosquitto

**Kafka, Zookeper -- Data Flow tier**

"Go to Kafka directory where kafka is installed"

bin\windows\zookeeper-server-start.bat config\zookeeper.properties

bin\windows\kafka-server-start.bat config\server.properties

**Note: create source topic and dest topic in kafka -- Data Store tier // processed data will be saved here:**

bin\windows\kafka-topics.bat --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic test1

bin\windows\kafka-topics.bat --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic dest1

To check:

bin\windows\kafka-topics.bat --zookeeper localhost:2181 --list

-- run simulator code

-- run Bridge\_btw\_MQTTServer\_Kafka

**streaming analysis tier:**

spark-submit --conf spark.jars.ivy=/C:/.ivy --packages org.apache.spark:spark-sql-kafka-0-10\_2.11:2.4.8 spark\_kafka\_q6.py >> demo1.txt

spark-submit --conf spark.jars.ivy=/C:/.ivy --packages org.apache.spark:spark-sql-kafka-0-10\_2.11:2.4.8 spark\_kafka.py >> demo2.txt

**Delivary/reporting Tier:**

run consumer.py

bin\windows\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic dest1