



Load data from Kafka to Hadoop

<Steps to run the python file to load data from Kafka>

We create a python file "spark_local_flatten.py" to load data from kafka server on hdfs path "/user/hadoop/click_stream_data" in csv format.

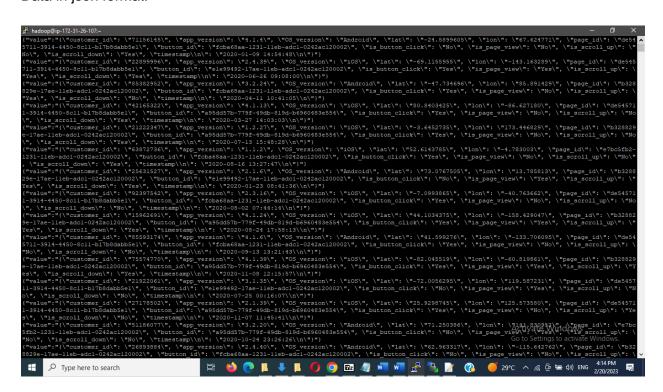
Command to execute Python file:-

export SPARK_KAFKA_VERSION=0.10
spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.11:2.4.5
spark_kafka_to_local.py
spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.11:2.4.5
spark_local_flatten.py
<Steps to load the data into Hadoop>

Check the files after importing data from kafka Hdfs dfs -ls /user/capstone/carride

<Screenshot of the data>

Data in json format:-







Data in CSV format:-

