1. Go to root user

sudo -i

2. create a file named

vi spark_kafka_to_local.py and write code into it

3. execute the .py file using following command

spark2-submit --jars spark-sql-kafka-0-10_2.11-2.3.0.jar spark_kafka_to_local.py

5. create another .py file

vi spark_local_flatten.py and write code

6. execute the .py file using following command

spark2-submit --jars spark-sql-kafka-0-10_2.11-2.3.0.jar spark_local_flatten.py

7. Use commands to transfer file from local system to hadoop

hdfs dfs -cp /home/clickstream_data_flatten /user/root/clickstream_data_flatten_new

8. see no of records and validate against validation doc

hadoop fs -cat /user/root/clickstream_data_flatten_new/part-00000-9d710548-36db-4474-a44a-cd09a42fce3e -c000.   csv | wc -l