Load data from Kafka to Hadoop

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

Step 1: You need to create two directories "checkpoint" and "clickstreamdump", first one to store the metadata of the job running and clickstreamdump is where the actual data will be dumped.
After creating the directories, you can replace the variables in the third and fourth line of the program to their respective filepath.

Step 2: Please run the below command to download the Spark-SQL-Kafka jar file. This jar will be used to run the Spark Streaming-Kafka codes. Please copy-paste the below command in your EC2 instance terminal.

"wget https://ds-spark-sql-kafka-jar.s3.amazonaws.com/spark-sql-kafka-0-10_2.11-2.3.0.jar"

Step 3: After you download the jar file you can run the python file by giving the following command to import data from kafka to hdfs.
"spark2-submit –jars spark-sql-kafka-0-10_2.11-2.3.0.jar spark_kafka_to_local.py"

<Steps to load the data into Hadoop>

The data is loaded into the hdfs in step3, now from running the "spark_local_flatten.py" script will clean the data into more structured format.

"spark2-submit spark_local_flatten.py"

<Screenshot of the data>

```
cusomter_id,app_version,OS_version,lat,lon,page_id,button_id,is_button_click,is_page_view,is_scroll_up,is_scroll_down,timestamp
26564820,3.2.35,Android,16.4454865,99.902065,de545711-3914-4450-8c11-b17b8dabb5e1,fcba68aa-1231-11eb-adc1-0242ac120002,No,Yes,No,Yes,
31906387,2.4.7,iOS,-64.813749,-133.527040,de545711-3914-4450-8c11-b17b8dabb5e1,a95dd57b-779f-49db-819d-b6960483e554,No,No,Yes,Yes,
25713677,3.4.12,Android,89.943435,127.313415,b328829e-17ae-11eb-adc1-0242ac120002,fcba68aa-1231-11eb-adc1-0242ac120002,No,No,Yes,No,
83474293,3.1.8,Android,-69.939070,-36.451670,e7bc5fb2-1231-11eb-adc1-0242ac120002,e1e99492-17ae-11eb-adc1-0242ac120002,Yes,No,Yes,No,
63727807,2.2.9,iOS,64.082108,-81.822078,e7bc5fb2-1231-11eb-adc1-0242ac120002,fcba68aa-1231-11eb-adc1-0242ac120002,No,Yes,Yes,Yes,
73737907,4.3.19,Android,-18.850508,-116.358375,b328829e-17ae-11eb-adc1-0242ac120002,e1e99492-17ae-11eb-adc1-0242ac120002,No,Yes,No,Yes,
3927433,3.2.26,iOS,-84.6857245,-146.507678,de545711-3914-4450-8c11-b17b8dabb5e1,a95dd57b-779f-49db-819d-b6960483e554,Yes,Yes,No,Yes,
12691783,3.3.11,Android,54.3852925,-37.411814,de545711-3914-4450-8c11-b17b8dabb5e1,e1e99492-17ae-11eb-adc1-0242ac120002,Yes,Yes,No,No,
22635021,4.4.36,iOS,-31.805500,150.655650,e7bc5fb2-1231-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,No,No,No,No,
23593546,1.2.16,Android,8.8918475,-83.929878,de545711-3914-4450-8c11-b17b8dabb5e1,e1e99492-17ae-11eb-adc1-0242ac120002,Yes,No,Yes,No,
16929816,1.2.25,Android,-34.8576385,-136.639835,e7bc5fb2-1231-11eb-adc1-0242ac120002,fcba68aa-1231-11eb-adc1-0242ac120002,No,No,No,Yes,
41929865,1.1.14,iOS,-34.3613155,-137.467833,b328829e-17ae-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,No,Yes,No,No,
68940320,1.1.17,Android,81.738239,-104.207763,de545711-3914-4450-8c11-b17b8dabb5e1,fcba68aa-1231-11eb-adc1-0242ac120002,Yes,No,Yes,No,
95405928,1.2.10,Android,44.185627,78.125488,e7bc5fb2-1231-11eb-adc1-0242ac120002,e1e99492-17ae-11eb-adc1-0242ac120002,No,No,No,No,
56235860,1.4.16,Android,-61.178407,8.534126,de545711-3914-4450-8c11-b17b8dabb5e1,a95dd57b-779f-49db-819d-b6960483e554,Yes,Yes,Yes,Yes,
81758845,4.4.30,Android,78.904984,167.315596,e7bc5fb2-1231-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,Yes,Yes,No,No,
46822067,1.2.26,Android,45.8993985,135.895118,de545711-3914-4450-8c11-b17b8dabb5e1,fcba68aa-1231-11eb-adc1-0242ac120002,Yes,Yes,Yes,Yes,
37623067,3.2.27,iOS,-74.1131505,-170.715421,b328829e-17ae-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,No,Yes,No,No,
14571693,3.4.21,iOS,-28.8329055,161.132369,e7bc5fb2-1231-11eb-adc1-0242ac120002,fcba68aa-1231-11eb-adc1-0242ac120002,No,Yes,No,No,
62605529,2.1.36,Android,44.196239,-15.354515,e7bc5fb2-1231-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,Yes,Yes,No,No,
83657253,2.1.2,iOS,38.7036005,-59.081929,b328829e-17ae-11eb-adc1-0242ac120002,fcba68aa-1231-11eb-adc1-0242ac120002,No,Yes,No,Yes,
47086145,3.3.16,Android,62.2950045,47.379520,de545711-3914-4450-8c11-b17b8dabb5e1,e1e99492-17ae-11eb-adc1-0242ac120002,No,No,No,Yes,
40025212,4.4.12,iOS,-87.4878555,-139.785288,b328829e-17ae-11eb-adc1-0242ac120002,fcba68aa-1231-11eb-adc1-0242ac120002,No,No,Yes,No,
19283684,2.4.23,iOS,47.447583,31.679116,e7bc5fb2-1231-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,Yes,No,No,No,
21563520,1.3.31,iOS,-24.616214,-66.741365,e7bc5fb2-1231-11eb-adc1-0242ac120002,fcba68aa-1231-11eb-adc1-0242ac120002,Yes,No,Yes,No,
35705060,2.2.13,iOS,-11.2298655,8.010661,de545711-3914-4450-8c11-b17b8dabb5e1,fcba68aa-1231-11eb-adc1-0242ac120002,No,No,Yes,No,
12252116,2.1.13,iOS,-7.7110045,-49.999098,e7bc5fb2-1231-11eb-adc1-0242ac120002,e1e99492-17ae-11eb-adc1-0242ac120002,Yes,Yes,No,Yes,
88636479,3.2.22,iOS,-46.236733,-21.683563,b328829e-17ae-11eb-adc1-0242ac120002,fcba68aa-1231-11eb-adc1-0242ac120002,No,Yes,No,Yes,
41664607,4.3.9,Android,-0.2391655,-77.952798,e7bc5fb2-1231-11eb-adc1-0242ac120002,fcba68aa-1231-11eb-adc1-0242ac120002,Yes,No,Yes,Yes,
72578040,3.3.25,Android,49.2911795,-152.594458,e7bc5fb2-1231-11eb-adc1-0242ac120002,e1e99492-17ae-11eb-adc1-0242ac120002,No,Yes,No,Yes,
17535587,4.2.13,iOS,-43.4202565,24.324168,e7bc5fb2-1231-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,No,No,Yes,No,
14734830,4.2.29,Android,-25.862119,65.408339,e7bc5fb2-1231-11eb-adc1-0242ac120002,fcba68aa-1231-11eb-adc1-0242ac120002,No,Yes,Yes,No,
68194999,2.3.4,iOS,51.4876125,73.580584,e7bc5fb2-1231-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,No,No,No,Yes,
```