Load data from Kafka to Hadoop

Steps to run the python file to load data from Kafka

1. Import required libraries
   from pyspark.sql import SparkSession
   from pyspark.sql.functions import *

2. Create a Spark Session
   spark = SparkSession \
        .builder \
        .appName("Kafka-to-local") \
        .getOrCreate()

   3. Read data from Kafka topic from bootstrap server 18.211.252.152:9092 by subscribing to de-capstone3
        df = spark.readStream \
        .format("kafka") \
        .option("kafka.bootstrap.servers", "18.211.252.152:9092") \
        .option("startingOffsets", "earliest") \
        .option("subscribe", "de-capstone3") \
        .load()

4. Drop irrelevant columns
        df= df \
        .withColumn('value_str',df['value'].cast('string').alias('key_str')).drop('value') \
        .drop('key','topic','partition','offset','timestamp','timestampType')

Steps to load the data into Hadoop
        df.writeStream \
        .format("json") \
        .outputMode("append") \
        .option("path", "/user/root/clickstream_dump_op") \
        .option("checkpointLocation", "/user/root/clickstream_dump_cp") \
        .start() \
        .awaitTermination()

Screenshots of the data

Data before flattening

estamp\\n\": \"2020-01-16 08:33:01\\n\"}"}
{"value_str":"{\"customer_id\": \"13532158\", \"app_version\": \"1.2.18\", \"OS_version\": \"Android\", \"lat\": \"22.191640\", \"lon\": \"-142.982899\", \"page_id\": \"de545711-3914-4450-8c11-b17b8dabb5e1\", \"button_id\": \"e1e99492-17ae-11eb-adc1-0242ac120002\", \"is_button_click\": \"No\", \"is_page_view\": \"Yes\", \"is_scroll_up\": \"No\", \"is_scroll_down\": \"Yes\", \"timestamp\\n\": \"2020-11-03 20:04:01\\n\"}"}
{"value_str":"{\"customer_id\": \"94779920\", \"app_version\": \"2.1.40\", \"OS_version\": \"Android\", \"lat\": \"-61.9647605\", \"lon\": \"-0.887637\", \"page_id\": \"e7bc5fb2-1231-11eb-adc1-0242ac120002\", \"button_id\": \"a95dd57b-779f-49db-819d-b6960483e554\", \"is_button_click\": \"Yes\", \"is_page_view\": \"No\", \"is_scroll_up\": \"Yes\", \"is_scroll_down\": \"Yes\", \"timestamp\\n\": \"2020-07-19 08:25:18\\n\"}"}
{"value_str":"{\"customer_id\": \"12802178\", \"app_version\": \"4.3.11\", \"OS_version\": \"iOS\", \"lat\": \"-67.7758015\", \"lon\": \"-46.871853\", \"page_id\": \"b328829e-17ae-11eb-adc1-0242ac120002\", \"button_id\": \"fcba68aa-1231-11eb-adc1-0242ac120002\", \"is_button_click\": \"Yes\", \"is_page_view\": \"Yes\", \"is_scroll_up\": \"Yes\", \"is_scroll_down\": \"Yes\", \"timestamp\\n\": \"2020-05-11 20:11:53\\n\"}"}
{"value_str":"{\"customer_id\": \"17132950\", \"app_version\": \"1.1.17\", \"OS_version\": \"iOS\", \"lat\": \"-81.4539895\", \"lon\": \"-146.000945\", \"page_id\": \"de545711-3914-4450-8c11-b17b8dabb5e1\", \"button_id\": \"fcba68aa-1231-11eb-adc1-0242ac120002\", \"is_button_click\": \"No\", \"is_page_view\": \"Yes\", \"is_scroll_up\": \"Yes\", \"is_scroll_down\": \"No\", \"timestamp\\n\": \"2020-01-26 02:05:01\\n\"}"}
{"value_str":"{\"customer_id\": \"89444185\", \"app_version\": \"1.1.23\", \"OS_version\": \"Android\", \"lat\": \"-75.4926585\", \"lon\": \"116.182403\", \"page_id\": \"b328829e-17ae-11eb-adc1-0242ac120002\", \"button_id\": \"e1e99492-17ae-11eb-adc1-0242ac120002\", \"is_button_click\": \"Yes\", \"is_page_view\": \"Yes\", \"is_scroll_up\": \"No\", \"is_scroll_down\": \"No\", \"timestamp\\n\": \"2020-09-02 10:49:26\\n\"}"}
{"value_str":"{\"customer_id\": \"29809194\", \"app_version\": \"2.4.35\", \"OS_version\": \"Android\", \"lat\": \"47.9158065\", \"lon\": \"58.444928\", \"page_id\": \"e7bc5fb2-1231-11eb-adc1-0242ac120002\", \"button_id\": \"a95dd57b-779f-49db-819d-b6960483e554\", \"is_button_click\": \"No\", \"is_page_view\": \"No\", \"is_scroll_up\": \"Yes\", \"is_scroll_down\": \"No\", \"timestamp\\n\": \"2020-08-06 05:23:48\\n\"}"}
{"value_str":"{\"customer_id\": \"31264188\", \"app_version\": \"2.2.2\", \"OS_version\": \"iOS\", \"lat\": \"21.2509885\", \"lon\": \"173.175022\", \"page_id\": \"e7bc5fb2-1231-11eb-adc1-0242ac120002\", \"button_id\": \"a95dd57b-779f-49db-819d-b6960483e554\", \"is_button_click\": \"No\", \"is_page_view\": \"No\", \"is_scroll_up\": \"Yes\", \"is_scroll_down\": \"No\", \"timestamp\\n\": \"2020-08-22 14:34:08\\n\"}"}
{"value_str":"{\"customer_id\": \"70397626\", \"app_version\": \"4.1.18\", \"OS_version\": \"iOS\", \"lat\": \"85.040241\", \"lon\": \"162.424346\", \"page_id\": \"b328829e-17ae-11eb-adc1-0242ac120002\", \"button_id\": \"e1e99492-17ae-11eb-adc1-0242ac120002\", \"is_button_click\": \"No\", \"is_page_view\": \"No\", \"is_scroll_up\": \"Yes\", \"is_scroll_down\": \"Yes\", \"timestamp\\n\": \"2020-02-10 03:33:21\\n\"}"}
{"value_str":"{\"customer_id\": \"83555907\", \"app_version\": \"2.1.8\", \"OS_version\": \"Android\", \"lat\": \"-67.2131345\", \"lon\": \"-14.820256\", \"page_id\": \"b328829e-17ae-11eb-adc1-0242ac120002\", \"button_id\": \"e1e99492-17ae-11eb-adc1-0242ac120002\", \"is_button_click\": \"Yes\", \"is_page_view\": \"Yes\", \"is_scroll_up\": \"Yes\", \"is_scroll_down\": \"No\", \"timestamp\\n\": \"2020-04-14 14:42:36\\n\"}"}
{"value_str":"{\"customer_id\": \"18607720\", \"app_version\": \"2.1.36\", \"OS_version\": \"iOS\", \"lat\": \"1.284805\", \"lon\": \"55.568818\", \"page_id\": \"b328829e-17ae-11eb-adc1-0242ac120002\", \"button_id\": \"e1e99492-17ae-11eb-adc1-0242ac120002\", \"is_button_click\": \"No\", \"is_page_view\": \"Yes\", \"is_scroll_up\": \"No\", \"is_scroll_down\": \"No\", \"timestamp\\n\": \"2020-05-27 11:33:12\\n\"}"}
{"value_str":"{\"customer_id\": \"53000914\", \"app_version\": \"1.4.21\", \"OS_version\": \"Android\", \"lat\": \"-61.983406\", \"lon\": \"10.156343\", \"page_id\": \"b328829e-17ae-11eb-adc1-0242ac120002\", \"button_id\": \"fcba68aa-1231-11eb-adc1-0242ac120002\", \"is_button_click\": \"No\", \"is_page_view\": \"No\", \"is_scroll_up\": \"Yes\", \"is_scroll_down\": \"Yes\", \"timestamp\\n\": \"2020-01-28 09:44:38\\n\"}"}
{"value_str":"{\"customer_id\": \"14058847\", \"app_version\": \"3.3.5\", \"OS_version\": \"iOS\", \"lat\": \"10.525097\", \"lon\": \"-168.264407\", \"page_id\": \"b328829e-17ae-11eb-adc1-0242ac120002\", \"button_id\": \"e1e99492-17ae-11eb-adc1-0242ac120002\", \"is_button_click\": \"Yes\", \"is_page_view\": \"Yes\", \"is_scroll_up\": \"Yes\", \"is_scroll_down\": \"No\", \"timestamp\\n\": \"2020-05-26 00:55:32\\n\"}"}
{"value_str":"{\"customer_id\": \"66911192\", \"app_version\": \"1.3.4\", \"OS_version\": \"Android\", \"lat\": \"58.6922195\", \"lon\": \"101.989138\", \"page_id\": \"b328829e-17ae-11eb-adc1-0242ac120002\", \"button_id\": \"fcba68aa-1231-11eb-adc1-0242ac120002\", \"is_button_click\": \"Yes\", \"is_page_view\": \"Yes\", \"is_scroll_up\": \"No\", \"is_scroll_down\": \"No\", \"timestamp\\n\": \"2020-08-30 05:29:39\\n\"}"}
{"value_str":"{\"customer_id\": \"21592555\", \"app_version\": \"2.1.19\", \"OS_version\": \"Android\", \"lat\": \"-56.3204135\", \"lon\": \"149.467240\", \"page_id\": \"de545711-3914-4450-8c11-b17b8dabb5e1\", \"button_id\": \"a95dd57b-779f-49db-819d-b6960483e554\", \"is_button_click\": \"Yes\", \"is_page_view\": \"Yes\", \"is_scroll_up\": \"No\", \"is_scroll_down\": \"No\", \"timestamp\\n\": \"2020-05-06 20:40:59\\n\"}"}
{"value_str":"{\"customer_id\": \"84122684\", \"app_version\": \"2.4.37\", \"OS_version\": \"iOS\", \"lat\": \"25.619935\", \"lon\": \"-150.280041\", \"page_id\": \"e7bc5fb2-1231-11eb-adc1-0242ac120002\", \"button_id\": \"a95dd57b-779f-49db-819d-b6960483e554\", \"is_button_click\": \"No\", \"is_page_view\": \"Yes\", \"is_scroll_up\": \"No\", \"is_scroll_down\": \"No\", \"timestamp\\n\": \"2020-04-02 06:07:41\\n\"}"}
{"value_str":"{\"customer_id\": \"23425453\", \"app_version\": \"4.4.1\", \"OS_version\": \"Android\", \"lat\": \"21.4624825\", \"lon\": \"116.578345\", \"page_id\": \"e7bc5fb2-1231-11eb-adc1-0242ac120002\", \"button_id\": \"e1e99492-17ae-11eb-adc1-0242ac120002\", \"is_button_click\": \"Yes\", \"is_page_view\": \"Yes\", \"is_scroll_up\": \"Yes\", \"is_scroll_down\": \"No\", \"timestamp\\n\": \"2020-08-22 16:00:11\\n\"}"}
{"value_str":"{\"customer_id\": \"40925979\", \"app_version\": \"4.4.12\", \"OS_version\": \"Android\", \"lat\": \"83.2222865\", \"lon\": \"-175.942391\", \"page_id\": \"de545711-3914-4450-

Data after structuring

...13550,2,2.218,Android,..,94.2470193,19.700640,de545711-3914-4450-8c11-b17b8dabb5e1,a95dd57b-779f-49db-819d-b6960483e554,No,No,Yes,Yes,""
48927284,1.3.12,iOS,5.9788155,151.064029,e7bc5fb2-1231-11eb-adc1-0242ac120002,fcba68aa-1231-11eb-adc1-0242ac120002,No,Yes,No,No,""
85177527,4.2.10,iOS,-51.8301325,-70.399319,e7bc5fb2-1231-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,Yes,No,No,Yes,""
47743867,3.3.20,iOS,-13.990733,-178.175285,b328829e-17ae-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,Yes,No,No,No,""
46894536,4.2.31,iOS,-59.551650,115.343099,e7bc5fb2-1231-11eb-adc1-0242ac120002,fcba68aa-1231-11eb-adc1-0242ac120002,Yes,No,Yes,Yes,""
87861168,4.4.27,iOS,-9.0299885,87.753166,de545711-3914-4450-8c11-b17b8dabb5e1,fcba68aa-1231-11eb-adc1-0242ac120002,Yes,No,No,Yes,""
98750961,1.4.24,Android,-67.056611,-42.210252,b328829e-17ae-11eb-adc1-0242ac120002,e1e99492-17ae-11eb-adc1-0242ac120002,Yes,Yes,Yes,No,""
34179709,2.4.1,iOS,-65.870590,112.347735,de545711-3914-4450-8c11-b17b8dabb5e1,e1e99492-17ae-11eb-adc1-0242ac120002,Yes,Yes,Yes,No,""
35577566,2.2.18,Android,28.5281265,28.317597,e7bc5fb2-1231-11eb-adc1-0242ac120002,e1e99492-17ae-11eb-adc1-0242ac120002,Yes,Yes,Yes,No,""
39790044,3.3.32,iOS,60.0575415,-36.785274,b328829e-17ae-11eb-adc1-0242ac120002,fcba68aa-1231-11eb-adc1-0242ac120002,No,Yes,No,No,""
49451715,2.3.8,iOS,77.0398825,174.942366,e7bc5fb2-1231-11eb-adc1-0242ac120002,e1e99492-17ae-11eb-adc1-0242ac120002,No,No,Yes,No,""
87559592,2.2.17,Android,-72.2719865,-66.732794,e7bc5fb2-1231-11eb-adc1-0242ac120002,fcba68aa-1231-11eb-adc1-0242ac120002,No,Yes,No,No,""
20011383,2.1.30,Android,58.3581915,27.525051,b328829e-17ae-11eb-adc1-0242ac120002,e1e99492-17ae-11eb-adc1-0242ac120002,Yes,No,No,Yes,""
83519078,4.2.9,iOS,78.4456805,-83.204209,b328829e-17ae-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,No,No,Yes,No,""
35953253,1.3.11,iOS,29.422319,106.217372,de545711-3914-4450-8c11-b17b8dabb5e1,e1e99492-17ae-11eb-adc1-0242ac120002,No,No,Yes,Yes,""
79134831,1.3.25,iOS,-43.676692,-13.693995,de545711-3914-4450-8c11-b17b8dabb5e1,e1e99492-17ae-11eb-adc1-0242ac120002,Yes,Yes,Yes,No,""
74932660,2.4.11,Android,-16.258038,-67.903992,b328829e-17ae-11eb-adc1-0242ac120002,e1e99492-17ae-11eb-adc1-0242ac120002,No,No,No,Yes,""
48948016,1.2.37,Android,28.116246,80.131224,b328829e-17ae-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,No,No,No,Yes,""
70449656,2.3.23,Android,-38.571418,102.352612,de545711-3914-4450-8c11-b17b8dabb5e1,e1e99492-17ae-11eb-adc1-0242ac120002,No,No,Yes,Yes,""
41398860,4.1.10,Android,-58.4709445,-56.966952,e7bc5fb2-1231-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,No,No,Yes,Yes,""
35844721,2.3.22,Android,0.997151,-56.869833,de545711-3914-4450-8c11-b17b8dabb5e1,fcba68aa-1231-11eb-adc1-0242ac120002,Yes,Yes,Yes,Yes,""
49195530,4.1.38,Android,-65.173138,-45.282215,de545711-3914-4450-8c11-b17b8dabb5e1,fcba68aa-1231-11eb-adc1-0242ac120002,No,No,Yes,No,""
92467259,1.1.28,iOS,-13.9285225,166.976040,e7bc5fb2-1231-11eb-adc1-0242ac120002,e1e99492-17ae-11eb-adc1-0242ac120002,Yes,No,No,No,""
39856207,4.4.21,iOS,-45.9151095,-93.708618,de545711-3914-4450-8c11-b17b8dabb5e1,e1e99492-17ae-11eb-adc1-0242ac120002,Yes,No,No,No,""
47664043,3.3.14,Android,-13.842446,117.938952,e7bc5fb2-1231-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,Yes,Yes,Yes,Yes,""
38729595,1.1.32,iOS,-86.7711605,-133.680955,e7bc5fb2-1231-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,Yes,Yes,Yes,No,""
52083661,1.3.6,Android,-20.051056,-113.408925,b328829e-17ae-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,No,No,Yes,No,""
21248817,2.2.14,iOS,25.7465345,-115.096705,e7bc5fb2-1231-11eb-adc1-0242ac120002,e1e99492-17ae-11eb-adc1-0242ac120002,No,No,Yes,No,""
32887975,2.3.23,Android,-14.8909615,-168.408479,b328829e-17ae-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,Yes,No,Yes,No,""
79298369,4.1.11,iOS,-3.7870705,128.420903,b328829e-17ae-11eb-adc1-0242ac120002,e1e99492-17ae-11eb-adc1-0242ac120002,No,No,No,Yes,""
74700232,3.1.6,iOS,38.679099,-26.254784,b328829e-17ae-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,No,No,No,No,""
52434368,4.4.13,Android,86.2803345,-4.684262,e7bc5fb2-1231-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,No,No,Yes,No,""
83053436,1.1.25,iOS,69.6784265,-93.018506,e7bc5fb2-1231-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,No,Yes,Yes,No,""
36425257,3.4.9,iOS,88.945103,171.605415,b328829e-17ae-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,Yes,No,No,No,""
09571102,1.1.39,iOS,-64.4794635,9.497410,de545711-3914-4450-8c11-b17b8dabb5e1,e1e99492-17ae-11eb-adc1-0242ac120002,No,No,No,Yes,""
07051710,4.2.3,iOS,-5.041729,29.906932,e7bc5fb2-1231-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,Yes,No,No,No,""
84077369,1.3.36,Android,83.702958,41.236421,e7bc5fb2-1231-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,No,Yes,Yes,No,""
30901771,1.3.32,iOS,33.601921,-18.663633,de545711-3914-4450-8c11-b17b8dabb5e1,fcba68aa-1231-11eb-adc1-0242ac120002,No,No,Yes,Yes,""
67047855,4.3.3,iOS,50.817671,21.621528,de545711-3914-4450-8c11-b17b8dabb5e1,fcba68aa-1231-11eb-adc1-0242ac120002,Yes,No,Yes,Yes,""
50949577,2.3.21,Android,-19.0183995,-20.792102,de545711-3914-4450-8c11-b17b8dabb5e1,fcba68aa-1231-11eb-adc1-0242ac120002,No,No,No,Yes,""
67036669,4.1.9,Android,-68.149560,37.836674,b328829e-17ae-11eb-adc1-0242ac120002,e1e99492-17ae-11eb-adc1-0242ac120002,No,No,No,No,""
27315595,1.4.18,iOS,-52.566865,-177.174549,de545711-3914-4450-8c11-b17b8dabb5e1,a95dd57b-779f-49db-819d-b6960483e554,No,No,No,No,""
70426942,2.3.13,iOS,76.747663,93.546816,b328829e-17ae-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,Yes,No,Yes,No,""
10895363,1.1.1,Android,-10.027164,46.705580,de545711-3914-4450-8c11-b17b8dabb5e1,e1e99492-17ae-11eb-adc1-0242ac120002,Yes,No,No,No,""
99247595,2.3.15,Android,15.9023165,49.306042,e7bc5fb2-1231-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,No,Yes,Yes,No,""
63762620,3.2.1,Android,60.507770,56.985170,de545711-3914-4450-8c11-b17b8dabb5e1,e1e99492-17ae-11eb-adc1-0242ac120002,No,No,Yes,No,""
84418220,4.1.40,iOS,64.6043925,-53.285161,b328829e-17ae-11eb-adc1-0242ac120002,fcba68aa-1231-11eb-adc1-0242ac120002,No,No,No,Yes,""
27581545,1.2.21,Android,-63.363132,-71.337368,b328829e-17ae-11eb-adc1-0242ac120002,e1e99492-17ae-11eb-adc1-0242ac120002,Yes,Yes,Yes,Yes,""
68526728,1.1.35,Android,28.0262385,118.632557,e7bc5fb2-1231-11eb-adc1-0242ac120002,fcba68aa-1231-11eb-adc1-0242ac120002,Yes,Yes,Yes,Yes,""
68929628,1.1.10,iOS,-38.562532,120.214812,e7bc5fb2-1231-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,No,No,No,Yes,""