

Load data from Kafka to Hadoop

Commands

spark_kafka_to_local.py is needed to load data to Hadoop. Load this file to root level in ec2-instance.

Run this command in terminal

```
spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.11:2.4.5 spark_kafka_to_local.py
```

After running the command, wait for a while until we see a batch table showing in the console.

When we see the batch table, check if the data is correctly loaded to HDFS in another terminal using this command.

Hadoop fs -ls clickstream-data

```
hadoop@ip-172-31-1-153:~  
[hadoop@ip-172-31-1-153 ~]$ hadoop fs -ls  
Found 2 items  
drwxr-xr-x - hadoop hadoop 0 2022-06-15 13:22 clickstream-data  
drwxr-xr-x - hadoop hadoop 0 2022-06-15 13:22 clickstream-data-cp  
[hadoop@ip-172-31-1-153 ~]$ hadoop fs -ls clickstream-data  
Found 2 items  
drwxr-xr-x - hadoop hadoop 0 2022-06-15 13:22 clickstream-data/ spark metadata  
-rw-r--r-- 1 hadoop hadoop 1000605 2022-06-15 13:22 clickstream-data/part-00000-a17b6f03-d71b-4373-ba2d-e5a96dff8d71-c000.json  
[hadoop@ip-172-31-1-153 ~]$ hadoop fs -ls clickstream-data  
Found 2 items  
drwxr-xr-x - hadoop hadoop 0 2022-06-15 13:22 clickstream-data/ spark metadata  
-rw-r--r-- 1 hadoop hadoop 1000605 2022-06-15 13:22 clickstream-data/part-00000-a17b6f03-d71b-4373-ba2d-e5a96dff8d71-c000.json
```

```
hadoop@ip-172-31-1-153:~  
[hadoop@ip-172-31-1-153 ~]$ hadoop fs -cat clickstream-data/part-00000-a17b6f03-d71b-4373-ba2d-e5a96dff8d71-c000.json |more  
{  
  "customer_id": "26564820", "app_version": "3.2.35", "os_version": "Android", "lat": "16.445465", "lon": "99.902065", "page_id": "de545711-3914-4450-8c11-b17b8dabb5e1", "button_id":  
  "fcb6a8aa-1231-11eb-adc1-0242ac120002", "is_button_click": "No", "is page view": "Yes", "is scroll up": "No", "is scroll down": "Yes", "timestamp": "2020-09-14 09:59:07\n"}  
{  
  "customer_id": "31906387", "app_version": "2.4.7", "os_version": "iOS", "lat": "-64.813749", "lon": "-133.527040", "page_id": "de545711-3914-4450-8c11-b17b8dabb5e1", "button_id":  
  "a95dd57b-779f-49db-b19d-b6960483e554", "is_button_click": "No", "is page view": "No", "is scroll up": "Yes", "is scroll down": "Yes", "timestamp": "2020-05-16 16:30:21\n"}  
{  
  "customer_id": "25713677", "app_version": "3.4.12", "os_version": "Android", "lat": "89.943435", "lon": "127.313415", "page_id": "b328829e-17ae-11eb-adc1-0242ac120002", "button_id":  
  "fcb6a8aa-1231-11eb-adc1-0242ac120002", "is_button_click": "No", "is page view": "No", "is scroll up": "Yes", "is scroll down": "No", "timestamp": "2020-02-09 00:52:13\n"}  
{  
  "customer_id": "83474293", "app_version": "3.1.8", "os_version": "Android", "lat": "-69.939070", "lon": "-36.451670", "page_id": "e7bc5fb2-1231-11eb-adc1-0242ac120002", "button_id":  
  "e1e99492-17ae-11eb-adc1-0242ac120002", "is_button_click": "Yes", "is page view": "No", "is scroll up": "Yes", "is scroll down": "No", "timestamp": "2020-06-17 10:42:50\n"}  
{  
  "customer_id": "63727807", "app_version": "2.2.9", "os_version": "iOS", "lat": "64.082108", "lon": "-81.822078", "page_id": "e7bc5fb2-1231-11eb-adc1-0242ac120002", "button_id": "fcb6a8aa-1231-11eb-adc1-0242ac120002", "is_button_click": "No", "is page view": "Yes", "is scroll up": "Yes", "is scroll down": "Yes", "timestamp": "2020-07-06 02:51:53\n"}  
{  
  "customer_id": "73737907", "app_version": "4.3.15", "os_version": "Android", "lat": "-15.850508", "lon": "-116.358375", "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": "Yes", "timestamp": "2020-04-26 06:18:14\n"}  
{  
  "customer_id": "36927433", "app_version": "2.2.25", "os_version": "iOS", "lat": "-84.6857245", "lon": "-146.507678", "page_id": "de545711-3914-4450-8c11-b17b8dabb5e1", "button_id":  
  "a95dd57b-779f-49db-b19d-b6960483e554", "is_button_click": "Yes", "is page view": "Yes", "is scroll up": "No", "is scroll down": "Yes", "timestamp": "2020-02-06 10:21:18\n"}  
{  
  "customer_id": "12691783", "app_version": "3.3.11", "os_version": "Android", "lat": "54.3852925", "lon": "-37.411814", "page_id": "de545711-3914-4450-8c11-b17b8dabb5e1", "button_id":  
  "e1e99492-17ae-11eb-adc1-0242ac120002", "is_button_click": "Yes", "is page view": "Yes", "is scroll up": "No", "is scroll down": "No", "timestamp": "2020-08-08 04:23:56\n"}  
{  
  "customer_id": "22635021", "app_version": "4.4.36", "os_version": "iOS", "lat": "-31.805500", "lon": "150.655650", "page_id": "e7bc5fb2-1231-11eb-adc1-0242ac120002", "button_id":  
  "a95dd57b-779f-49db-b19d-b6960483e554", "is_button_click": "No", "is page view": "No", "is scroll up": "No", "is scroll down": "No", "timestamp": "2020-08-02 00:33:50\n"}  
{  
  "customer_id": "23539346", "app_version": "1.2.16", "os_version": "Android", "lat": "8.391849", "lon": "-83.929578", "page_id": "de545711-3914-4450-8c11-b17b8dabb5e1", "button_id":  
  "e1e99492-17ae-11eb-adc1-0242ac120002", "is_button_click": "Yes", "is page view": "No", "is scroll up": "Yes", "is scroll down": "No", "timestamp": "2020-07-23 23:59:19\n"}  
{  
  "customer_id": "16929816", "app_version": "1.2.25", "os_version": "Android", "lat": "-34.8576385", "lon": "-136.639835", "page_id": "e7bc5fb2-1231-11eb-adc1-0242ac120002", "button_id":  
  "fcb6a8aa-1231-11eb-adc1-0242ac120002", "is_button_click": "No", "is page view": "No", "is scroll up": "No", "is scroll down": "Yes", "timestamp": "2020-05-13 16:35:17\n"}  
{  
  "customer_id": "41929865", "app_version": "1.1.14", "os_version": "iOS", "lat": "-34.3613155", "lon": "-137.467833", "page_id": "b328829e-17ae-11eb-adc1-0242ac120002", "button_id":  
  "a95dd57b-779f-49db-b19d-b6960483e554", "is_button_click": "No", "is page view": "Yes", "is scroll up": "No", "is scroll down": "No", "timestamp": "2020-04-13 07:58:26\n"}  
{  
  "customer_id": "60940300", "app_version": "1.1.17", "os_version": "Android", "lat": "81.730235", "lon": "104.207743", "page_id": "de545711-3914-4450-8c11-b17b8dabb5e1", "button_id":  
  "fcb6a8aa-1231-11eb-adc1-0242ac120002", "is_button_click": "Yes", "is page view": "No", "is scroll up": "Yes", "is scroll down": "No", "timestamp": "2020-10-14 00:39:05\n"}  
{  
  "customer_id": "95405928", "app_version": "1.2.10", "os_version": "Android", "lat": "44.185627", "lon": "78.125488", "page_id": "e7bc5fb2-1231-11eb-adc1-0242ac120002", "button_id":  
  "e1e99492-17ae-11eb-adc1-0242ac120002", "is_button_click": "No", "is page view": "No", "is scroll up": "No", "is scroll down": "No", "timestamp": "2020-07-03 19:15:34\n"}  
{  
  "customer_id": "56235860", "app_version": "1.4.16", "os_version": "Android", "lat": "-61.178407", "lon": "8.534126", "page_id": "de545711-3914-4450-8c11-b17b8dabb5e1", "button_id":  
  "a95dd57b-779f-49db-b19d-b6960483e554", "is_button_click": "Yes", "is page view": "Yes", "is scroll up": "Yes", "is scroll down": "Yes", "timestamp": "2020-06-02 21:32:44\n"}  
{  
  "customer_id": "81758815", "app_version": "4.4.30", "os_version": "Android", "lat": "78.904984", "lon": "167.315596", "page_id": "e7bc5fb2-1231-11eb-adc1-0242ac120002", "button_id":  
  "a95dd57b-779f-49db-b19d-b6960483e554", "is_button_click": "Yes", "is page view": "Yes", "is scroll up": "No", "is scroll down": "No", "timestamp": "2020-04-17 16:52:16\n"}  
{  
  "customer_id": "46822067", "app_version": "1.2.26", "os_version": "Android", "lat": "45.893985", "lon": "135.895118", "page_id": "de545711-3914-4450-8c11-b17b8dabb5e1", "button_id":  
  "fcb6a8aa-1231-11eb-adc1-0242ac120002", "is_button_click": "Yes", "is page view": "Yes", "is scroll up": "Yes", "is scroll down": "Yes", "timestamp": "2020-08-09 12:02:09\n"}  
{  
  "customer_id": "37623067", "app_version": "3.2.27", "os_version": "iOS", "lat": "-74.1131505", "lon": "-170.715421", "page_id": "b328829e-17ae-11eb-adc1-0242ac120002", "button_id":  
  "a95dd57b-779f-49db-b19d-b6960483e554", "is_button_click": "No", "is page view": "Yes", "is scroll up": "No", "is scroll down": "No", "timestamp": "2020-01-20 00:06:06\n"}  
{  
  "customer_id": "14571639", "app_version": "3.7.21", "os_version": "iOS", "lat": "-28.8329055", "lon": "141.133169", "page_id": "e7bc5fb2-1231-11eb-adc1-0242ac120002", "button_id":  
  "fcb6a8aa-1231-11eb-adc1-0242ac120002", "is_button_click": "No", "is page view": "Yes", "is scroll up": "No", "is scroll down": "No", "timestamp": "2020-08-16 11:20:40\n"}  
{  
  "customer_id": "66065529", "app_version": "2.1.36", "os_version": "Android", "lat": "44.196239", "lon": "-15.354515", "page_id": "e7bc5fb2-1231-11eb-adc1-0242ac120002", "button_id":  
  "a95dd57b-779f-49db-b19d-b6960483e554", "is_button_click": "Yes", "is page view": "No", "is scroll up": "No", "is scroll down": "No", "timestamp": "2020-06-24 01:55:32\n"}  
{  
  "customer_id": "83657253", "app_version": "2.1.2", "os_version": "iOS", "lat": "38.7036005", "lon": "-59.081929", "page_id": "b328829e-17ae-11eb-adc1-0242ac120002", "button_id": "fcb6a8aa-1231-11eb-adc1-0242ac120002", "is_button_click": "Yes", "is page view": "Yes", "is scroll up": "Yes", "is scroll down": "Yes", "timestamp": "2020-08-09 12:02:09\n"}  
}
```