

## 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:-

```

{"value":{"\\"customer_id\\":\\"71156145\\",\\"app_version\\":\\"4.1.14\\",\\"OS_version\\":\\"Android\\",\\"lat\\":\\"-24.5899605\\",\\"lon\\":\\"67.624771\\",\\"page_id\\":\\"de54711-3914-4450-8c11-b17b8dabb5e1\\",\\"button_id\\":\\"fcfcb68aa-1231-11eb-adc1-0242ac120002\\",\\"is_button_click\\":\\"No\\",\\"is_page_view\\":\\"No\\",\\"is_scroll_up\\":\\"No\\",\\"is_scroll_down\\":\\"Yes\\",\\"timestamp\\":\\"2020-01-09 15:54:48[\\n\\"]\\"}}
{"value":{"\\"customer_id\\":\\"72250597\\",\\"app_version\\":\\"4.1.14\\",\\"OS_version\\":\\"iOS\\",\\"lat\\":\\"-69.1155955\\",\\"lon\\":\\"-143.163298\\",\\"page_id\\":\\"de54511-3914-4450-8c11-b17b8dabb5e1\\",\\"button_id\\":\\"e1e599492-17ae-11eb-adc1-0242ac120002\\",\\"is_button_click\\":\\"Yes\\",\\"is_page_view\\":\\"No\\",\\"is_scroll_up\\":\\"No\\",\\"is_scroll_down\\":\\"Yes\\",\\"timestamp\\":\\"2020-06-26 09:08:00[\\n\\"]\\"}}
{"value":{"\\"customer_id\\":\\"85382952\\",\\"app_version\\":\\"3.2.24\\",\\"OS_version\\":\\"Android\\",\\"lat\\":\\"-47.734696\\",\\"lon\\":\\"85.891429\\",\\"page_id\\":\\"b32829e-17ae-11eb-adc1-0242ac120002\\",\\"button_id\\":\\"fcfcb68aa-1231-11eb-adc1-0242ac120002\\",\\"is_button_click\\":\\"Yes\\",\\"is_page_view\\":\\"No\\",\\"is_scroll_up\\":\\"Yes\\",\\"is_scroll_down\\":\\"No\\",\\"timestamp\\":\\"2020-04-11 10:41:05[\\n\\"]\\"}}
{"value":{"\\"customer_id\\":\\"421615322\\",\\"app_version\\":\\"4.1.13\\",\\"OS_version\\":\\"iOS\\",\\"lat\\":\\"80.8403425\\",\\"lon\\":\\"-86.627180\\",\\"page_id\\":\\"de545711-3914-4450-8c11-b17b8dabb5e1\\",\\"button_id\\":\\"a95dd57b-779f-49db-819d-b6960483e554\\",\\"is_button_click\\":\\"Yes\\",\\"is_page_view\\":\\"No\\",\\"is_scroll_up\\":\\"No\\",\\"is_scroll_down\\":\\"Yes\\",\\"timestamp\\":\\"2020-03-27 14:03:03[\\n\\"]\\"}}
{"value":{"\\"customer_id\\":\\"12222347\\",\\"app_version\\":\\"11.2.27\\",\\"OS_version\\":\\"iOS\\",\\"lat\\":\\"-3.6452735\\",\\"lon\\":\\"173.446829\\",\\"page_id\\":\\"b328829e-17ae-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\\":\\"2020-07-19 15:48:28[\\n\\"]\\"}}
{"value":{"\\"customer_id\\":\\"63979784\\",\\"app_version\\":\\"3.2.24\\",\\"OS_version\\":\\"iOS\\",\\"lat\\":\\"52.6143785\\",\\"lon\\":\\"-4.783003\\",\\"page_id\\":\\"b328829e-17ae-11eb-adc1-0242ac120002\\",\\"button_id\\":\\"fcfcb68aa-1231-11eb-adc1-0242ac120002\\",\\"is_button_click\\":\\"Yes\\",\\"is_page_view\\":\\"No\\",\\"is_scroll_up\\":\\"No\\",\\"is_scroll_down\\":\\"Yes\\",\\"timestamp\\":\\"2020-08-16 13:47:47[\\n\\"]\\"}}
{"value":{"\\"customer_id\\":\\"25631527\\",\\"app_version\\":\\"12.1.6\\",\\"OS_version\\":\\"Android\\",\\"lat\\":\\"73.0767505\\",\\"lon\\":\\"13.785813\\",\\"page_id\\":\\"b328829e-17ae-11eb-adc1-0242ac120002\\",\\"button_id\\":\\"e1e599492-17ae-11eb-adc1-0242ac120002\\",\\"is_button_click\\":\\"No\\",\\"is_page_view\\":\\"Yes\\",\\"is_scroll_up\\":\\"Yes\\",\\"is_scroll_down\\":\\"Yes\\",\\"timestamp\\":\\"2020-01-23 08:41:36[\\n\\"]\\"}}
{"value":{"\\"customer_id\\":\\"92397542\\",\\"app_version\\":\\"2.3.16\\",\\"OS_version\\":\\"iOS\\",\\"lat\\":\\"-7.0993865\\",\\"lon\\":\\"-40.763662\\",\\"page_id\\":\\"de545711-3914-4450-8c11-b17b8dabb5e1\\",\\"button_id\\":\\"fcfcb68aa-1231-11eb-adc1-0242ac120002\\",\\"is_button_click\\":\\"Yes\\",\\"is_page_view\\":\\"No\\",\\"is_scroll_up\\":\\"No\\",\\"is_scroll_down\\":\\"No\\",\\"timestamp\\":\\"2020-08-02 07:44:14[\\n\\"]\\"}}
{"value":{"\\"customer_id\\":\\"15962691\\",\\"app_version\\":\\"4.1.14\\",\\"OS_version\\":\\"iOS\\",\\"lat\\":\\"44.1034375\\",\\"lon\\":\\"-158.429047\\",\\"page_id\\":\\"b328829e-17ae-11eb-adc1-0242ac120002\\",\\"button_id\\":\\"a95dd57b-779f-49db-819d-b6960483e554\\",\\"is_button_click\\":\\"Yes\\",\\"is_page_view\\":\\"Yes\\",\\"is_scroll_up\\":\\"Yes\\",\\"is_scroll_down\\":\\"Yes\\",\\"timestamp\\":\\"2020-08-17 15:58:13[\\n\\"]\\"}}
{"value":{"\\"customer_id\\":\\"75574770\\",\\"app_version\\":\\"4.1.39\\",\\"OS_version\\":\\"iOS\\",\\"lat\\":\\"-82.045519\\",\\"lon\\":\\"-60.819861\\",\\"page_id\\":\\"b328829e-17ae-11eb-adc1-0242ac120002\\",\\"button_id\\":\\"a95dd57b-779f-49db-819d-b6960483e554\\",\\"is_button_click\\":\\"Yes\\",\\"is_page_view\\":\\"Yes\\",\\"is_scroll_up\\":\\"Yes\\",\\"is_scroll_down\\":\\"Yes\\",\\"timestamp\\":\\"2020-11-08 12:15:57[\\n\\"]\\"}}
{"value":{"\\"customer_id\\":\\"21922061\\",\\"app_version\\":\\"3.1.35\\",\\"OS_version\\":\\"iOS\\",\\"lat\\":\\"-72.0056295\\",\\"lon\\":\\"119.587231\\",\\"page_id\\":\\"de545711-3914-4450-8c11-b17b8dabb5e1\\",\\"button_id\\":\\"e1e599492-17ae-11eb-adc1-0242ac120002\\",\\"is_button_click\\":\\"No\\",\\"is_page_view\\":\\"No\\",\\"is_scroll_up\\":\\"No\\",\\"is_scroll_down\\":\\"No\\",\\"timestamp\\":\\"2020-07-25 10:16:07[\\n\\"]\\"}}
{"value":{"\\"customer_id\\":\\"27178502\\",\\"app_version\\":\\"2.1.39\\",\\"OS_version\\":\\"iOS\\",\\"lat\\":\\"25.9298745\\",\\"lon\\":\\"125.573580\\",\\"page_id\\":\\"de545711-3914-4450-8c11-b17b8dabb5e1\\",\\"button_id\\":\\"a95dd57b-779f-49db-819d-b6960483e554\\",\\"is_button_click\\":\\"Yes\\",\\"is_page_view\\":\\"No\\",\\"is_scroll_up\\":\\"Yes\\",\\"is_scroll_down\\":\\"Yes\\",\\"timestamp\\":\\"2020-10-17 15:41:41[\\n\\"]\\"}}
{"value":{"\\"customer_id\\":\\"55160974\\",\\"app_version\\":\\"3.2.20\\",\\"OS_version\\":\\"Android\\",\\"lat\\":\\"71.250396\\",\\"lon\\":\\"114.330952\\",\\"page_id\\":\\"b328829e-17ae-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\\":\\"Yes\\",\\"timestamp\\":\\"2020-10-24 23:26:26[\\n\\"]\\"}}
{"value":{"\\"customer_id\\":\\"26893884\\",\\"app_version\\":\\"2.4.40\\",\\"OS_version\\":\\"Android\\",\\"lat\\":\\"62.963317\\",\\"lon\\":\\"-115.642762\\",\\"page_id\\":\\"b328829e-17ae-11eb-adc1-0242ac120002\\",\\"button_id\\":\\"fcfcb68aa-1231-11eb-adc1-0242ac120002\\",\\"is_button_click\\":\\"No\\",\\"is_page_view\\":\\"No\\",\\"is_scroll_up\\":\\"No\\",\\"is_scroll_down\\":\\"No\\",\\"timestamp\\":\\"2020-09-13 13:21:43[\\n\\"]\\"}}
{"value":{"\\"customer_id\\":\\"75574770\\",\\"app_version\\":\\"4.1.39\\",\\"OS_version\\":\\"iOS\\",\\"lat\\":\\"-82.045519\\",\\"lon\\":\\"-60.819861\\",\\"page_id\\":\\"b328829e-17ae-11eb-adc1-0242ac120002\\",\\"button_id\\":\\"a95dd57b-779f-49db-819d-b6960483e554\\",\\"is_button_click\\":\\"Yes\\",\\"is_page_view\\":\\"Yes\\",\\"is_scroll_up\\":\\"Yes\\",\\"is_scroll_down\\":\\"Yes\\",\\"timestamp\\":\\"2020-11-08 12:15:57[\\n\\"]\\"}}
{"value":{"\\"customer_id\\":\\"21922061\\",\\"app_version\\":\\"3.1.35\\",\\"OS_version\\":\\"iOS\\",\\"lat\\":\\"-72.0056295\\",\\"lon\\":\\"119.587231\\",\\"page_id\\":\\"de545711-3914-4450-8c11-b17b8dabb5e1\\",\\"button_id\\":\\"e1e599492-17ae-11eb-adc1-0242ac120002\\",\\"is_button_click\\":\\"No\\",\\"is_page_view\\":\\"No\\",\\"is_scroll_up\\":\\"No\\",\\"is_scroll_down\\":\\"No\\",\\"timestamp\\":\\"2020-07-25 10:16:07[\\n\\"]\\"}}
{"value":{"\\"customer_id\\":\\"27178502\\",\\"app_version\\":\\"2.1.39\\",\\"OS_version\\":\\"iOS\\",\\"lat\\":\\"25.9298745\\",\\"lon\\":\\"125.573580\\",\\"page_id\\":\\"de545711-3914-4450-8c11-b17b8dabb5e1\\",\\"button_id\\":\\"a95dd57b-779f-49db-819d-b6960483e554\\",\\"is_button_click\\":\\"Yes\\",\\"is_page_view\\":\\"No\\",\\"is_scroll_up\\":\\"Yes\\",\\"is_scroll_down\\":\\"Yes\\",\\"timestamp\\":\\"2020-10-17 15:41:41[\\n\\"]\\"}}
{"value":{"\\"customer_id\\":\\"55160974\\",\\"app_version\\":\\"3.2.20\\",\\"OS_version\\":\\"Android\\",\\"lat\\":\\"71.250396\\",\\"lon\\":\\"114.330952\\",\\"page_id\\":\\"b328829e-17ae-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\\":\\"Yes\\",\\"timestamp\\":\\"2020-10-24
```

Data in CSV format:-

```
hadoop@ip-172-31-26-107:~$ cat /dev/null
23/02/20 11:42:04 INFO StateStoreCoordinatorRef: Registered StateStoreCoordinator endpoint
```

customer_id	app_version	OS_version	lat	lon	page_id	button_id	is_button_click	is_page_view	is_scroll_up	is_scroll_down	date_timestamp
26564820	3.2.35	Android	16.4454865	99.902065	de545711-3914-445... fcb68aa-1231-11e...		No	Yes	No	Yes	2020-09-1
31906387	2.4.7	iOS	-64.813749	-133.527040	de545711-3914-445... a95dd57b-779f-49d...		No	No	Yes	Yes	2020-05-1
25713677	3.4.12	Android	89.943435	127.313415	b328829e-17ae-11e... fcb68aa-1231-11e...		No	No	Yes	No	2020-02-0
83474293	3.1.8	Android	-69.939070	-36.451670	e7bc5fb2-1231-11e... ele99492-17ae-11e...		Yes	No	Yes	No	2020-06-1
63727807	2.2.9	iOS	64.082108	-81.822078	e7bc5fb2-1231-11e... fcb68aa-1231-11e...		No	Yes	Yes	Yes	2020-07-0
73737907	4.3.19	Android	-18.850508	-116.358375	b328829e-17ae-11e... ele99492-17ae-11e...		No	Yes	No	Yes	2020-04-2
36927433	3.2.26	iOS	-84.6857245	-146.507678	de545711-3914-445... a95dd57b-779f-49d...		Yes	Yes	No	Yes	2020-02-0
12691793	3.3.11	Android	54.3852925	-37.411814	de545711-3914-445... ele99492-17ae-11e...		Yes	Yes	No	No	2020-08-0
22635021	4.4.36	iOS	-31.805500	150.655650	e7bc5fb2-1231-11e... a95dd57b-779f-49d...		No	No	No	No	2020-08-0
23593546	1.2.16	Android	8.8918475	-83.929878	de545711-3914-445... ele99492-17ae-11e...		Yes	No	Yes	No	2020-07-2
16929816	1.2.25	Android	-34.8576385	-136.639835	e7bc5fb2-1231-11e... fcb68aa-1231-11e...		No	No	No	Yes	2020-05-1
41929865	1.1.14	iOS	-34.3613155	-137.467833	b328829e-17ae-11e... a95dd57b-779f-49d...		No	Yes	Yes	No	2020-04-1
68940320	1.1.17	Android	81.738239	-104.207763	de545711-3914-445... fcb68aa-1231-11e...		Yes	No	Yes	No	2020-10-1