

# 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.

Make sure that there exists Kafka version 0.10 inside Cloudera Manager.

If not, go to **Cloudera Manager**, click on **Spark2** service and search for Kafka configuration in it. Change 0.9 to 0.10 and save.

The screenshot shows the Cloudera Manager interface for the Spark 2 service. The top navigation bar includes 'Status', 'Instances', 'Configuration' (selected), 'Commands', 'Charts Library', 'Audits', 'History Server Web UI', and 'Quick Links'. The main content area is titled 'Spark 2' and shows the 'Configuration' tab. A search bar at the top left contains the text 'kafka'. On the left side, there is a 'Filters' panel with 'SCOPE' and 'CATEGORY' sections. The 'SCOPE' section lists 'Spark 2 (Service-Wide)' with a count of 0, 'Gateway' with a count of 1, and 'History Server' with a count of 0. The 'CATEGORY' section lists various categories like 'Advanced', 'Cloudera Navigator', 'Logs', 'Main', 'Monitoring', 'Performance', 'Ports and Addresses', 'Resource Management', 'Security', and 'Stacks Collection', all with counts of 0. The main configuration area displays the 'Default Kafka Version' as 'spark\_kafka\_version' with a dropdown menu showing '0.9', '0.10' (selected), and 'None'. There is also a 'Gateway Default Group' dropdown. A 'Show All Descriptions' link is visible. At the bottom right, there is a 'Save Changes (CTRL+S)' button.

Run this command in terminal

```
spark2-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.11:2.3.2 spark_kafka_to_local.py 18.211.252.152 9092 de-capstone3
```

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 -cat /user/ec2-user/capstone-  
project/warehouse/capstone/*
```

When the data is already there, we should stop the process gracefully.

Open another terminal and use the command below to stop gracefully:

```
ps -ef | grep spark | grep spark_kafka_to_local | awk '{print $2}'  
| xargs kill -SIGTERM
```

# Screenshots of the data

```
root@ip-10-0-0-174:~/capstoi x + v
[root@ip-10-0-0-174 capstone_project]# hadoop fs -ls /user/ec2-user/capstone_project/warehouse/capstone/
Found 2 items
drwxr-xr-x - root ec2-user 0 2022-04-19 20:37 /user/ec2-user/capstone_project/warehouse/capstone/_spark_metadata
-rw-r--r-- 3 root ec2-user 469605 2022-04-19 20:37 /user/ec2-user/capstone_project/warehouse/capstone/part-00000-654591dd-add4-4d38-8389-b814e89f1d63-c000.csv
[root@ip-10-0-0-174 capstone_project]# |
```

```
root@ip-10-0-0-174:~/capstoi x + v
91124640,4.1.1,Android,3.6346725,-174.664812,e7bc5fb2-1231-11eb-adc1-0242ac120002,fcba68aa-1231-11eb-adc1-0242ac120002,No,No,Yes,Yes,2020-03-05T15:51:05.000Z
22354331,2.4.25,Android,-67.422789,41.484604,e7bc5fb2-1231-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,No,Yes,No,Yes,2020-06-29T17:40:33.000Z
24561396,4.1.6,iOS,-40.7626955,-159.885739,de545711-3914-4450-8c11-b17b8dabb5e1,fcba68aa-1231-11eb-adc1-0242ac120002,No,No,Yes,Yes,2020-02-07T19:48:31.000Z
15596069,2.4.12,iOS,-72.1381845,-40.040386,e7bc5fb2-1231-11eb-adc1-0242ac120002,fcba68aa-1231-11eb-adc1-0242ac120002,No,Yes,No,Yes,2020-02-20T01:51:41.000Z
91676599,4.4.3,Android,-15.242505,-108.412877,b328829e-17ae-11eb-adc1-0242ac120002,fcba68aa-1231-11eb-adc1-0242ac120002,Yes,No,No,Yes,2020-07-24T12:23:01.000Z
97278161,1.2.4,Android,-52.458265,52.090323,de545711-3914-4450-8c11-b17b8dabb5e1,e1e99492-17ae-11eb-adc1-0242ac120002,No,No,Yes,Yes,2020-02-15T21:11:41.000Z
86307402,2.4.20,iOS,-27.603681,132.654253,e7bc5fb2-1231-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,Yes,Yes,No,No,2020-02-27T04:04:22.000Z
80445384,2.3.29,Android,-74.4209645,-171.230004,de545711-3914-4450-8c11-b17b8dabb5e1,e1e99492-17ae-11eb-adc1-0242ac120002,No,Yes,No,No,2020-03-09T06:06:38.000Z
23303502,2.1.33,Android,-72.527739,-77.178381,de545711-3914-4450-8c11-b17b8dabb5e1,e1e99492-17ae-11eb-adc1-0242ac120002,Yes,No,Yes,Yes,2020-09-17T21:43:14.000Z
81061650,3.3.1,iOS,32.234933,154.131783,de545711-3914-4450-8c11-b17b8dabb5e1,e1e99492-17ae-11eb-adc1-0242ac120002,Yes,No,Yes,No,2020-03-30T04:32:53.000Z
19413541,2.1.17,iOS,33.9260585,106.127469,e7bc5fb2-1231-11eb-adc1-0242ac120002,fcba68aa-1231-11eb-adc1-0242ac120002,No,No,Yes,Yes,2020-01-25T16:04:03.000Z
65047706,2.4.27,iOS,-77.108233,140.420846,b328829e-17ae-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,No,Yes,Yes,No,2020-04-25T16:12:37.000Z
92497528,3.2.31,iOS,-69.007902,46.259543,b328829e-17ae-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,No,Yes,Yes,No,2020-05-24T16:05:43.000Z
53978698,1.3.34,Android,0.224584,-8.528708,de545711-3914-4450-8c11-b17b8dabb5e1,e1e99492-17ae-11eb-adc1-0242ac120002,Yes,Yes,Yes,Yes,2020-09-26T12:26:34.000Z
22360184,3.1.11,iOS,-50.9895835,-15.754777,b328829e-17ae-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,Yes,Yes,No,Yes,2020-05-20T21:32:54.000Z
44585909,2.4.37,iOS,-23.0985865,-156.438962,b328829e-17ae-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,No,No,Yes,No,2020-02-07T00:57:01.000Z
54823847,4.3.2,iOS,-48.6597475,-63.722102,e7bc5fb2-1231-11eb-adc1-0242ac120002,fcba68aa-1231-11eb-adc1-0242ac120002,Yes,No,Yes,Yes,2020-03-08T14:03:15.000Z
21621758,2.2.22,iOS,-56.2635185,33.939368,e7bc5fb2-1231-11eb-adc1-0242ac120002,fcba68aa-1231-11eb-adc1-0242ac120002,No,Yes,No,Yes,2020-08-01T02:46:47.000Z
33090503,3.1.4,Android,-53.8831745,150.937637,e7bc5fb2-1231-11eb-adc1-0242ac120002,fcba68aa-1231-11eb-adc1-0242ac120002,No,No,No,Yes,2020-02-10T07:12:55.000Z
83436917,1.3.19,iOS,29.6107005,91.528804,de545711-3914-4450-8c11-b17b8dabb5e1,e1e99492-17ae-11eb-adc1-0242ac120002,Yes,No,No,Yes,2020-07-30T05:04:13.000Z
68796997,2.3.29,iOS,43.2997955,-42.645686,b328829e-17ae-11eb-adc1-0242ac120002,e1e99492-17ae-11eb-adc1-0242ac120002,Yes,Yes,Yes,Yes,2020-10-17T20:00:19.000Z
84392327,1.3.12,iOS,40.7522335,-105.599661,de545711-3914-4450-8c11-b17b8dabb5e1,fcba68aa-1231-11eb-adc1-0242ac120002,Yes,No,No,No,2020-01-07T12:48:35.000Z
39542434,2.1.11,iOS,-8.8883795,31.780290,e7bc5fb2-1231-11eb-adc1-0242ac120002,fcba68aa-1231-11eb-adc1-0242ac120002,Yes,Yes,Yes,No,2020-04-05T18:00:34.000Z
16813062,3.2.18,Android,-88.342371,-176.291730,b328829e-17ae-11eb-adc1-0242ac120002,e1e99492-17ae-11eb-adc1-0242ac120002,No,Yes,No,No,2020-05-05T23:08:24.000Z
13307480,4.3.36,iOS,26.6645385,109.842184,b328829e-17ae-11eb-adc1-0242ac120002,e1e99492-17ae-11eb-adc1-0242ac120002,No,Yes,Yes,Yes,2020-06-12T19:23:21.000Z
66537285,3.4.37,Android,-43.4804705,108.716577,de545711-3914-4450-8c11-b17b8dabb5e1,e1e99492-17ae-11eb-adc1-0242ac120002,Yes,Yes,No,Yes,2020-09-07T17:10:58.000Z
48434687,4.1.28,Android,-7.888734,-54.732169,b328829e-17ae-11eb-adc1-0242ac120002,a95dd57b-779f-49db-819d-b6960483e554,Yes,Yes,No,No,2020-02-18T19:48:48.000Z
98030364,3.3.35,iOS,59.4450305,-66.155279,e7bc5fb2-1231-11eb-adc1-0242ac120002,e1e99492-17ae-11eb-adc1-0242ac120002,Yes,Yes,No,Yes,2020-04-15T17:33:44.000Z
[root@ip-10-0-0-174 capstone_project]# |
```