

## Deliverable 2: Kafka Producer and Consumer

### Creating a Kafka topic

To create a Kafka topic, you can use the following command inside the Kafka container:

```
kafka-topics --create --topic transactions --partitions 1
→ --replication-factor 1 --bootstrap-server localhost:9092
```

Then, we need to connect the topic to HDFS. To do that, we need to create a connector. The connector is a configuration file that tells Kafka how to connect to HDFS.

```
curl -X POST -H "Content-Type: application/json" --data '{
  "name": "hdfs-sink-connector",
  "config": {
    "connector.class":
→   "io.confluent.connect.hdfs.HdfsSinkConnector",
    "tasks.max": "1",
    "topics": "transactions",
    "hdfs.url": "hdfs://namenode:9000",
    "flush.size": "10",
    "hdfs.authentication.kerberos": "false",
    "format.class":
→   "io.confluent.connect.hdfs.json.JsonFormat",
    "partitioner.class":
→   "io.confluent.connect.storage.partitionner.DefaultPartitioner",
    "rotate.interval.ms": "60000",
    "locale": "en",
    "timezone": "UTC",
    "value.converter.schemas.enable": "false"
  }
}' http://localhost:8083/connectors
```

### Consuming data from CSV file and sending it to Kafka

The code available in `src/prod-cons/deliverable/producer.py` reads a CSV with the standard csv library available in all python installations.

It saves all the registers in a list via the use of a `NamedTuple`, which is a lightweight object that can be used to store data.

Then, it iterates over all the registers, dumps them into a JSON format and sends them to the Kafka topic.

## Evidences

[illegible]

Figure 1: Sending data to kafka

## Browse Directory

Show

25

entries

Search:


<input type="checkbox"/>	 Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	<input type="checkbox"/>
<input type="checkbox"/>	-rw-r--r--	appuser	supergroup	29.6 KB	Mar 30 17:43	3	128 MB	transactions+0+0000000000+0000000009.json	<input type="checkbox"/>
<input type="checkbox"/>	-rw-r--r--	appuser	supergroup	27.16 KB	Mar 30 17:43	3	128 MB	transactions+0+0000000010+0000000019.json	<input type="checkbox"/>
<input type="checkbox"/>	-rw-r--r--	appuser	supergroup	35.09 KB	Mar 30 17:43	3	128 MB	transactions+0+0000000020+0000000029.json	<input type="checkbox"/>

Figure 2: First three entries of data in HDFS

## Challenges encountered

Overall the deliverable was quite easy and straight forward. It was basically replicating what we saw in the classes and adapting it to our needs.