Assignment 08 - Kafka Producer

Name: Kesav Adithya Venkidusamy

Course: DSC650 - Big Data

Instructor: Amirfarrokh Iranitalab

In [17]:
```python
## Loading required libraries for this assignment

import json
import uuid
import pandas as pd

from kafka import KafkaProducer, KafkaAdminClient
from kafka.admin.new_topic import NewTopic
from kafka.errors import TopicAlreadyExistsError

import warnings
warnings.filterwarnings("ignore")
```

Configuration Parameters

In [3]:
```python
## Defining the configuration

config = dict(
    bootstrap_servers=['kafka.kafka.svc.cluster.local:9092'],
    first_name='KesavAdithya',
    last_name='Venkidusamy'
)

config['client_id'] = '{}{}'.format(
    config['last_name'],
    config['first_name']
)
config['topic_prefix'] = '{}{}'.format(
    config['last_name'],
    config['first_name']
)
```

```
config
```

Out[3]: {'bootstrap_servers': ['kafka.kafka.svc.cluster.local:9092'],
        'first_name': 'KesavAdithya',
        'last_name': 'Venkidusamy',
        'client_id': 'VenkidusamyKesavAdithya',
        'topic_prefix': 'VenkidusamyKesavAdithya'}

Create Topic Utility Function

The `create_kafka_topic` helps create a Kafka topic based on your configuration settings. For instance, if your first name is *John* and your last name is *Doe*, `create_kafka_topic('locations')` will create a topic with the name `DoeJohn-locations`. The function will not create the topic if it already exists.

In [4]:
```python
def create_kafka_topic(topic_name, config=config, num_partitions=1, replication_factor=1):
    bootstrap_servers = config['bootstrap_servers']
    client_id = config['client_id']
    topic_prefix = config['topic_prefix']
    name = '{}-{}'.format(topic_prefix, topic_name)

    admin_client = KafkaAdminClient(
        bootstrap_servers=bootstrap_servers,
        client_id=client_id
    )

    topic = NewTopic(
        name=name,
        num_partitions=num_partitions,
        replication_factor=replication_factor
    )

    topic_list = [topic]
    try:
        admin_client.create_topics(new_topics=topic_list)
        print('Created topic "{}"'.format(name))
    except TopicAlreadyExistsError as e:
        print('Topic "{}" already exists'.format(name))

create_kafka_topic('locations')
```

```
Created topic "VenkidusamyKesavAdithya-locations"
```

```
In [5]:   ## Creating Topic for accelerations
          create_kafka_topic('accelerations')
```

```
Created topic "VenkidusamyKesavAdithya-accelerations"
```

Kafka Producer

The following code creates a `KafkaProducer` object which you can use to send Python objects that are serialized as JSON.

Note: This producer serializes Python objects as JSON. This means that object must be JSON serializable. As an example, Python `DateTime` values are not JSON serializable and must be converted to a string (e.g. ISO 8601) or a numeric value (e.g. a Unix timestamp) before being sent.

```
In [7]:   producer = KafkaProducer(
              bootstrap_servers=config['bootstrap_servers'],
              value_serializer=lambda x: json.dumps(x).encode('utf-8')
          )
```

Send Data Function

The `send_data` function sends a Python object to a Kafka topic. This function adds the `topic_prefix` to the topic so `send_data('locations', data)` sends a JSON serialized message to `DoeJohn-locations`. The function also registers callbacks to let you know if the message has been sent or if an error has occured.

```
In [8]:   def on_send_success(record_metadata):
              print('Message sent:\n    Topic: "{}"\n    Partition: {}\n    Offset: {}'.format(
                  record_metadata.topic,
                  record_metadata.partition,
                  record_metadata.offset
              ))

          def on_send_error(excp):
              print('I am an errback', exc_info=excp)


          def send_data(topic, data, config=config, producer=producer, msg_key=None):
              topic_prefix = config['topic_prefix']
              topic_name = '{}-{}'.format(topic_prefix, topic)

              if msg_key is not None:
                  key = msg_key
```

```python
    else:
        key = uuid.uuid4().hex

    producer.send(
        topic_name,
        value=data,
        key=key.encode('utf-8')
    ).add_callback(on_send_success).add_errback(on_send_error)
```

In [11]:
```python
## Load the source data inot dataframe
data_dir = '/home/jovyan/dsc650/data/processed/bdd/locations'
df = pd.read_parquet(data_dir)
```

In [12]:
```python
## Displaying few records in dataframe
df.head()
```

Out[12]:

	id	ride_id	uuid	timestamp	offset	cou
0	58682c5d48cad9d9e103431d773615bf	c9a2b46c9aa515b632eddc45c4868482	19b9aa10588646b3bf22c9b4865a7995	1970-01-01 00:25:03.882586	1.525061	299.619
1	85c61911b7fe2ced1000c33c9e932706	6760ffa3f41908695d1405b776c3e8d5	dad7eae44e784b549c8c5a3aa051a8c7	1970-01-01 00:25:07.320453	1.077913	158.203
2	58682c5d48cad9d9e103431d773615bf	c9a2b46c9aa515b632eddc45c4868482	19b9aa10588646b3bf22c9b4865a7995	1970-01-01 00:25:03.882583	4.525061	299.619
3	85c61911b7fe2ced1000c33c9e932706	6760ffa3f41908695d1405b776c3e8d5	dad7eae44e784b549c8c5a3aa051a8c7	1970-01-01 00:25:07.320449	5.077913	159.960
4	58682c5d48cad9d9e103431d773615bf	c9a2b46c9aa515b632eddc45c4868482	19b9aa10588646b3bf22c9b4865a7995	1970-01-01 00:25:03.882579	8.525061	299.619

In [13]:
```python
## Printing the column and datatypes
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 478 entries, 0 to 477
Data columns (total 14 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   id         327 non-null    object
 1   ride_id    478 non-null    object
 2   uuid       478 non-null    object
 3   timestamp  478 non-null    datetime64[ns]
 4   offset     478 non-null    float64
 5   course     478 non-null    float64
 6   latitude   478 non-null    float64
 7   longitude  478 non-null    float64
 8   geohash    478 non-null    object
 9   speed      478 non-null    float64
 10  accuracy   478 non-null    float64
 11  timelapse  478 non-null    bool
 12  filename   478 non-null    object
 13  t          478 non-null    category
dtypes: bool(1), category(1), datetime64[ns](1), float64(6), object(5)
memory usage: 47.4+ KB
```

In [18]:
```python
## Convert timestamp to str datatype and convert df to dictionary
df['timestamp'] = df['timestamp'].astype('str')
data_dict = df.set_index('t').transpose().to_dict()
```

In [19]:
```python
## Sending data to producer
for k,v in data_dict.items():
    send_data(topic='locations', data=v, config=config, producer=producer, msg_key=k)
```

```
Message sent:
    Topic: "VenkidusamyKesavAdithya-locations"
    Partition: 0
    Offset: 0
Message sent:
    Topic: "VenkidusamyKesavAdithya-locations"
    Partition: 0
    Offset: 1
Message sent:
    Topic: "VenkidusamyKesavAdithya-locations"
    Partition: 0
    Offset: 2
Message sent:
    Topic: "VenkidusamyKesavAdithya-locations"
    Partition: 0
    Offset: 3
```

```
Message sent:
    Topic: "VenkidusamyKesavAdithya-locations"
    Partition: 0
    Offset: 4
Message sent:
    Topic: "VenkidusamyKesavAdithya-locations"
    Partition: 0
    Offset: 5
Message sent:
    Topic: "VenkidusamyKesavAdithya-locations"
    Partition: 0
    Offset: 6
Message sent:
    Topic: "VenkidusamyKesavAdithya-locations"
    Partition: 0
    Offset: 7
Message sent:
    Topic: "VenkidusamyKesavAdithya-locations"
    Partition: 0
    Offset: 8
Message sent:
    Topic: "VenkidusamyKesavAdithya-locations"
    Partition: 0
    Offset: 9
Message sent:
    Topic: "VenkidusamyKesavAdithya-locations"
    Partition: 0
    Offset: 10
Message sent:
    Topic: "VenkidusamyKesavAdithya-locations"
    Partition: 0
    Offset: 11
Message sent:
    Topic: "VenkidusamyKesavAdithya-locations"
    Partition: 0
    Offset: 12
Message sent:
    Topic: "VenkidusamyKesavAdithya-locations"
    Partition: 0
    Offset: 13
Message sent:
    Topic: "VenkidusamyKesavAdithya-locations"
    Partition: 0
    Offset: 14
Message sent:
    Topic: "VenkidusamyKesavAdithya-locations"
    Partition: 0
    Offset: 15
Message sent:
```

```
        Topic: "VenkidusamyKesavAdithya-locations"
        Partition: 0
        Offset: 16
    Message sent:
        Topic: "VenkidusamyKesavAdithya-locations"
        Partition: 0
        Offset: 17
    Message sent:
        Topic: "VenkidusamyKesavAdithya-locations"
        Partition: 0
        Offset: 18
    Message sent:
        Topic: "VenkidusamyKesavAdithya-locations"
        Partition: 0
        Offset: 19
    Message sent:
        Topic: "VenkidusamyKesavAdithya-locations"
        Partition: 0
        Offset: 20
    Message sent:
        Topic: "VenkidusamyKesavAdithya-locations"
        Partition: 0
        Offset: 21
    Message sent:
        Topic: "VenkidusamyKesavAdithya-locations"
        Partition: 0
        Offset: 22
    Message sent:
        Topic: "VenkidusamyKesavAdithya-locations"
        Partition: 0
        Offset: 23
    Message sent:
        Topic: "VenkidusamyKesavAdithya-locations"
        Partition: 0
        Offset: 24
    Message sent:
        Topic: "VenkidusamyKesavAdithya-locations"
        Partition: 0
        Offset: 25
    Message sent:
        Topic: "VenkidusamyKesavAdithya-locations"
        Partition: 0
        Offset: 26
    Message sent:
        Topic: "VenkidusamyKesavAdithya-locations"
        Partition: 0
        Offset: 27
    Message sent:
        Topic: "VenkidusamyKesavAdithya-locations"
```

```
        Partition: 0
        Offset: 28
Message sent:
        Topic: "VenkidusamyKesavAdithya-locations"
        Partition: 0
        Offset: 29
Message sent:
        Topic: "VenkidusamyKesavAdithya-locations"
        Partition: 0
        Offset: 30
Message sent:
        Topic: "VenkidusamyKesavAdithya-locations"
        Partition: 0
        Offset: 31
Message sent:
        Topic: "VenkidusamyKesavAdithya-locations"
        Partition: 0
        Offset: 32
Message sent:
        Topic: "VenkidusamyKesavAdithya-locations"
        Partition: 0
        Offset: 33
```

In [20]:
```python
## Load the accelerations data inot dataframe
data_dir_acc = '/home/jovyan/dsc650/data/processed/bdd/accelerations'
acc_df = pd.read_parquet(data_dir_acc)
```

In [21]:
```python
## showing fre records from acc_df dataframe
acc_df.head()
```

Out[21]:

	id	ride_id	uuid	timestamp	offset	x
0	58682c5d48cad9d9e103431d773615bf	c9a2b46c9aa515b632eddc45c4868482	19b9aa10588646b3bf22c9b4865a7995	1970-01-01 00:25:03.882586	0.822061	-0.994
1	58682c5d48cad9d9e103431d773615bf	c9a2b46c9aa515b632eddc45c4868482	19b9aa10588646b3bf22c9b4865a7995	1970-01-01 00:25:03.882586	0.842061	-0.998
2	58682c5d48cad9d9e103431d773615bf	c9a2b46c9aa515b632eddc45c4868482	19b9aa10588646b3bf22c9b4865a7995	1970-01-01 00:25:03.882586	0.862061	-0.999

	id	ride_id	uuid	timestamp	offset	x
3	58682c5d48cad9d9e103431d773615bf	c9a2b46c9aa515b632eddc45c4868482	19b9aa10588646b3bf22c9b4865a7995	1970-01-01 00:25:03.882586	0.882061	-0.999
4	58682c5d48cad9d9e103431d773615bf	c9a2b46c9aa515b632eddc45c4868482	19b9aa10588646b3bf22c9b4865a7995	1970-01-01 00:25:03.882586	0.902061	-0.999

In [22]:
```python
## Displaying column and its datatypes
acc_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 23512 entries, 0 to 23511
Data columns (total 11 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   id         16056 non-null  object
 1   ride_id    23512 non-null  object
 2   uuid       23512 non-null  object
 3   timestamp  23512 non-null  datetime64[ns]
 4   offset     23512 non-null  float64
 5   x          23512 non-null  float64
 6   y          23512 non-null  float64
 7   z          23512 non-null  float64
 8   timelapse  23512 non-null  bool
 9   filename   23512 non-null  object
 10  t          23512 non-null  category
dtypes: bool(1), category(1), datetime64[ns](1), float64(4), object(4)
memory usage: 1.7+ MB
```

In [23]:
```python
## Convert timestamp to str datatype and convert acc_df to dictionary
acc_df['timestamp'] = acc_df['timestamp'].astype('str')
data_dict_acc = acc_df.set_index('t').transpose().to_dict()
```

In [24]:
```python
## Sending data to the topic
for k,v in data_dict_acc.items():
    send_data(topic='accelerations', data=v, config=config, producer=producer, msg_key=k)
```

```
Message sent:
    Topic: "VenkidusamyKesavAdithya-accelerations"
```

```
        Partition: 0
        Offset: 0
    Message sent:
        Topic: "VenkidusamyKesavAdithya-accelerations"
        Partition: 0
        Offset: 1
    Message sent:
        Topic: "VenkidusamyKesavAdithya-accelerations"
        Partition: 0
        Offset: 2
    Message sent:
        Topic: "VenkidusamyKesavAdithya-accelerations"
        Partition: 0
        Offset: 3
    Message sent:
        Topic: "VenkidusamyKesavAdithya-accelerations"
        Partition: 0
        Offset: 4
    Message sent:
        Topic: "VenkidusamyKesavAdithya-accelerations"
        Partition: 0
        Offset: 5
    Message sent:
        Topic: "VenkidusamyKesavAdithya-accelerations"
        Partition: 0
        Offset: 6
    Message sent:
        Topic: "VenkidusamyKesavAdithya-accelerations"
        Partition: 0
        Offset: 7
    Message sent:
        Topic: "VenkidusamyKesavAdithya-accelerations"
        Partition: 0
        Offset: 8
    Message sent:
        Topic: "VenkidusamyKesavAdithya-accelerations"
        Partition: 0
        Offset: 9
    Message sent:
        Topic: "VenkidusamyKesavAdithya-accelerations"
        Partition: 0
        Offset: 10
    Message sent:
        Topic: "VenkidusamyKesavAdithya-accelerations"
        Partition: 0
        Offset: 11
    Message sent:
        Topic: "VenkidusamyKesavAdithya-accelerations"
        Partition: 0
```

```
        Offset: 12
Message sent:
        Topic: "VenkidusamyKesavAdithya-accelerations"
        Partition: 0
        Offset: 13
Message sent:
        Topic: "VenkidusamyKesavAdithya-accelerations"
        Partition: 0
        Offset: 14
Message sent:
        Topic: "VenkidusamyKesavAdithya-accelerations"
        Partition: 0
        Offset: 15
Message sent:
        Topic: "VenkidusamyKesavAdithya-accelerations"
        Partition: 0
        Offset: 16
Message sent:
        Topic: "VenkidusamyKesavAdithya-accelerations"
        Partition: 0
        Offset: 17
Message sent:
        Topic: "VenkidusamyKesavAdithya-accelerations"
        Partition: 0
        Offset: 18
Message sent:
        Topic: "VenkidusamyKesavAdithya-accelerations"
        Partition: 0
        Offset: 19
Message sent:
        Topic: "VenkidusamyKesavAdithya-accelerations"
        Partition: 0
        Offset: 20
Message sent:
        Topic: "VenkidusamyKesavAdithya-accelerations"
        Partition: 0
        Offset: 21
Message sent:
        Topic: "VenkidusamyKesavAdithya-accelerations"
        Partition: 0
        Offset: 22
Message sent:
        Topic: "VenkidusamyKesavAdithya-accelerations"
        Partition: 0
        Offset: 23
Message sent:
        Topic: "VenkidusamyKesavAdithya-accelerations"
        Partition: 0
        Offset: 24
```

```
Message sent:
     Topic: "VenkidusamyKesavAdithya-accelerations"
     Partition: 0
     Offset: 25
Message sent:
     Topic: "VenkidusamyKesavAdithya-accelerations"
     Partition: 0
     Offset: 26
Message sent:
     Topic: "VenkidusamyKesavAdithya-accelerations"
     Partition: 0
     Offset: 27
Message sent:
     Topic: "VenkidusamyKesavAdithya-accelerations"
     Partition: 0
     Offset: 28
Message sent:
     Topic: "VenkidusamyKesavAdithya-accelerations"
     Partition: 0
     Offset: 29
Message sent:
     Topic: "VenkidusamyKesavAdithya-accelerations"
     Partition: 0
     Offset: 30
Message sent:
     Topic: "VenkidusamyKesavAdithya-accelerations"
     Partition: 0
     Offset: 31
Message sent:
     Topic: "VenkidusamyKesavAdithya-accelerations"
     Partition: 0
     Offset: 32
Message sent:
     Topic: "VenkidusamyKesavAdithya-accelerations"
     Partition: 0
     Offset: 33
```

In [ ]: