Assignment 08 - Kafka Producer

Author: Anjani Bonda

Date: 5/6/2023

In [19]:
```
# Load necessary modules/libraries.

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
```

Configuration Parameters

Updated with appropriate values

In [20]:
```
config = dict(
    bootstrap_servers=['kafka.kafka.svc.cluster.local:9092'],
    first_name='Anjani',
    last_name='Bonda'
)

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

config
```

Out[20]:
```
{'bootstrap_servers': ['kafka.kafka.svc.cluster.local:9092'],
 'first_name': 'Anjani',
 'last_name': 'Bonda',
 'client_id': 'BondaAnjani',
 'topic_prefix': 'BondaAnjani'}
```

Create Topic Utility Function

In [21]:
```
def create_kafka_topic(topic_name, config=config, num_partitions=1, replicatior
    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 topic for locations.
create_kafka_topic('locations')
```

```
Topic "BondaAnjani-locations" already exists
```

In [22]:
```
# Create topic for accelerations.
create_kafka_topic('accelerations')
```

```
Topic "BondaAnjani-accelerations" already exists
```

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 [23]:
```
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 [24]:
```
def on_send_success(record_metadata):
    print('Message sent:\n    Topic: "{}"\n    Partition: {}\n    Offset: {}'.f
        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
    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 [25]:
```
# Load 'locations'data.
locations_data_dir = '/home/jovyan/dsc650/data/processed/bdd/locations'
locations_df = pd.read_parquet(locations_data_dir)
```

In [26]:
```
# Check columns and datatypes for locations_df since few datatypes are not jso
locations_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: 47.2+ KB
```

In [27]:
```
# Change datatype timestamp to string and dataframe to dictionary for sending c
locations_df['timestamp'] = locations_df['timestamp'].astype('str')
data_dict = locations_df.set_index('t').transpose().to_dict()
```

```
/tmp/ipykernel_281/1817710994.py:3: UserWarning: DataFrame columns are not uni
que, some columns will be omitted.
  data_dict = locations_df.set_index('t').transpose().to_dict()
```

In [28]:
```
for key,value in data_dict.items():
    send_data(topic='locations', data=value, config=config, producer=producer,
```

```
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 34
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 35
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 36
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 37
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 38
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 39
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 40
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 41
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 42
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 43
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 44
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 45
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 46
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 47
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 48
```

```
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 49
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 50
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 51
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 52
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 53
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 54
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 55
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 56
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 57
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 58
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 59
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 60
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 61
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 62
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 63
```

```
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 64
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 65
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 66
Message sent:
    Topic: "BondaAnjani-locations"
    Partition: 0
    Offset: 67
```

In [29]:
```
# Load 'accelerations' data.
accelerations_data_dir = '/home/jovyan/dsc650/data/processed/bdd/accelerations'
accelerations_df = pd.read_parquet(accelerations_data_dir)
```

In [30]:
```
# Check columns and datatypes for accelerations_df as few datatypes are not jso
accelerations_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 [31]:
```
# Change datatype timestamp to string and dataframe to dictionary for sending d
accelerations_df['timestamp'] = accelerations_df['timestamp'].astype('str')
data_dict = accelerations_df.set_index('t').transpose().to_dict()
```

```
/tmp/ipykernel_281/1633942791.py:3: UserWarning: DataFrame columns are not uni
que, some columns will be omitted.
  data_dict = accelerations_df.set_index('t').transpose().to_dict()
```

In [32]:
```
for key,value in data_dict.items():
    send_data(topic='accelerations', data=value, config=config, producer=produc
```

```
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 34
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 35
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 36
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 37
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 38
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 39
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 40
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 41
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 42
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 43
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 44
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 45
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 46
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 47
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 48
```

```
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 49
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 50
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 51
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 52
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 53
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 54
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 55
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 56
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 57
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 58
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 59
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 60
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 61
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 62
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 63
```

```
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 64
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 65
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 66
Message sent:
    Topic: "BondaAnjani-accelerations"
    Partition: 0
    Offset: 67
```

In [ ]: