

# assignment08\_SafsafiAchraf

May 16, 2021

Achraf Safsafi

DSC650

Assignment 8

```
[1]: import json
import uuid
import time
from kafka import KafkaProducer, KafkaAdminClient, KafkaConsumer
from kafka.admin.new_topic import NewTopic
from kafka.errors import TopicAlreadyExistsError
```

```
[2]: import pandas as pd
import glob

files = glob.glob('dsc650/data/processed/bdd')
df = pd.concat([pd.read_parquet(fp) for fp in files])
df['timestamp'] = df['timestamp'].apply(lambda x: pd.Timestamp(x).
    ↳strftime('%Y-%m-%d %H:%M:%S'))
df.head()
```

```
[2]:
```

	id	ride_id	\
0	58682c5d48cad9d9e103431d773615bf	c9a2b46c9aa515b632eddc45c4868482	
1	58682c5d48cad9d9e103431d773615bf	c9a2b46c9aa515b632eddc45c4868482	
2	58682c5d48cad9d9e103431d773615bf	c9a2b46c9aa515b632eddc45c4868482	
3	58682c5d48cad9d9e103431d773615bf	c9a2b46c9aa515b632eddc45c4868482	
4	58682c5d48cad9d9e103431d773615bf	c9a2b46c9aa515b632eddc45c4868482	

  

	uuid	timestamp	offset	x	\
0	19b9aa10588646b3bf22c9b4865a7995	1970-01-01 00:25:03	0.822061	-0.994	
1	19b9aa10588646b3bf22c9b4865a7995	1970-01-01 00:25:03	0.842061	-0.998	
2	19b9aa10588646b3bf22c9b4865a7995	1970-01-01 00:25:03	0.862061	-0.999	
3	19b9aa10588646b3bf22c9b4865a7995	1970-01-01 00:25:03	0.882061	-0.999	
4	19b9aa10588646b3bf22c9b4865a7995	1970-01-01 00:25:03	0.902061	-0.999	

	y	z	timelapse	filename	t
0	0.045	-0.036	False	e2f795a7-6a7d-4500-b5d7-4569de996811.mov	000.0
1	0.046	-0.040	False	e2f795a7-6a7d-4500-b5d7-4569de996811.mov	000.0
2	0.047	-0.036	False	e2f795a7-6a7d-4500-b5d7-4569de996811.mov	000.0
3	0.045	-0.034	False	e2f795a7-6a7d-4500-b5d7-4569de996811.mov	000.0
4	0.048	-0.033	False	e2f795a7-6a7d-4500-b5d7-4569de996811.mov	000.0

```
[3]: data = df.head(1).to_json()
```

### 0.0.1 Configuration Parameters

```
[4]: config = dict(
    bootstrap_servers=['kafka.kafka.svc.cluster.local:9092'],
    first_name='Achraf',
    last_name='Safsafi'
)

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

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

config
```

```
[4]: {'bootstrap_servers': ['kafka.kafka.svc.cluster.local:9092'],
      'first_name': 'Achraf',
      'last_name': 'Safsafi',
      'client_id': 'SafsafiAchraf',
      'topic_prefix': 'SafsafiAchraf'}
```

### 0.0.2 Create Topic Utility Function

```
[5]: 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')

```

Topic "SafsafiAchraf-locations" already exists

### 0.0.3 Kafka Producer

```

[6]: producer = KafkaProducer(bootstrap_servers='kafka.kafka.svc.cluster.local:9092',
                               value_serializer=lambda v: json.dumps(v).
                               ↪encode('utf-8'))

producer.send('locations',data)
time.sleep(1)
producer.flush(100)

print("Message Sent to locations")

```

Message Sent to locations

### 0.0.4 Kafka Consumer

```

[ ]: consumer = KafkaConsumer('locations',
                               bootstrap_servers=['kafka.kafka.svc.cluster.local:
                               ↪9092'],
                               value_deserializer=lambda m: json.loads(m.
                               ↪decode('utf-8')))

for message in consumer:
    print(message)

consumer.close()

```

ConsumerRecord(topic='locations', partition=0, offset=51518143,

```
timestamp=1621201682882, timestamp_type=0, key=None, value='{"id":{"0":"58682c5d48cad9d9e103431d773615bf"},"ride_id":{"0":"c9a2b46c9aa515b632eddc45c4868482"},"uid":{"0":"19b9aa10588646b3bf22c9b4865a7995"},"timestamp":{"0":"1970-01-01 00:25:03"},"offset":{"0":0.8220608865},"x":{"0":-0.994},"y":{"0":0.045},"z":{"0":-0.036},"timelapse":{"0":false},"filename":{"0":"e2f795a7-6a7d-4500-b5d7-4569de996811.mov"},"t":{"0":"000.0"}}', headers=[], checksum=None, serialized_key_size=-1, serialized_value_size=424, serialized_header_size=-1)
```