**Kafka Assignment -**

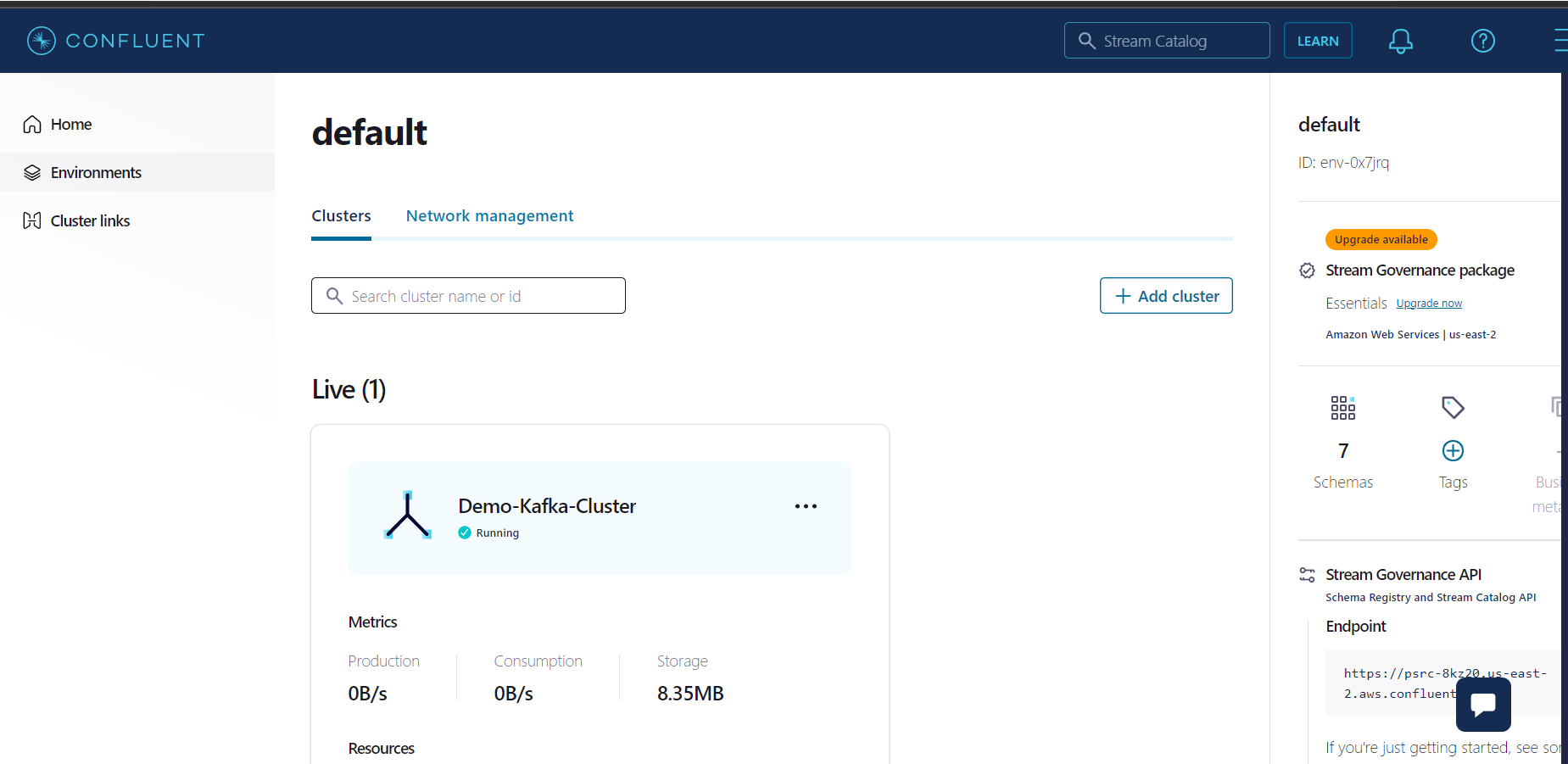
# download restaurent data from below mentioned link

Download Data Link -> <https://github.com/shashank-mishra219/Confluent-Kafka-Setup/blob/main/restaurant_orders.csv> --Done

Complete the given below task to finish this assignment.

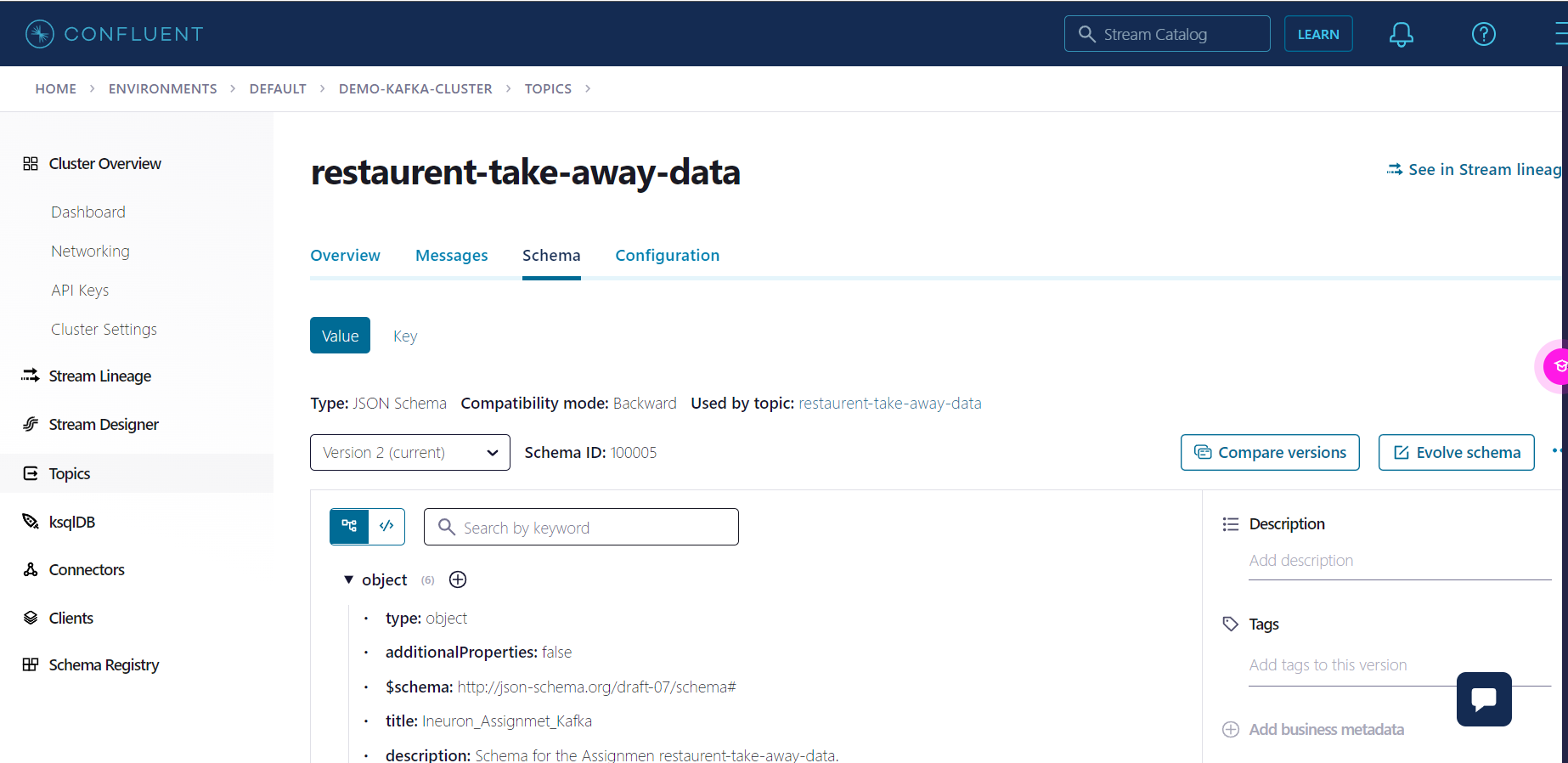
1.Setup Confluent Kafka Account –

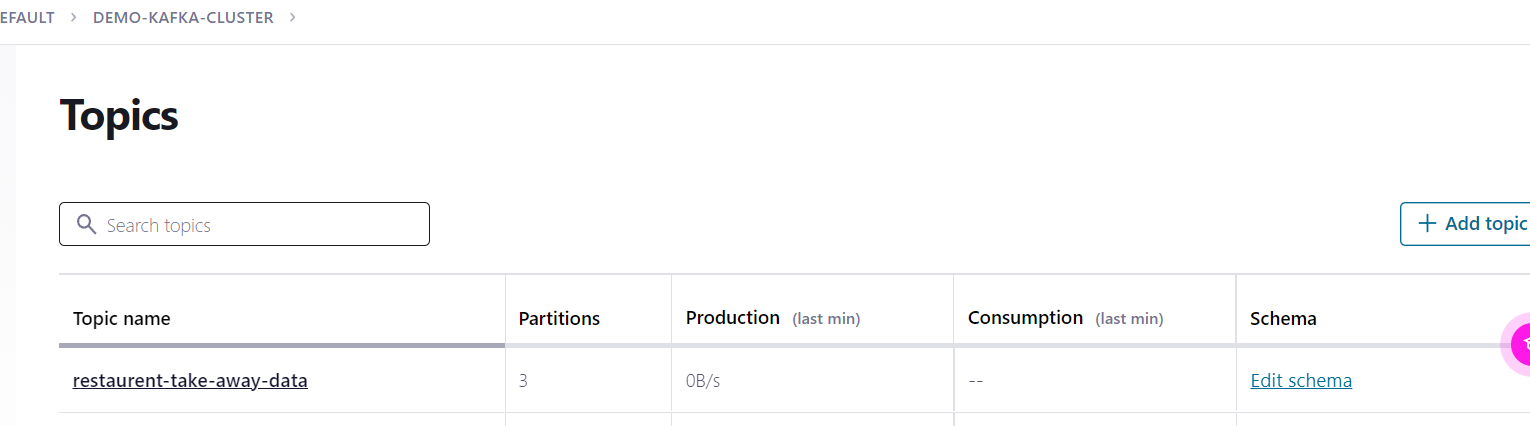
Action -



2. Create one kafka topic named as "restaurent-take-away-data" with 3 partitions

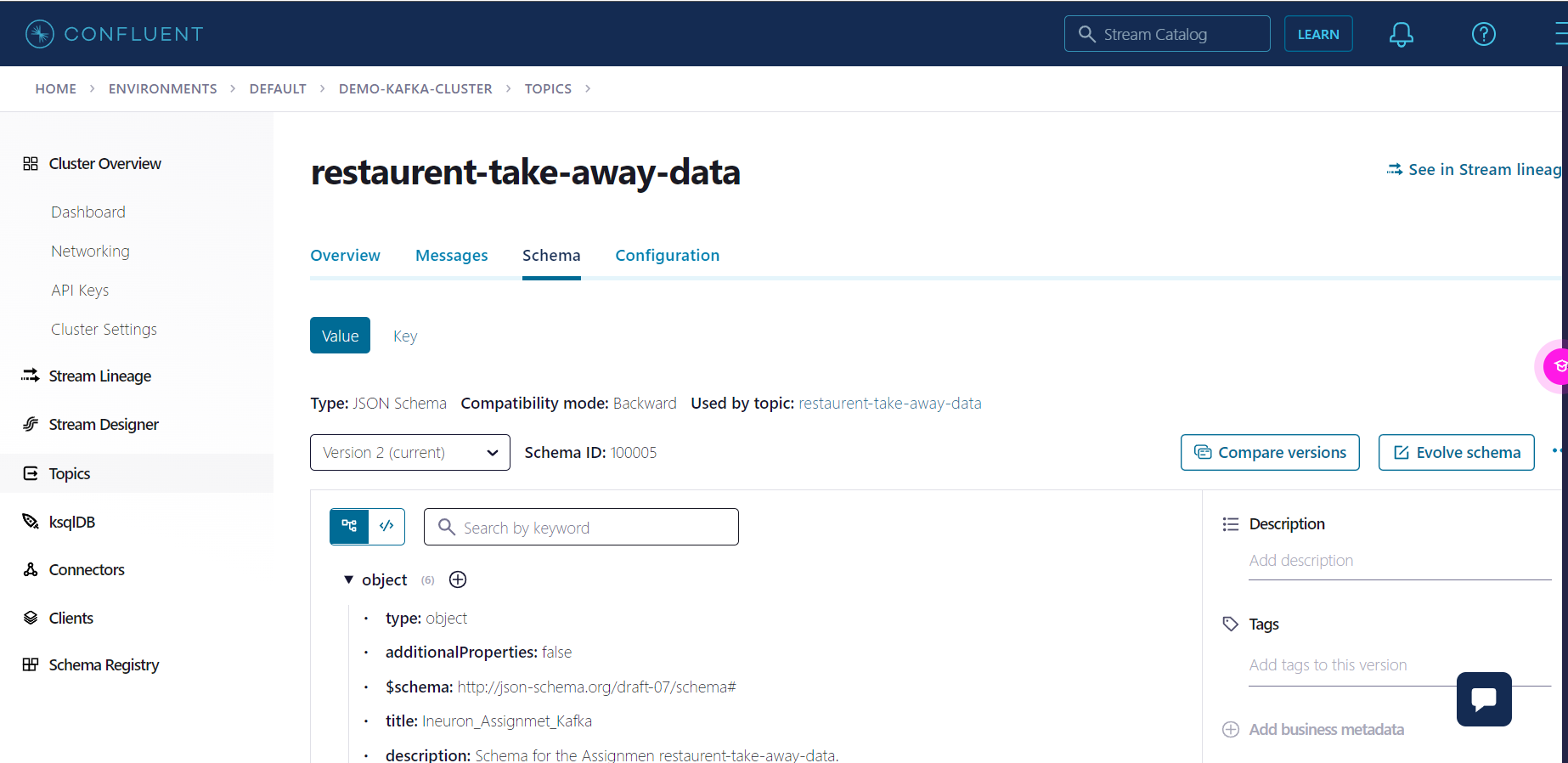
Action -

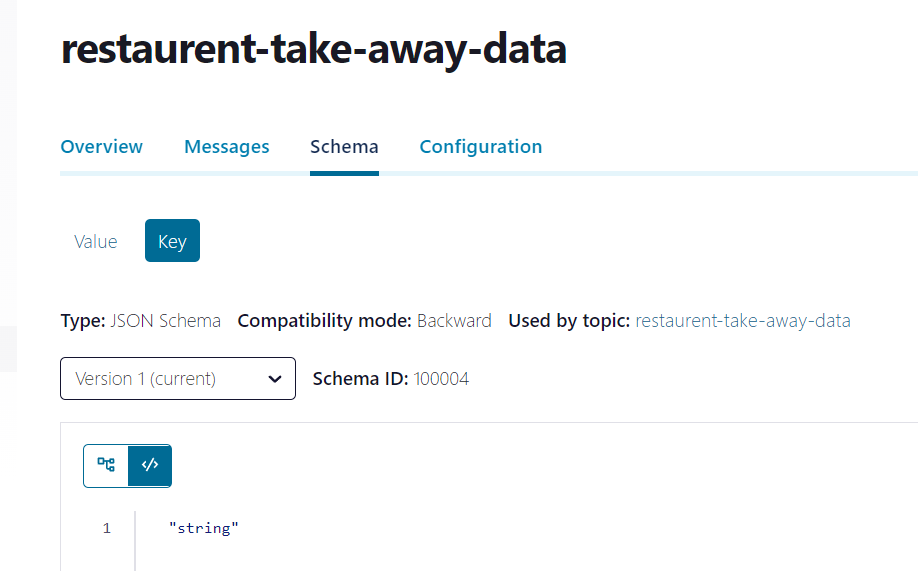




3. Setup key (string) & value (json) schema in the confluent schema registry

Action -





4. Write a kafka producer program (python or any other language) to read data records from restaurent data csv file, make sure schema is not hardcoded in the producer code, read the latest version of schema and schema\_str from schema registry and use it for data serialization.

Action –

Code is writer and placed over path (Below is precisely piece of code requested) -

schema\_str=schema\_registry\_client.get\_latest\_version("restaurent-take-away-data-value").schema.schema\_str

5. From producer code, publish data in Kafka Topic one by one and use dynamic key while publishing the records into the Kafka Topic

Action –

Code is writer and placed over path

<https://github.com/Mayurthele/Kafka_Assignment/blob/main/kafka_json_producer.py>

(Below is precisely piece of code requested) -

producer.produce(topic=topic,  
 key=string\_serializer(str(uuid4()), restaurent\_to\_dict), # MUR - str(uuid4() USED FOR creting the key , you can use name as well like 'mayur' or 'xyz'  
 value=json\_serializer(restaurent, SerializationContext(topic, MessageField.VALUE)),  
 on\_delivery=delivery\_report)

6. Write kafka consumer code and create two copies of same consumer code and save it with different names (kafka\_consumer\_1.py & kafka\_consumer\_2.py), again make sure lates schema version and schema\_str is not hardcoded in the consumer code, read it automatically from the schema registry to desrialize the data.

Now test two scenarios with your consumer code:

1. Use "group.id" property in consumer config for both consumers and mention different group\_ids in kafka\_consumer\_1.py & kafka\_consumer\_2.py, apply "earliest" offset property in both consumers and run these two consumers from two different terminals. Calculate how many records each consumer consumed and printed on the terminal

Action –

Code is writer and placed over path

https://github.com/Mayurthele/Kafka\_Assignment/blob/main/ [kafka\_json\_consumer\_1.py](https://github.com/Mayurthele/Kafka_Assignment/blob/main/kafka_json_consumer_1.py) https://github.com/Mayurthele/Kafka\_Assignment/blob/main/ [kafka\_json\_consumer\_2.py](https://github.com/Mayurthele/Kafka_Assignment/blob/main/kafka_json_consumer_1.py)

(Below is precisely piece of code requested) -

**Observation -** As the group id are same the output count was same for each kafka\_consumer\_1.py & kafka\_consumer\_2.py

consumer\_conf.update({  
 'group.id': 'group1',  
 'auto.offset.reset': "earliest"})

consumer\_conf.update({  
 'group.id': 'group2',  
 'auto.offset.reset': "earliest"})

1. Use "group.id" property in consumer config for both consumers and mention same group\_ids in kafka\_consumer\_1.py & kafka\_consumer\_2.py, apply "earliest" offset property in both consumers and run these two consumers from two different terminals. Calculate how many records each consumer consumed and printed on the terminal

Action –

Code is writer and placed over path

https://github.com/Mayurthele/Kafka\_Assignment/blob/main/ [kafka\_json\_consumer\_1.py](https://github.com/Mayurthele/Kafka_Assignment/blob/main/kafka_json_consumer_1.py)

https://github.com/Mayurthele/Kafka\_Assignment/blob/main/ [kafka\_json\_consumer\_2.py](https://github.com/Mayurthele/Kafka_Assignment/blob/main/kafka_json_consumer_1.py)

(Below is precisely piece of code requested) -

**Observation -** As the group id are same the output count was different for each kafka\_consumer\_1.py & kafka\_consumer\_2.py

consumer\_conf.update({  
 'group.id': 'group1',  
 'auto.offset.reset': "earliest"})

consumer\_conf.update({  
 'group.id': 'group1',  
 'auto.offset.reset': "earliest"})

7. Once above questions are done, write another kafka consumer to read data from kafka topic and from the consumer code create one csv file "output.csv" and append consumed records output.csv file.

Action –

Code is writer and placed over path - https://github.com/Mayurthele/Kafka\_Assignment/blob/main/kafka\_json\_consumer.py (Below is precisely piece of code requested) -

myDict = []  
if restaurent is not None:  
 print("User record {}: restaurent: {}\n"  
 .format(msg.key(), restaurent))  
 myDict.append(restaurent.\_\_dict\_\_)  
 print('list',myDict)  
  
FieldName=[]  
for m in myDict[0]:  
 FieldName.append(m)  
  
myFile = open('Streaming\_Data.csv','w+')  
writer = csv.DictWriter(myFile, fieldnames=list(FieldName))  
writer.writeheader()  
for i in myDict:  
 writer.writerow(i)  
myFile.close()