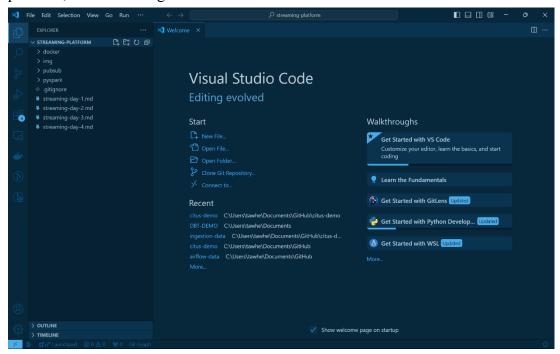
Stream Processing Task

1. Pertama lakukan clone repository Immersive-DataEngineer-Resource / streaming-platform, lalu buka dengan visual studio code.



2. Kemudian buka aplikasi Ubuntu pada windows.



3. Kemudian masukkan secure shell "ssh <u>raja_rahmanakmaludin@34.101.224.54</u>" dan password "mentoralterra2024".

```
fiqar_de4@IbnuEmil:~$ ssh raja_rahmanakmaludin@34.101.224.54
raja_rahmanakmaludin@34.101.224.54's password:
Linux instance-20240714-035051 6.1.0-23-cloud-amd64 #1 SMP PREEMPT_DYNAMIC Debian 6.1.99-1 (2024-07-15) x86_64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Thu Aug 1 12:03:38 2024 from 203.175.125.135
```

Berikut ini tampilan setelah melakukan langkah nomor 3.

4. Kemudian lakukan perintah "ls" untuk melihat apa saja folder / file yang ada.

```
raja_rahmanakmaludin@instance-20240714-035051:~$ ls airflow-data dbt-demo extract-load-demo ingestion-data streaming-platform Berikut ini folder yang tersedia.
```

5. Lalu masuk ke folder "streaming-platform" dengan menjalankan perintah "cd streaming-platform"

```
raja_rahmanakmaludin@instance-20240714-035051:~$ cd streaming-platform/raja_rahmanakmaludin@instance-20240714-035051:~/streaming-platform$ |
```

Sekarang posisinya sudah masuk ke folder "streaming-platform"

6. Di dalam folder tersebut, jalankan perintah "ls" lagi untuk melihat apa saja folder / file yang ada. Kemudian jika ada folder "docker", masuk ke folder tersebut dengan menjalankan perintah "cd docker". Di dalam folder tersebut jalankan perintah "ls" lagi untuk melihat folder / file yan tersedia. Jika ada folder redpanda, maka masuk ke folder tersebut dengan menjalankan perintah "cd redpanda". Setelah itu di dalam folder "redpanda" lakukan perintah "ls" untuk melihat file apa yang ada di folder tersebut. Berikut ini adalah gambaran pada server saat melakukan langkah – langkah yang sudah dijelaskan sebelumnya.

```
raja_rahmanakmaludin@instance-20240714-035051:~/streaming-platform$ ls
docker img pubsub pyspark streaming-day-1.md streaming-day-2.md streaming-day-3.md streaming-day-4.md
raja_rahmanakmaludin@instance-20240714-035051:~/streaming-platform$ cd docker/
raja_rahmanakmaludin@instance-20240714-035051:~/streaming-platform/docker$ ls
redpanda spark
raja_rahmanakmaludin@instance-20240714-035051:~/streaming-platform/docker$ cd redpanda/
raja_rahmanakmaludin@instance-20240714-035051:~/streaming-platform/docker/redpanda$ ls
docker-compose-spark.yml docker-compose.yml
raja_rahmanakmaludin@instance-20240714-035051:~/streaming-platform/docker/redpanda$ |
```

7. Pada folder tersebut jalankan docker dengan perintah "sudo docker compose -f docker-compose.yml up - d".

```
raja_rahmanakmaludin@instance-20240714-035051:~/streaming-platform/docker/redpanda$ sudo docker compose -f docker-compose.yml up -d WARN[0000] /home/raja_rahmanakmaludin/streaming-platform/docker/redpanda/docker-compose.yml: `version` is obsolete

[+] Running 7/7

*Network redpanda_redpanda_network

*Container ksqldb-server

*Started

*Container connect

*Container postgres

*Container postgres

*Container redpanda

*Container redpanda

*Container ksqldb-cli

*Container ksqldb-cli

*Container console

*Started

*Container console

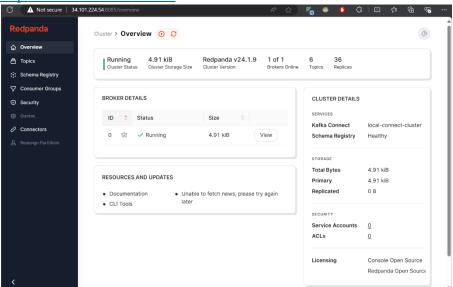
*Started

*Container console

*Container console
```

Berikut ini adalah tampilan saat perintah sudo docker compose -f docker-compose.yml up - d" berhasil dijalankan.

8. Setelah itu masuk ke website redpanda dengan memasukkan URL http://34.101.224.54:8085/.



9. Lalu kembali ke aplikasi visual studio code dan jalankan environment dengan perintah "python -m venv .venv" lalu aktifkan juga environment tersebut dengan perintah "source .venv/Scripts/activate.



10. Kemudian install "confluent_kafka" dengan menjalankan perintah "pip install confluent kafka".

```
tawhe@Ibnu_Emil MINGW64 ~/Documents/streaming-platform/pubsub/json (main)

$ pip install confluent_kafka
Collecting confluent_kafka
Downloading confluent_kafka-2.5.0-cp312-cp312-win_amd64.whl.metadata (2.4 kB)
Downloading confluent_kafka-2.5.0-cp312-cp312-win_amd64.whl (3.5 MB)

3.5/3.5 MB 848.1 kB/s eta 0:00:00

Installing collected packages: confluent_kafka
Successfully installed confluent_kafka-2.5.0

[notice] A new release of pip is available: 24.0 -> 24.2
[notice] To update, run: python.exe -m pip install --upgrade pip
(.venv)
tawhe@Ibnu_Emil MINGW64 ~/Documents/streaming-platform/pubsub/json (main)
```

Berikut ini tampilan setelah install "confluent kafka".

11. Pada folder "pubsub\json\produce.py" buatlah script seperti berikut ini :

```
from confluent_kafka import Producer
from datetime import datetime
import random
import json
TOPIC="fiqarrachman_stock_json_topic"
         'bootstrap.servers': '34.101.224.54:19092', # Assuming you're running this on the same machine as the compo
        'client.id': 'python-producer'
    countdata = 0
                 'event_time': datetime.now().isoformat(),
'ticker': random.choice(['AAPL', 'AMZN', 'MSFT', 'INTC', 'TBV']),
'price': round(random.random() * 100, 2)
            p.produce(TOPIC, key=str(uuid.uuid4), value=json.dumps(stock), callback=delivery_report)
            countdata+=1
            if countdata ==2000:
                 break;
        print(str(e))
    # Wait for any outstanding messages to be delivered
def delivery_report(err, msg):
    if err is not None:
        print('Message delivery failed: {}'.format(err))
        print('Message delivered to {} [{}]'.format(msg.topic(), msg.partition()))
    produce()
if __name__ == "__main__":
```

12. Kemudian pada folder "pubsub\json" eksekusi script yang sudah dibuat dengan menjalankan perintah "python produce.py".

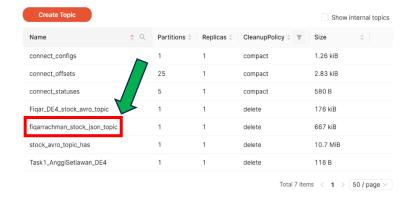
Berikut ini adalah tampilan bahwa perintah menjalankan file dengan nama "produce.py" sudah berhasil.

13. Setelah itu, buat script seperti berikut ini pada file dengan nama "consume.py".

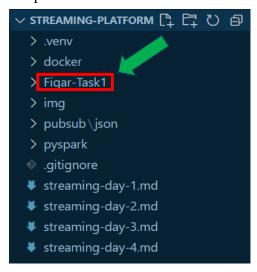
14. Lalu coba jalankan juga file dengan nama "consume.py" dengan menjalankan perintah "python consume.py"

```
(.venv)
  tawhe@Ibnu_Emil MINGW64 ~/Documents/streaming-platform/pubsub/json (main)
  $ python consume.py
```

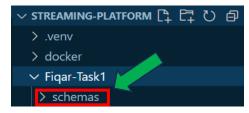
15. Kemudian kita cek apakah topic sudah berhasil dibuat dalam website redpanda.



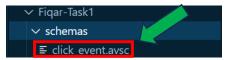
16. Setelah itu untuk tugas avro, sebelum membuat file ekstensi python, perlu membuat terlebih dahulu folder yang diinginkan. Dalam hal ini membuat folder dengan nama "Fiqar-Task1".



17. Setelah membuat folder, buat lagi di dalam folder "Fiqar-Task1" dengan nama folder "schemas".



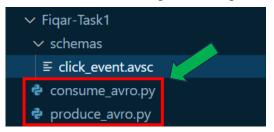
18. Di dalam folder "schemas" buat file dengan nama "click event.avsc".



19. Buat script seperti berikut ini pada file dengan nama "click event.avsc" tersebut.

Berikut ini adalah tampilan dari isi file "click event.avsc"

20. Setelah itu buat file dengan nama "produce avro.py" dan "consume avro.py".



21. Pada file "produce avro.py" buat script seperti berikut ini :

```
Fiqar-Task1 > 🍦 produce_avro.py > 😚 produce
       from uuid import uuid4
       from confluent_kafka import KafkaException
      from confluent_kafka.avro import AvroProducer
       from confluent_kafka import avro
       def delivery_callback(error, message):
            if error:
            else:
                 print(f"Message with the key {message.key()} has been produced to the topic {message.topic()}")
            key_schema_string = """
            {"type": "string"}
           key_schema = avro.loads(key_schema_string)
            value_schema = avro.load('/Users/tawhe/Documents/streaming-platform/Fiqar-Task1/schemas/click_event.avsc')
            return key_schema, value_schema
            config = {
                 'schema.registry.url': "http://34.101.224.54:18081"
            key_schema, value_schema = load_avro_schema_from_file()
            producer = AvroProducer(
                config,
                 default_key_schema=key_schema,
                 default_value_schema=value_schema
            event_types = ["CLICK", "PAGE_VIEW", "PURCHASE", "USER_REGISTRATION", "LOGIN", "LOGOUT"]
            countdata = 0
                 while countdata < 2000:
                      event_type = event_types[countdata % len(event_types)]
                          "user_id": 2,

"event_type": event_type,
                          "ts": "2023-12-12",

"page_url": "http://example.com" if event_type in ["PAGE_VIEW", "CLICK"] else None,

"product_id": "prod123" if event_type == "PURCHASE" else None,

"if weat type == "ILSEP_PEGISTRATION" else None,
                          "user_email": "user@example.com" if event_type == "USER_REGISTRATION" else None,
"location": "New York" if event_type in ["LOGIN", "LOGOUT"] else None,
"device_type": "mobile" if event_type in ["CLICK", "PAGE_VIEW"] else None
```

```
producer.produce(
topic="Fiqar_DE4_stock_avro_topic",
key=key,
value=value,
on_delivery=delivery_callback
)
countdata += 1

producer.poll(10000)
producer.flush()

except KafkaException as e:
print(f"Error occurred during message production: {e}")

print("Done!")

def main():
produce()

if __name__ == "__main__":
main()
```

Gambar diatas merupakan isi dari file "produce_avro.py".

22. Setelah itu buat juga script untuk file "consume_avro.py".

```
from confluent_kafka import KafkaError, KafkaException
from confluent_kafka.avro import AvroConsumer
from confluent_kafka.avro.serializer import SerializerError
    config = {
        'bootstrap.servers': '34.101.224.54:19092',
        'group.id': 'Fiqar_DE4_consumer_group',
'schema.registry.url': 'http://34.101.224.54:18081',
    consumer = AvroConsumer(config)
    consumer.subscribe(['Fiqar_DE4_stock_avro_topic'])
        while True:
                message = consumer.poll(1.0)
                if message is None:
                if message.error():
                     if message.error().code() == KafkaError._PARTITION_EOF:
                        print(f"Reached end at {message.topic()} [{message.partition()}] offset {message.offset()}")
                     else:
                         raise KafkaException(message.error())
                else:
                     value = message.value()
                     event_type = value.get("event_type")
                     if event_type == "CLICK":
                     elif event_type == "PAGE_VIEW":
                     elif event_type == "USER_REGISTRATION":
                         pass
```

```
elif event_type == "LOGIN":

# Handle LOGIN event
pass
elif event_type == "LOGOUT":

# Handle LOGOUT event
pass

print(f"Consumed message from topic {message.topic()}, partition {message.partition()}, offset {mi_
print(f"Key: {message.key()}, Value: {value}")

except SerializerError as e:
print(f"Message deserialization failed: {e}")
continue

except KeyboardInterrupt:
pass
finally:
consumer.close()
print("Consumer closed")

def main():
consume()

if __name__ == "__main__":
main()
```

Gambar diatas merupakan isi dari file "consume avro.py"

23. Setelah file "produce_avro.py" dan "consume_avro.py" sudah dibuat. Jalankan file tersebut dengan perintah "python produce_avro.py" untuk file "produce_avro.py". Dan perintah "python consume_avro.py" untuk file "consume_avro.py".



Consumed message from topic Fiqar_DE4_stock_avro_topic, partition 0, offset 1994

Key: 68943b56-53cd-4aae-a724-0af8aela388b, Value: {'user_id': 2, 'event_type': 'PURCHASE', 'ts': '2023-12-12', 'page_url': None, 'product_id': 'prod123', 'user_email': None, 'location': None, 'device_type': None}

Consumed message from topic Fiqar_DE4_stock_avro_topic, partition 0, offset 1994

Key: 68943b56-53cd-4aae-a724-0af8aela388b, Value: {'user_id': 2, 'event_type': 'None}

Consumed message from topic Fiqar_DE4_stock_avro_topic, partition 0, offset 1995

Key: 77e44bdd-1849-4ced-ac5b-fe684a190be0, Value: {'user_id': 2, 'event_type': 'USER_REGISTRATION', 'ts': '2023-12-12', 'page_url': None, 'product_id': None, 'user_email': 'user@example.com', 'location': None, 'device_type': None}

Consumed message from topic Fiqar_DE4_stock_avro_topic, partition 0, offset 1996

Key: 97e18035-0698-441e-8744-88744-80affd0502c, Value: {'user_id': 2, 'event_type': 'LOGIN', 'ts': '2023-12-12', 'page_url': None, 'product_id': None, 'user_email': None, 'location': 'New York', 'device_type': None}

Consumed message from topic Fiqar_DE4_stock_avro_topic, partition 0, offset 1998

Key: 2a861119-e44a-41b3-97a2-c6a5ac862bf7, Value: {'user_id': 2, 'event_type': 'CLOGUT', 'ts': '2023-12-12', 'page_url': 'http://e xample.com', 'product_id': None, 'user_email': None, 'location': None, 'device_type': 'mobile'}

Consumed message from topic Fiqar_DE4_stock_avro_topic, partition 0, offset 1999

Key: b4301983-86cf-453a-b49e-9de37663828, Value: {'user_id': 2, 'event_type': 'PAGE_VIEW', 'ts': '2023-12-12', 'page_url': 'http://example.com', 'product_id': None, 'user_email': None, 'location': None, 'device_type': 'mobile'}

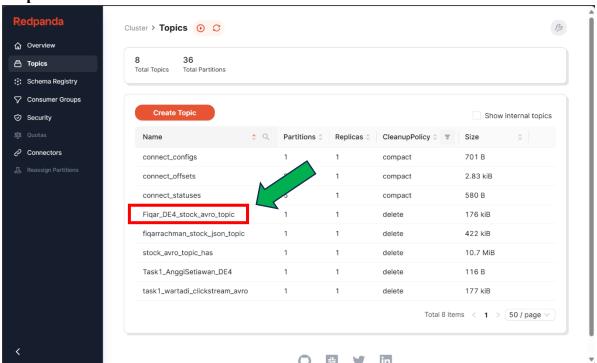
Consumed message from topic Fiqar_DE4_stock_avro_topic, partition 0, offset 1999

Key: b4301983-86cf-453a-b49e-9de37663828, Value: {'user_id': 2, 'event_type': 'PAGE_VIEW', 'ts': '2023-12-12', 'page_url': 'http://example.com', 'product_id': None, 'user_email': None, 'location': None, 'device_type': 'mobile'}

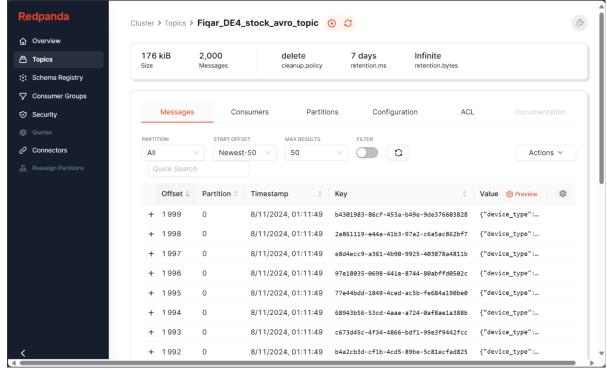
Berikut ini adalah proses saat menjalankan perintah "python consume avro.py"

24. Setelah menjalankan kedua file tersebut, langkah selanjutnya adalah memeriksa ke dalam website redpanda, apakah sudah berjalan sesuai yang diinginkan atau belum. Berikut ini adalah seluruh hasil dari script yang sudah dijalankan.

"Topics"



Berikut ini adalah tampilan list topic yang sudah dibuat, pada gambar diatas dapat terlihat bahwa topic dengan nama "Fiqar DE4 stock topic" berhasil dibuat.



Berikut ini adalah sebagian tampilan dari isi topic dengan nama "Fiqar_DE4_stock_topic".

Redpanda

○ Overview

○ Topics

○ Schema Registry

○ Consumer Groups

○ Security

№ Ountas

○ Connectors

○ Connectors

○ Figar_DE4_stock_avro_topic-value

figarrachman_stock_avro_topic-value

figarrachman_stock_avro_topic-value

figarrachman_stock_json_topic-value

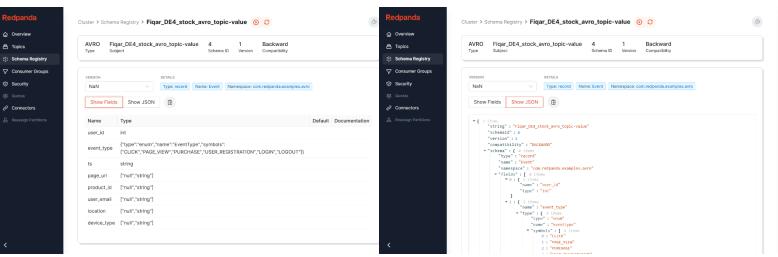
figarrachman_stock_json_topic-key

figarrachman_stock_json_topic-value

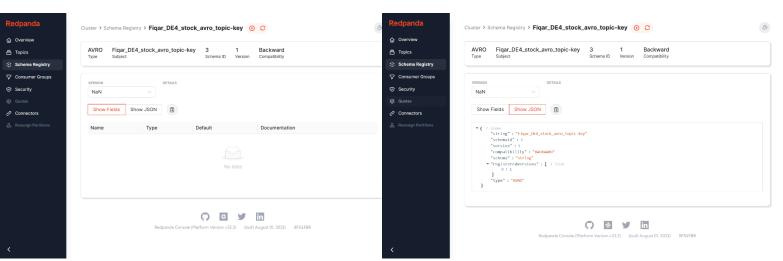
stock_avro_topic-value

task1_wartadl_clickstream_avro-key

task1_wartadl_clickstream_avro-value

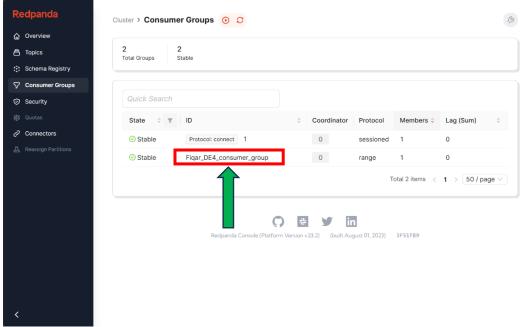


Berikut ini adalah tampilan fields dan json dari schema registry yang bernama "Fiqar_DE4_stock_avro_topic_value".

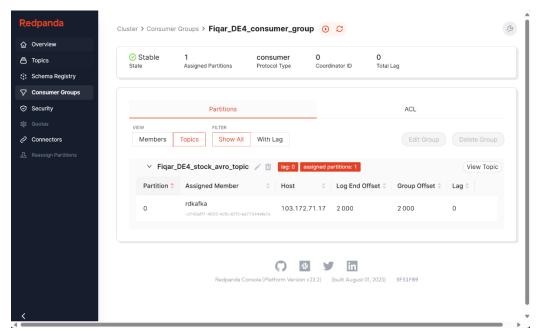


Berikut ini adalah tampilan fields dan json dari schema registry yang bernama "Fiqar_DE4_stock_avro_topic_key".

"Consumer Groups"



Berikut ini adalah tampilan list consumer groups yang sudah dibuat, pada gambar diatas dapat terlihat bahwa consumer groups dengan nama "Fiqar DE4 consumer group" berhasil dibuat.



Berikut ini adalah tampilan dari isi consumer groups dengan nama "Fiqar_DE4_consumer_group".

Dari hasil – hasil yang sudah dipaparkan diatas melalui gambar, dapat disimpulkan bahwa proses pembuatan **Topics**, **Schema Registry**, maupun **Consumer Groups berhasil** dan berjalan sesuai yang diharapkan.