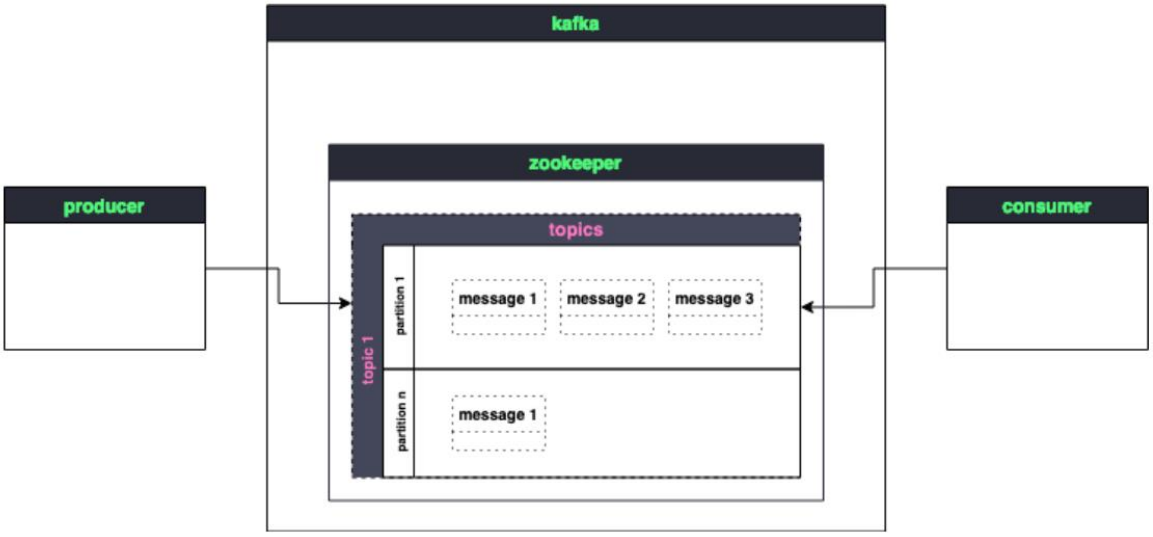
GeoStreaming using Apache Kafka producer and consumer with FastAPI and aiokafka

What is Apache Kafka?

Apache Kafka is a messaging system with a producer producing a message on the one side and a consumer consuming the message on the other side with a broker (zookeeper) in the middle.

When a producer sends a message, the message is pushed into Kafka topics. When a consumer consumes a message it is pulling the message from a Kafka topic. Kafka topics reside within a so-called broker (eg. Zookeeper). Zookeeper provides synchronization within distributed systems and in the case of Apache Kafka keeps track of the status of Kafka cluster nodes and Kafka topics.

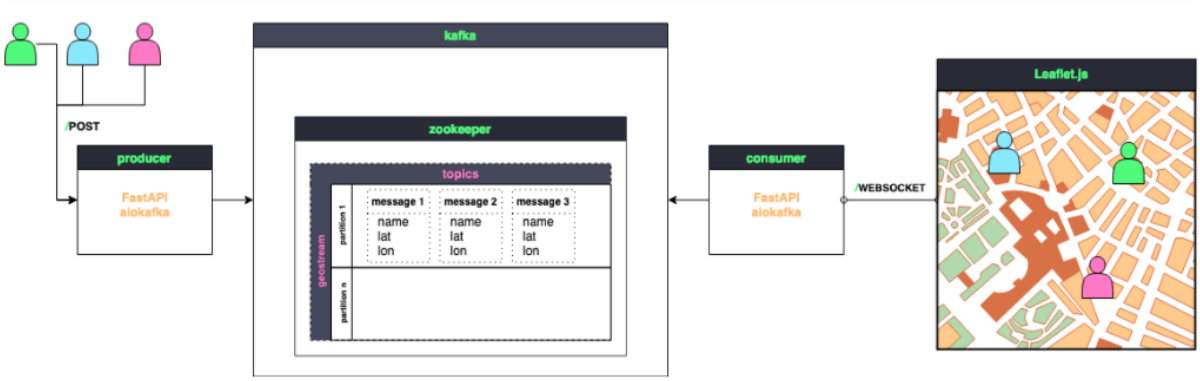
Kafka core principles



- When a producer sends a message, the message is pushed into Kafka topics. When a consumer consumes a message it is pulling the message from a Kafka topic.
- Kafka topics reside within a so-called broker (eg. Zookeeper). Zookeeper provides synchronization within distributed systems and in the case of Apache Kafka keeps track of the status of Kafka cluster nodes and Kafka topics.
- Multiple producers pushing messages into one topic, or you can have them push to different topics
- Messages within topics can be retained indefinitly or be discarded after a certain time, depending on the needs.
- When a consumer starts consuming a message, it starts from the first message that has been pushed to the topic and continues from thereon.
- Kafka stores the so-called *offset*, basically a pointer telling the consumer which messages have been consumed and what is still left to indulge.

- Messages will stay within the topic, yet when the same consumer pulls messages from the topic it will only receive messages from the offset onwards.

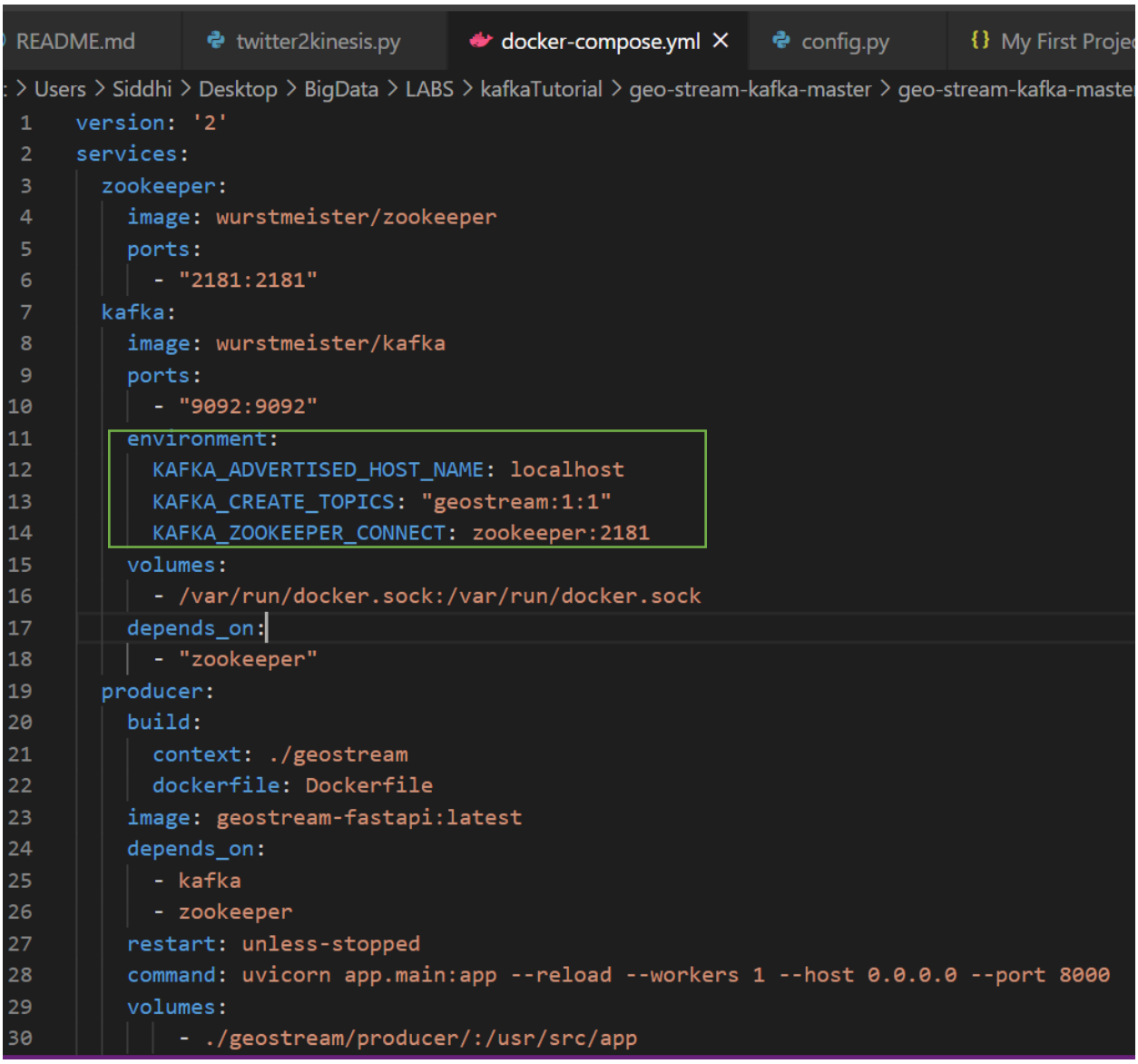
Geo Stream Kafka architecture

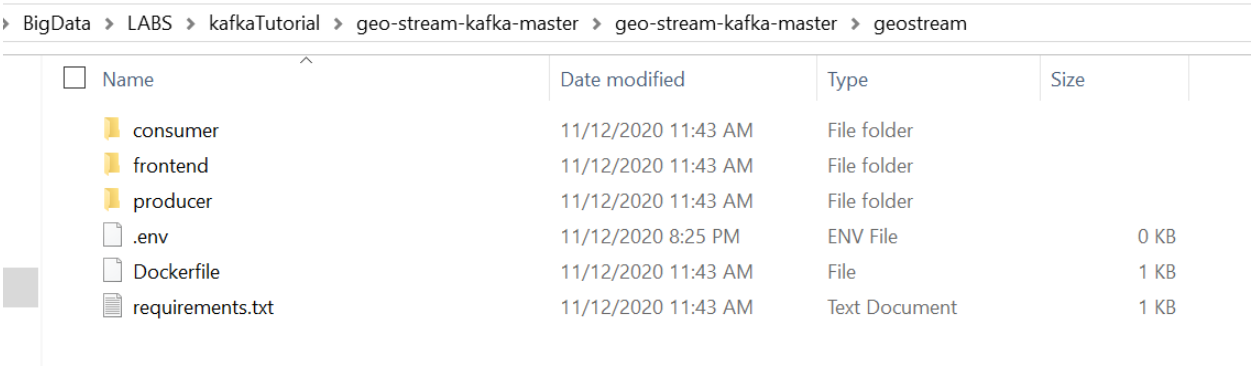


Steps to geostream latitude longitude data using Kafka on static website

Setup an Apache Kafka cluster

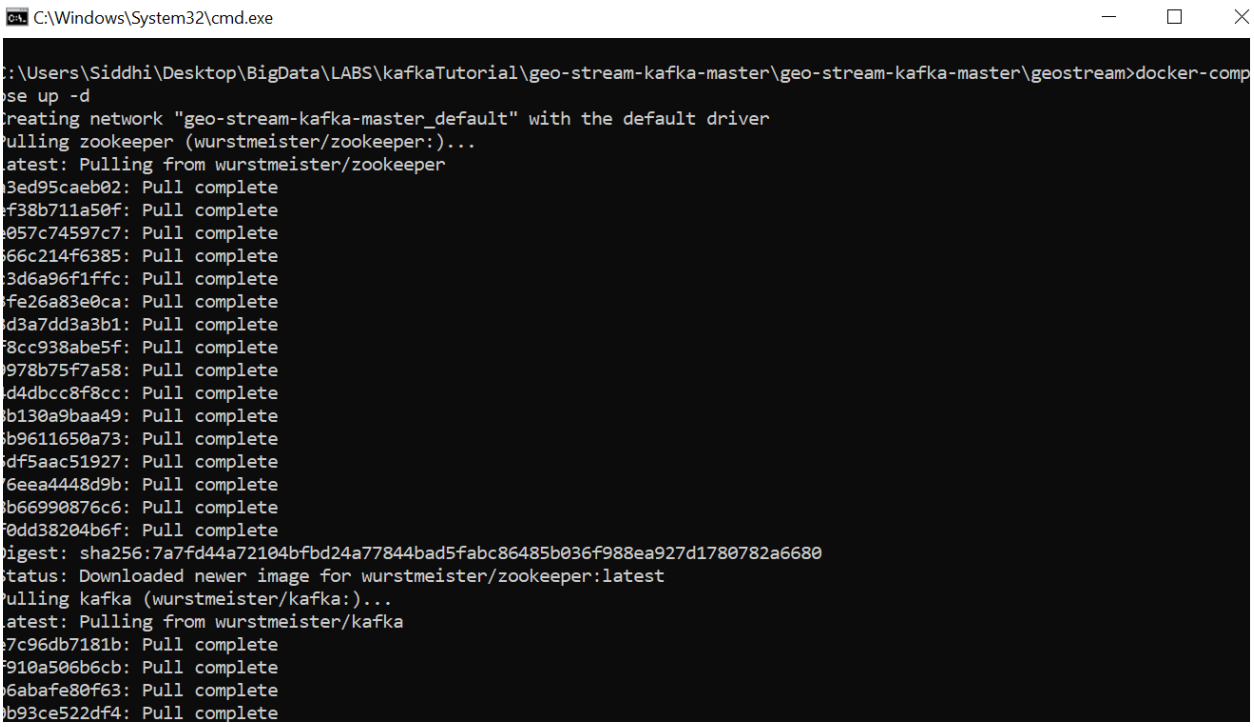
Change the KAFKA_ADVERTISED_HOST_NAME in the docker-compose.yaml and create a blank .env file in geostream folder





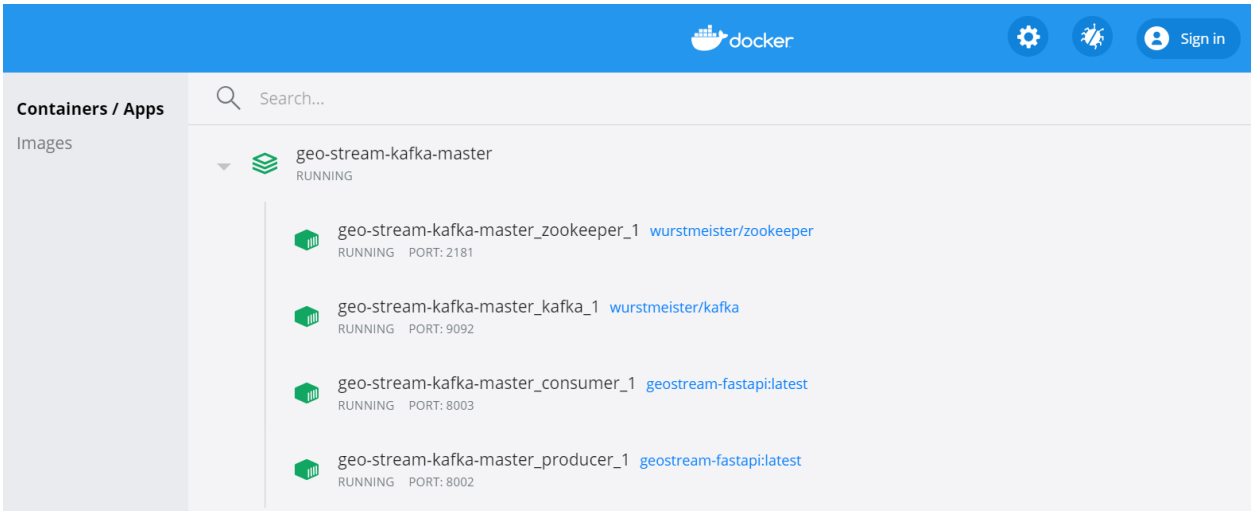
Run docker and use cmd to run the following command

This command would help spin up multiple docker containers for your kafka client, zookeeper, a producer and a consumer



```
C:\Windows\System32\cmd.exe
  Downloading idna-2.10-py2.py3-none-any.whl (58 kB)
Collecting certifi>=2017.4.17
  Downloading certifi-2020.11.8-py2.py3-none-any.whl (155 kB)
Collecting six
  Downloading six-1.15.0-py2.py3-none-any.whl (10 kB)
Collecting pyparsing>=2.0.2
  Downloading pyparsing-2.4.7-py2.py3-none-any.whl (67 kB)
Collecting zipp>=0.5
  Downloading zipp-3.4.0-py3-none-any.whl (5.2 kB)
Installing collected packages: kafka-python, aiokafka, loguru, toml, zipp, importlib-metadata, pluggy, py, attrs, six, p
yparsing, packaging, iniconfig, pytest, coverage, pytest-cov, pytest-asyncio, chardet, urllib3, idna, certifi, requests
Successfully installed aiokafka-0.7.0 attrs-20.3.0 certifi-2020.11.8 chardet-3.0.4 coverage-5.3 idna-2.10 importlib-meta
data-2.0.0 iniconfig-1.1.1 kafka-python-2.0.2 loguru-0.5.3 packaging-20.4 pluggy-0.13.1 py-1.9.0 pyparsing-2.4.7 pytest-
6.1.2 pytest-asyncio-0.14.0 pytest-cov-2.10.1 requests-2.25.0 six-1.15.0 toml-0.10.2 urllib3-1.26.2 zipp-3.4.0
+ rm -rf /root/.cache/pip
Removing intermediate container b893b4bae767
 ---> 481a58f45c3b

Successfully built 481a58f45c3b
Successfully tagged geostream-fastapi:latest
WARNING: Image for service producer was built because it did not already exist. To rebuild this image you must use `dock
er-compose build` or `docker-compose up --build`.
Creating geo-stream-kafka-master_zookeeper_1 ... done
Creating geo-stream-kafka-master_kafka_1     ... done
Creating geo-stream-kafka-master_consumer_1  ... done
Creating geo-stream-kafka-master_producer_1  ... done

C:\Users\Siddhi\Desktop\BigData\LABS\kafkaTutorial\geo-stream-kafka-master\geo-stream-kafka-master\geostream>
```

Docker dashboard has the containers running

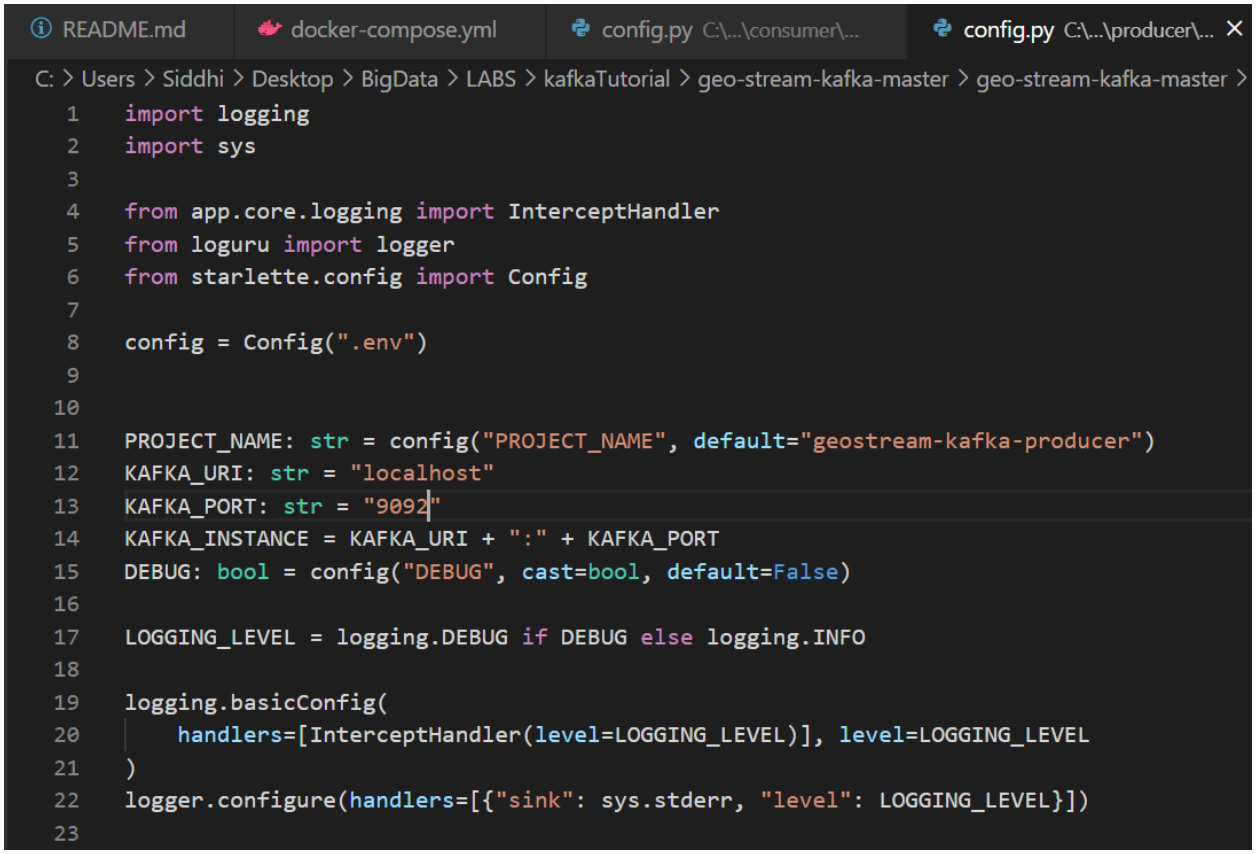


Change the configurations of producer and consumer

README.md | docker-compose.yml | config.py

C: > Users > Siddhi > Desktop > BigData > LABS > kafkaTutorial > geo-stream-kafka-master > geo-stream-kafka-mas

```python
import logging
import sys

from app.core.logging import InterceptHandler
from loguru import logger
from starlette.config import Config

config = Config(".env")


PROJECT_NAME: str = config("PROJECT_NAME", default="geostream-kafka-consumer")
KAFKA_URI: str = "localhost"
KAFKA_PORT: str = "9092"
KAFKA_INSTANCE = KAFKA_URI + ":" + KAFKA_PORT
DEBUG: bool = config("DEBUG", cast=bool, default=False)

LOGGING_LEVEL = logging.DEBUG if DEBUG else logging.INFO

logging.basicConfig(
    handlers=[InterceptHandler(level=LOGGING_LEVEL)], level=LOGGING_LEVEL
)
logger.configure(handlers=[{"sink": sys.stderr, "level": LOGGING_LEVEL}])

```



The KAFKA_URI and KAFKA_PORT need to be configured here(the above is for windows) the KAFKA_URI for Mac would be the same as KAFKA_ADVERTISED_HOST_NAME above. The KAFKA_PORT would be 9092 if you haven't made changes to the docker file.

FastAPI Apache Kafka producer

Wrap the producer into a FastAPI endpoint. This allows for more than one entity at a time to produce messages to a topic, but also enables to flexibly change topics that I want to produce messages to with FastAPI endpoint path parameters.

Used aiokafka to make use of FastAPIs async capabilities.

FastAPIs on_event("startup) and on_event("shutdown") make the use of a aiokafka producer easy.

```python
import asyncio
import json

from aiokafka import AIOKafkaProducer
from app.core.config import KAFKA_INSTANCE
from app.core.config import PROJECT_NAME
from app.core.models.model import ProducerMessage
from app.core.models.model import ProducerResponse
from fastapi import FastAPI
from loguru import logger

app = FastAPI(title=PROJECT_NAME)

loop = asyncio.get_event_loop()
aioproducer = AIOKafkaProducer(
    loop=loop, client_id=PROJECT_NAME, bootstrap_servers=KAFKA_INSTANCE
)


@app.on_event("startup")
async def startup_event():
    await aioproducer.start()


@app.on_event("shutdown")
async def shutdown_event():
    await aioproducer.stop()
```

Used `aioproducer` in the application

```python
@app.post("/producer/{topicname}")
async def kafka_produce(msg: ProducerMessage, topicname: str):
    """
    Produce a message into <topicname>

    This will produce a message into a Apache Kafka topic

    And this path operation will:

    * return ProducerResponse
    """

    await aioproducer.send(topicname, json.dumps(msg.dict()).encode("ascii"))
    response = ProducerResponse(
        name=msg.name, message_id=msg.message_id, topic=topicname
    )
    logger.info(response)
    return response


@app.get("/ping")
def ping():
    return {"ping": "pong!"}
```

FastAPI Apache Kafka consumer

Since FastAPI is built on-top of starlette we can use class-basedd endpoints and especially the WebsocketEndpoint to handle incoming `Websocket` Sessions.

When an application starts a websocket connection with out websocket endpoint we grab the event loop, use that to build and start the aiokafka consumer, start it and start a consumer task in the loop.

Once this is set, everytime the consumer pulls a new message it is forwarded to the application through the websocket.

```python
import asyncio
import json
import typing

from aiokafka import AIOKafkaConsumer
from app.core.config import KAFKA_INSTANCE
from app.core.config import PROJECT_NAME
from app.core.models.model import ConsumerResponse
from fastapi import FastAPI
from fastapi import WebSocket
from loguru import logger
from starlette.endpoints import WebSocketEndpoint
from starlette.middleware.cors import CORSMiddleware


app = FastAPI(title=PROJECT_NAME)
app.add_middleware(CORSMiddleware, allow_origins=["*"])


async def consume(consumer, topicname):
    async for msg in consumer:
        return msg.value.decode()
```

```python
@app.websocket_route("/consumer/{topicname}")
class WebsocketConsumer(WebSocketEndpoint):
    """
    Consume messages from <topicname>

    This will start a Kafka Consumer from a topic

    And this path operation will:

    * return ConsumerResponse
    """

    async def on_connect(self, websocket: WebSocket) -> None:
        topicname = websocket["path"].split("/")[2]  # until I figure out an alternative

        await websocket.accept()
        await websocket.send_json({"Message: ": "connected"})

        loop = asyncio.get_event_loop()
        self.consumer = AIOKafkaConsumer(
            topicname,
            loop=loop,
            client_id=PROJECT_NAME,
            bootstrap_servers=KAFKA_INSTANCE,
            enable_auto_commit=False,
        )
```

```
            bootstrap_servers=KAFKA_INSTANCE,
            enable_auto_commit=False,
        )

        await self.consumer.start()

        self.consumer_task = asyncio.create_task(
            self.send_consumer_message(websocket=websocket, topicname=topicname)
        )

        logger.info("connected")

    async def on_disconnect(self, websocket: WebSocket, close_code: int) -> None:
        self.consumer_task.cancel()
        await self.consumer.stop()
        logger.info(f"counter: {self.counter}")
        logger.info("disconnected")
        logger.info("consumer stopped")

    async def on_receive(self, websocket: WebSocket, data: typing.Any) -> None:
        await websocket.send_json({"Message: ": data})

    async def send_consumer_message(self, websocket: WebSocket, topicname: str) -> None:
        self.counter = 0
        while True:
            data = await consume(self.consumer, topicname)
            response = ConsumerResponse(topic=topicname, **json.loads(data))
            logger.info(response)
            await websocket.send_text(f"{response.json()}")
            self.counter = self.counter + 1


@app.get("/ping")
def ping():
    return {"ping": "pong!"}
```

Leaflet application

Declare our map and websocket connection to the `/consumer/{topicname}` endpoint.

```
var map = new L.Map('map');
var linesLayer = new L.LayerGroup();
var ws = new WebSocket("ws://127.0.0.1:8003/consumer/geostream");

var osmUrl = 'https://{s}.basemaps.cartocdn.com/light_all/{z}/{x}/{y}{r}.png',
    osmAttribution = '&copy; <a href="https://www.openstreetmap.org/copyright">OpenStreetMap</a> contributors &copy; <a href="https://carto.com/attribut
    osm = new L.tileLayer(osmUrl, {maxZoom: 18, attribution: osmAttribution});

var colors = ["#8be9fd", "#50fa7b", "#ffb86c", "#ff79c6", "#bd93f9", "#ff5555", "#f1fa8c"];

map.setView([52.521677, 13.391777], 15).addLayer(osm);

lines = {};
```

When you zoom on the map the `bounds` will be messed up and the events will not properly draw polylines along the trajectory of the entity. To fix this I added an `zoomend` trigger:

```
map.on("zoomend", function (e) { linesLayer.clearLayers() });
```

When client and backend established the silent agreement to use WebSockets, we can declare what we want to do, whenever the websocket receives a new message. Every message is

an `event`. And every `event` consists of metadata and `event.data` that can be parsed with `JSON.parse()`.

```
ws.onmessage = function(event) {
    console.log(event.data)
    obj = JSON.parse(event.data)
```

One of the requirements was to display more than one entity that pushes messages through the producer, Kafka and the consumer on the map as a live-event. For the leaflet application to associate an event to an entity, I hash events by the name of the entity that is sending them. If there is a new `name` in an event, it'll be hashed into a dictionary and added as a new layer on the map. As I wanted every entity to be represented with a different color, the color will be randomly grabbed from the list of `colors` and hashed alongside the position of the event/entity.

```
    if(!(obj.name in lines)) {
      lines[obj.name] = {"latlon": []};
      lines[obj.name]["latlon"].push([obj.lat, obj.lon]);
      lines[obj.name]["config"] = {"color": colors[Math.floor(Math.random()*colors.length)]};
    }
    else {
      lines[obj.name]["latlon"].push([obj.lat, obj.lon]);
    }

    line = L.polyline(lines[obj.name]["latlon"], {color: lines[obj.name]["config"]["color"]})
    linesLayer.addLayer(line)
    map.addLayer(linesLayer);

};
```

When the leafletjs application either specifically closes the websocket or the browser is closed, we close the websocket, cancel the consumer_task and stop the consumer.

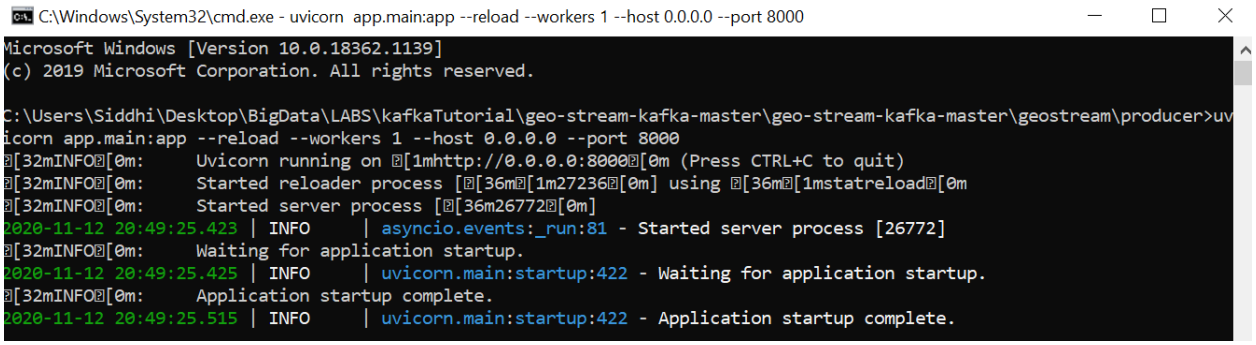
- Open to command prompts and change directory to respectively :

1. geo-stream-kafka-master\geostream\producer
2. geo-stream-kafka-master\geostream\consumer

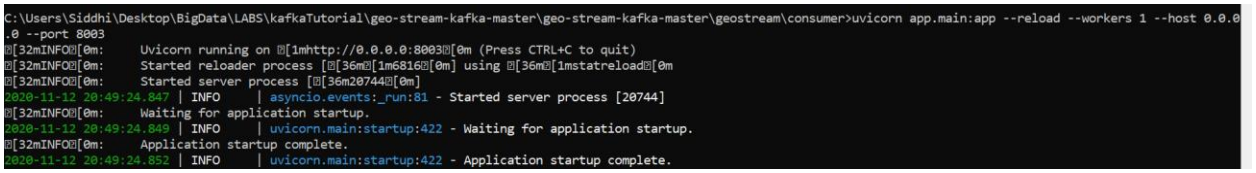
We will use the command prompts to run the producer and consumer simultaneously,

The producer should be sending messages for the consumer to receive messages hence both should be active when trying to test

Producer

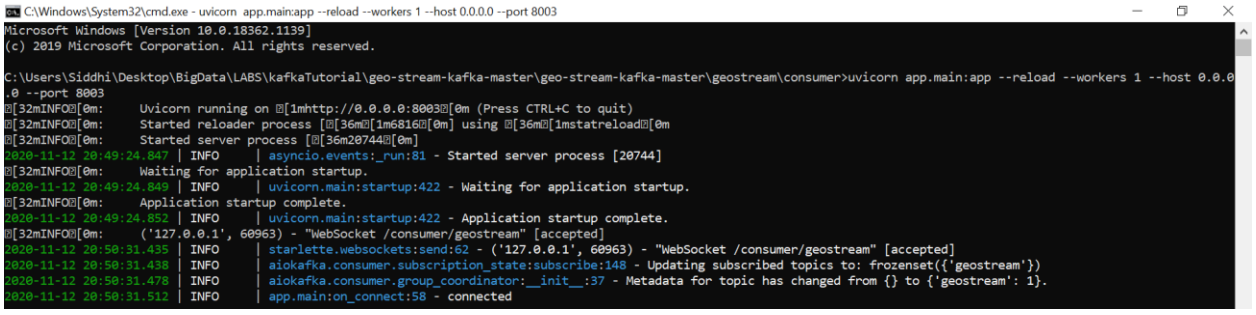


Consumer



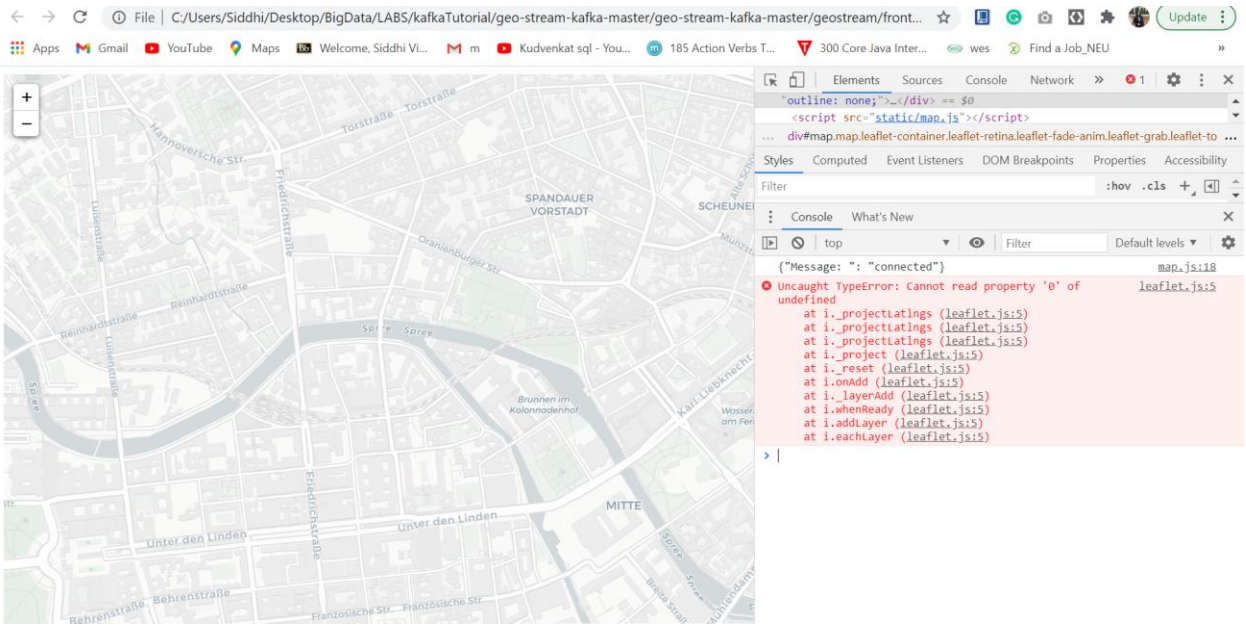
- For the consumer to be connected and ready to receive messages we need to access the static webpage that is configured as the consumer for the incoming websocket messages

After opening index.html page



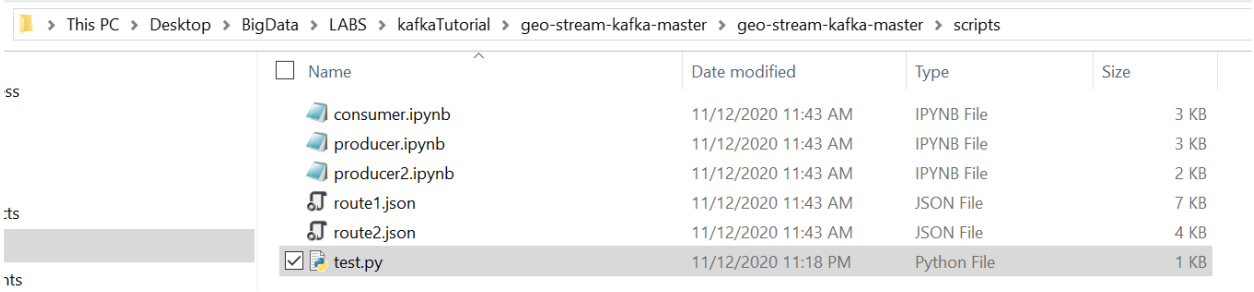
Successfully connected static webpage

The red errors are simply an indicator that there are no incoming messages to the consumer as of now.



- To stream data from producer to consumer configure a python file test.py as below in folder: geo-stream-kafka-master\scripts

```
import requests
import time
import json
import geojson

with open("route1.json", "r") as r:
    route = geojson.load(r)

msg = {"name": "Entity_ONE"}


for lon, lat in route["features"][0]["geometry"]["coordinates"]:
    msg["lat"] = lat
    msg["lon"] = lon
    requests.post("http://127.0.0.1:8000/producer/geostream", json=msg)
    time.sleep(0.2)
```

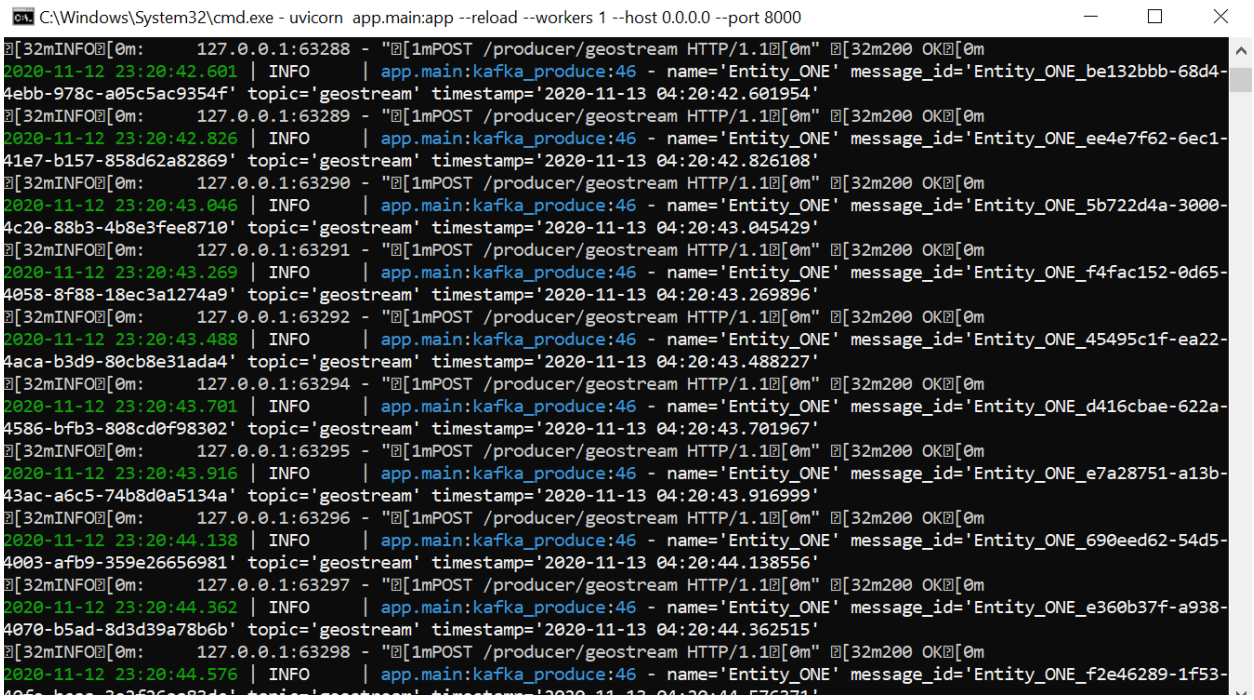


Install geojson package

```
C:\Users\Siddhi\Desktop\BigData\LABS\kafkaTutorial\geo-stream-kafka-master\geo-stream-kafka-master\scripts>pip install g
eojson
Collecting geojson
  Downloading geojson-2.5.0-py2.py3-none-any.whl (14 kB)
Installing collected packages: geojson
Successfully installed geojson-2.5.0

C:\Users\Siddhi\Desktop\BigData\LABS\kafkaTutorial\geo-stream-kafka-master\geo-stream-kafka-master\scripts>python test.p
y

C:\Users\Siddhi\Desktop\BigData\LABS\kafkaTutorial\geo-stream-kafka-master\geo-stream-kafka-master\scripts>
```

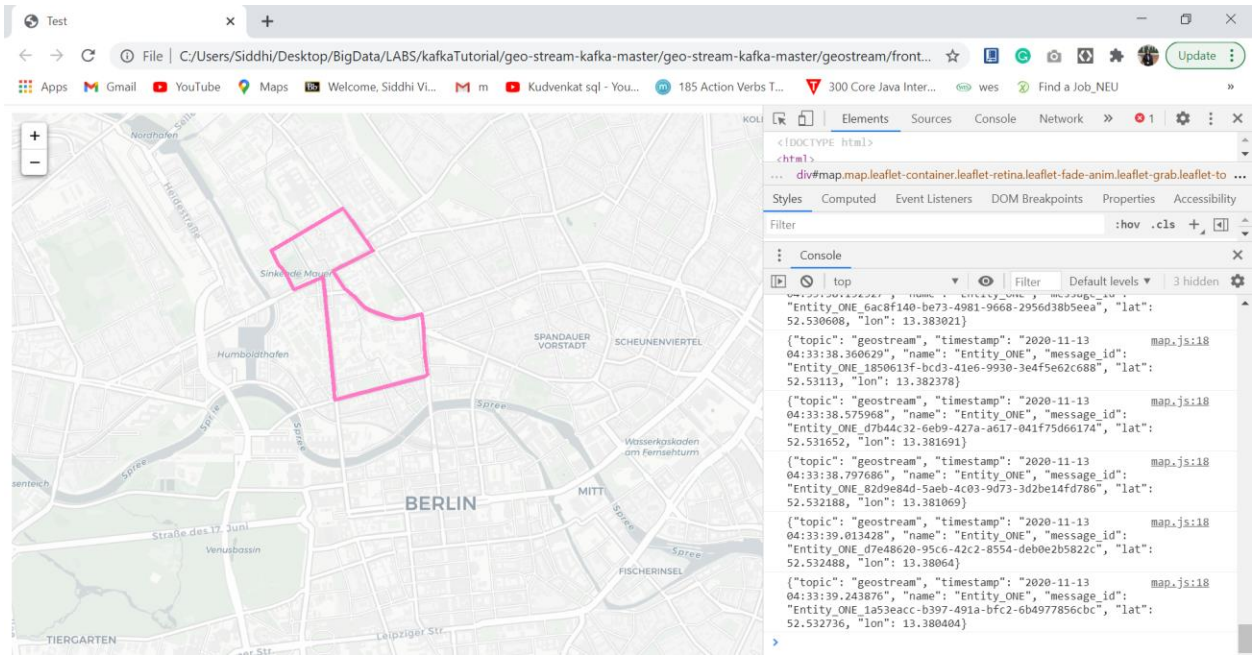
Run the python file:-

Producer producing messages after executing python script on producer's port



Consumer is consuming messages

We can see the path being traced in real-time when the script is ran.


Test
File | C:/Users/Siddhi/Desktop/BigData/LABS/kafkaTutorial/geo-stream-kafka-master/geo-stream-kafka-master/geostream/front...
Update
Apps
Gmail
YouTube
Maps
Welcome, Siddhi Vi...
m
Kudvenkat sql - You...
185 Action Verbs T...
300 Core Java Inter...
wes
Find a Job_NEU
Elements
Sources
Console
Network
<!DOCTYPE html>
<html>
div#map.map.leaflet-container.leaflet-retina.leaflet-fade-anim.leaflet-grab.leaflet-to
Styles
Computed
Event Listeners
DOM Breakpoints
Properties
Accessibility
Filter
:hov .cls
Console
top
Filter
Default levels
3 hidden
"Entity_ONE_6ac8f140-be73-4981-9668-2956d38b5eea", "lat": 52.530608, "lon": 13.383021}
{"topic": "geostream", "timestamp": "2020-11-13 04:33:38.360629", "name": "Entity_ONE", "message_id": "Entity_ONE_1850613f-bcd3-41e6-9930-3e4f5e62c688", "lat": 52.53113, "lon": 13.382378}
{"topic": "geostream", "timestamp": "2020-11-13 04:33:38.575968", "name": "Entity_ONE", "message_id": "Entity_ONE_d7b44c32-6eb9-427a-a617-041f75d66174", "lat": 52.531652, "lon": 13.381691}
{"topic": "geostream", "timestamp": "2020-11-13 04:33:38.797686", "name": "Entity_ONE", "message_id": "Entity_ONE_82d9e84d-5aeb-4c03-9d73-3d2be14fd786", "lat": 52.532188, "lon": 13.381069}
{"topic": "geostream", "timestamp": "2020-11-13 04:33:39.013428", "name": "Entity_ONE", "message_id": "Entity_ONE_d7e48620-95c6-42c2-8554-deb0e2b5822c", "lat": 52.532488, "lon": 13.38064}
{"topic": "geostream", "timestamp": "2020-11-13 04:33:39.243876", "name": "Entity_ONE", "message_id": "Entity_ONE_1a53eacc-b397-491a-bfc2-6b4977856cbc", "lat": 52.532736, "lon": 13.380404}
map.js:18
Nordhafen
Sinkende Mauer
Humboldthafen
SPANDAUER VORSTADT
SCHEUNENVIERTEL
Spree
Wasserkaskaden am Fernsehturm
BERLIN
MITT
Straße des 17. Juni
Venusbassin
FISCHERINSEL
TIERGARTEN
Leipziger Str