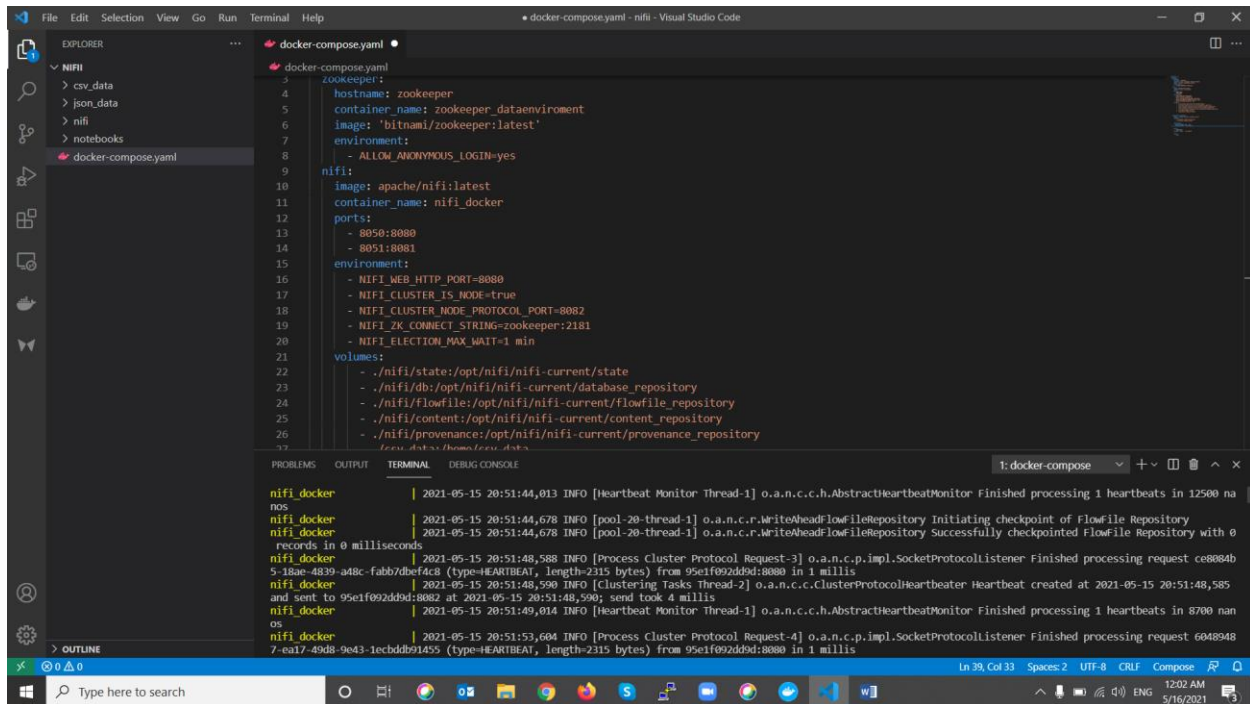


Mohd Tulaib

20208002

clone the git hub



The screenshot shows the Visual Studio Code interface with a Docker Compose file named `docker-compose.yml` open in the editor. The file defines two services: `zookeeper` and `nifi`. The `zookeeper` service uses the `bitnami/zookeeper:latest` image and has the environment variable `ALLOW_ANONYMOUS_LOGIN=yes`. The `nifi` service uses the `apache/nifi:latest` image and has several environment variables including `NIFI_WEB_HTTP_PORT=8080`, `NIFI_CLUSTER_IS_NODE=true`, `NIFI_CLUSTER_NODE_PROTOCOL_PORT=8082`, `NIFI_ZK_CONNECT_STRING=zookeeper:2181`, and `NIFI_ELECTION_MAX_WAIT=1 min`. It also has several volumes mapped to local directories. The terminal at the bottom shows the output of the `docker-compose up` command, indicating that the services are starting successfully. The output includes logs from the `nifi_docker` service, showing that the `AbstractHeartbeatMonitor` finished processing 1 heartbeat in 12500 ns, and the `FlowfileRepository` successfully checkpointed the repository with 0 records in 0 milliseconds.

```
docker-compose.yml
3 zookeeper:
4   hostname: zookeeper
5   container_name: zookeeper_dataenvironment
6   image: 'bitnami/zookeeper:latest'
7   environment:
8     - ALLOW_ANONYMOUS_LOGIN=yes
9
10  nifi:
11    image: apache/nifi:latest
12    container_name: nifi_docker
13    ports:
14      - 8050:8080
15      - 8051:8081
16    environment:
17      - NIFI_WEB_HTTP_PORT=8080
18      - NIFI_CLUSTER_IS_NODE=true
19      - NIFI_CLUSTER_NODE_PROTOCOL_PORT=8082
20      - NIFI_ZK_CONNECT_STRING=zookeeper:2181
21      - NIFI_ELECTION_MAX_WAIT=1 min
22    volumes:
23      - ./nifi/state:/opt/nifi/nifi-current/state
24      - ./nifi/db:/opt/nifi/nifi-current/database_repository
25      - ./nifi/flowfile:/opt/nifi/nifi-current/flowfile_repository
26      - ./nifi/content:/opt/nifi/nifi-current/content_repository
27      - ./nifi/provenance:/opt/nifi/nifi-current/provenance_repository
```

```
2021-05-15 20:51:44,013 INFO [Heartbeat Monitor Thread-1] o.a.n.c.c.h.AbstractHeartbeatMonitor Finished processing 1 heartbeats in 12500 ns
2021-05-15 20:51:44,678 INFO [pool-20-thread-1] o.a.n.c.r.WriteAheadFlowfileRepository Initiating checkpoint of Flowfile Repository
2021-05-15 20:51:44,678 INFO [pool-20-thread-1] o.a.n.c.r.WriteAheadFlowfileRepository Successfully checkpointed Flowfile Repository with 0
records in 0 milliseconds
2021-05-15 20:51:48,588 INFO [Process Cluster Protocol Request-3] o.a.n.c.p.impl.SocketProtocolListener Finished processing request ce808ab
5-18ae-4839-a48c-fabb7dbef4c8 (type=HEARTBEAT, length=2315 bytes) from 95e1f092dd9d:8080 in 1 millis
2021-05-15 20:51:48,590 INFO [Clustering Tasks Thread-2] o.a.n.c.c.ClusterProtocolHeartbeater Heartbeat created at 2021-05-15 20:51:48,585
and sent to 95e1f092dd9d:8082 at 2021-05-15 20:51:48,590; send took 4 millis
2021-05-15 20:51:49,014 INFO [Heartbeat Monitor Thread-1] o.a.n.c.c.h.AbstractHeartbeatMonitor Finished processing 1 heartbeats in 8700 ns
2021-05-15 20:51:53,604 INFO [Process Cluster Protocol Request-4] o.a.n.c.p.impl.SocketProtocolListener Finished processing request 6048948
7-ea17-49d8-9e43-1ecbdb91455 (type=HEARTBEAT, length=2315 bytes) from 95e1f092dd9d:8080 in 1 millis
```

\$ docker-compose up

JupyterLab <http://localhost:8880/>

Nifi <http://localhost:8050/nifi/>

Jupyter environment

The screenshot shows a Jupyter Notebook with the following code and output:

```
[1]: !pip install faker
```

Collecting faker
Downloading Faker-8.1.4-py3-none-any.whl (1.2 MB)
Collecting text-unidecode==1.3
Downloading text-unidecode-1.3-py2.py3-none-any.whl (78 kB)
Requirement already satisfied: python-dateutil<=2.4 in /opt/conda/lib/python3.9/site-packages (from faker) (2.8.1)
Requirement already satisfied: six>=1.5 in /opt/conda/lib/python3.9/site-packages (from python-dateutil<=2.4->faker) (1.16.0)
Installing collected packages: text-unidecode, faker
Successfully installed faker-8.1.4 text-unidecode-1.3

```
[2]: from faker import Faker
```

```
[5]: from faker import Faker
import csv
output=open('/home/csv_data/fakedata.csv','w')
fake=Faker()
headers=['name','age','street','city','state','zip','lng','lat']
mywriter=csv.writer(output)
mywriter.writerow(headers)
for r in range(10):
    row=[fake.name(),fake.random_int(min=18,max=80, step=1),
        fake.street_address(), fake.city(),fake.state(),
        fake.zipcode(),fake.longitude(),fake.latitude()]
    print(row)
    mywriter.writerow(row)

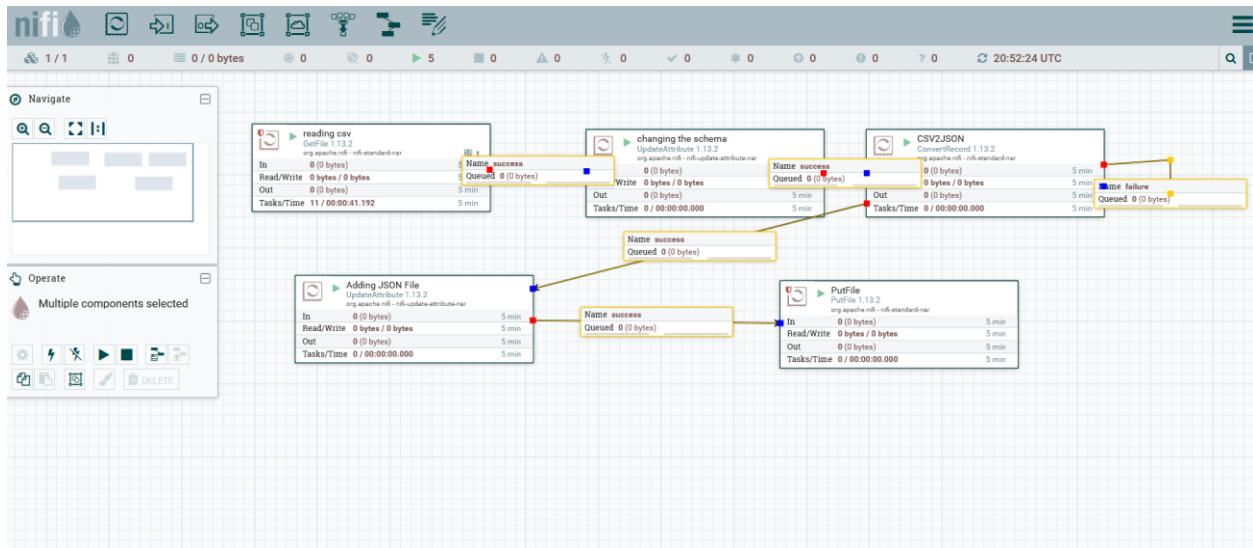
output.close()
```

The output of the notebook shows a list of 10 generated fake records, each containing a name, age, street address, city, state, zip code, longitude, and latitude.

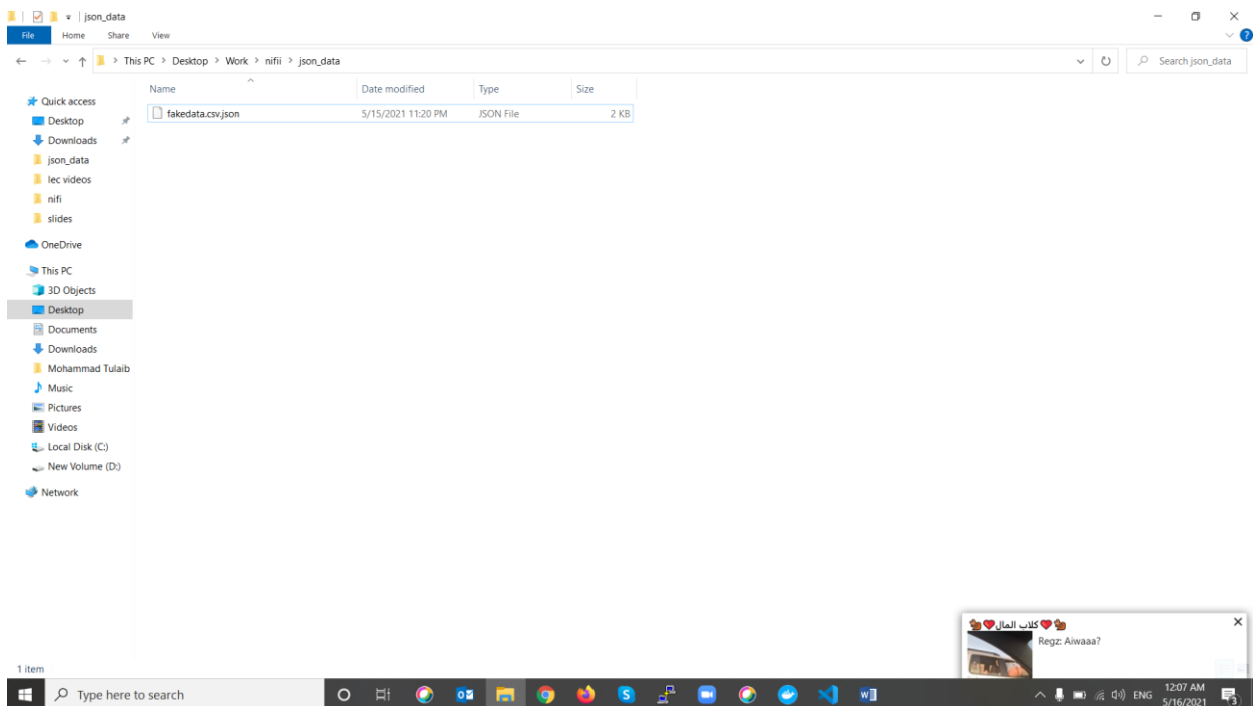
Data schema

```
{
  "type": "record",
  "name": "UserRecord",
  "fields" : [
    {"name": "name", "type": ["null", "string"]},
    {"name": "age", "type": ["null", "string"]},
    {"name": "street", "type": ["null", "string"]},
    {"name": "city", "type": ["null", "string"]},
    {"name": "state", "type": ["null", "string"]},
    {"name": "zip", "type": ["null", "string"]},
    {"name": "lng", "type": ["null", "string"]},
    {"name": "lat", "type": ["null", "string"]}
  ]
}
```

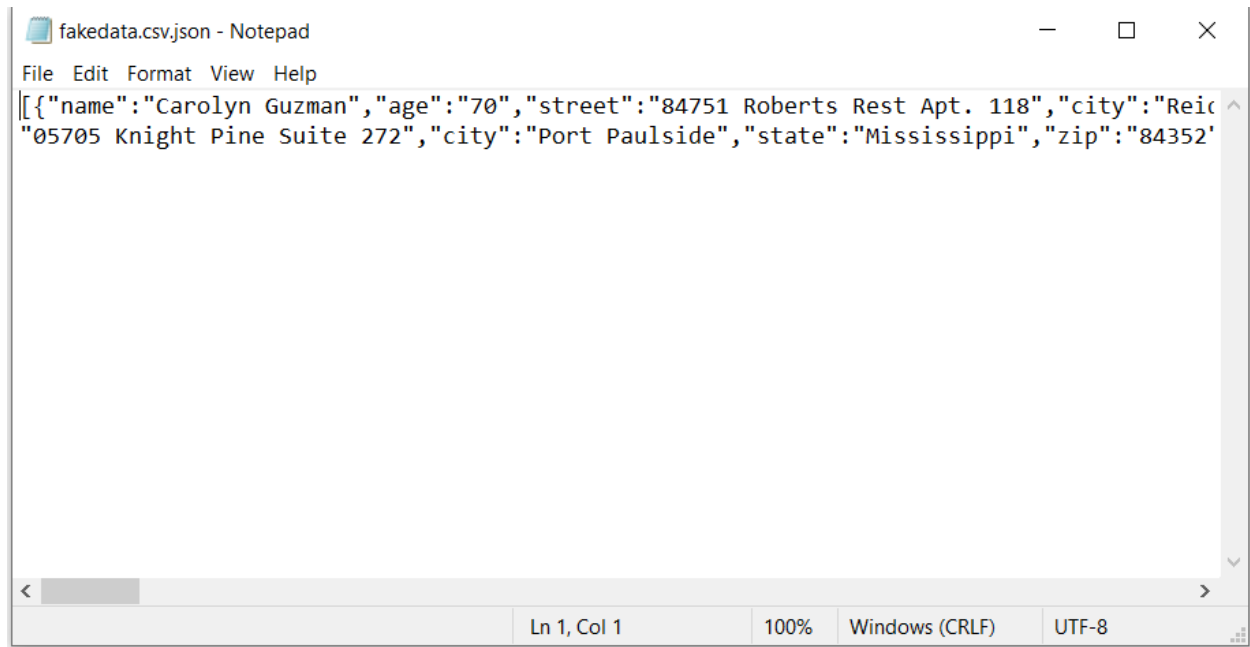
Nifi Environment



Output File



Output data



```
fakedata.csv,json - Notepad
File Edit Format View Help
[{"name":"Carolyn Guzman","age":"70","street":"84751 Roberts Rest Apt. 118","city":"Reic
"05705 Knight Pine Suite 272","city":"Port Paulside","state":"Mississippi","zip":"84352'

Ln 1, Col 1 100% Windows (CRLF) UTF-8
```