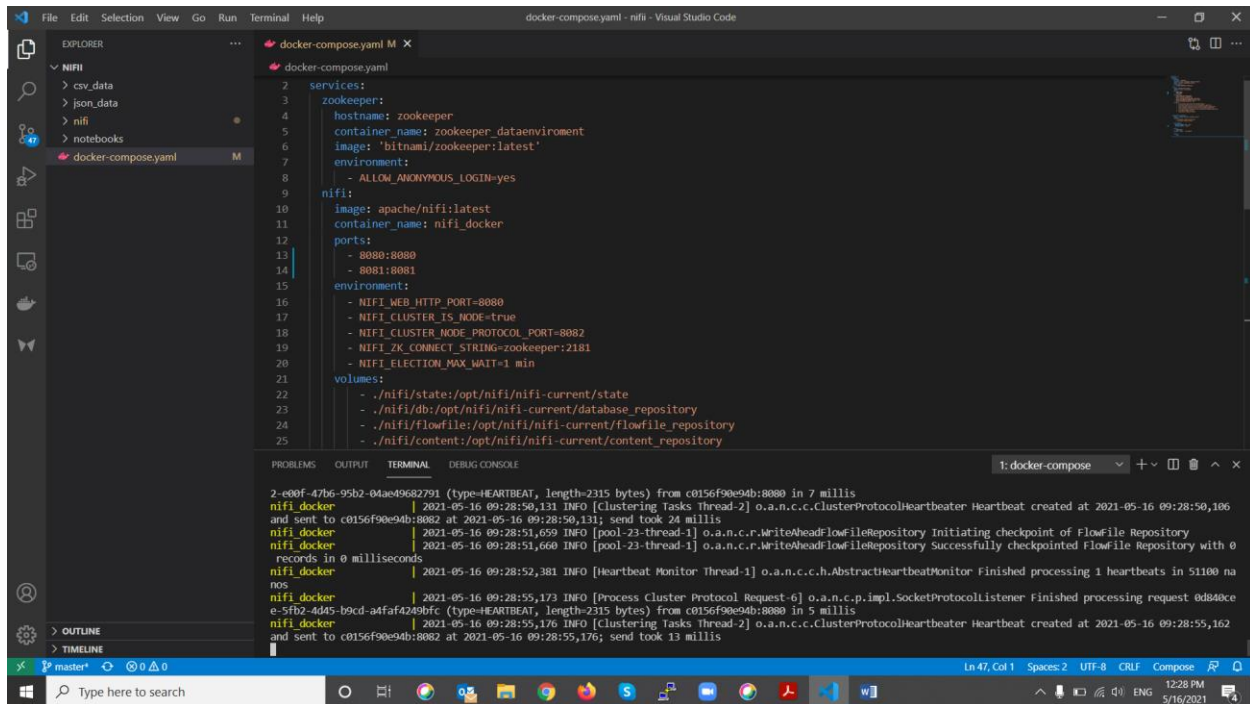


Mohd Tulaib

20208002

clone the git hub



The screenshot shows a Visual Studio Code editor with a Docker Compose file named `docker-compose.yml` open. The file defines two services: `zookeeper` and `nifi`. The `zookeeper` service uses the `bitnami/zookeeper:latest` image and has an environment variable `ALLOW_ANONYMOUS_LOGIN=yes`. The `nifi` service uses the `apache/nifi:latest` image and has several environment variables for configuration, 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 a `ports` section mapping `8080:8080` and `8081:8081`. The `volumes` section maps local directories to the container's `/opt/nifi/nifi-current/` directory.

```
services:
  zookeeper:
    hostname: zookeeper
    container_name: zookeeper_dataenvironment
    image: 'bitnami/zookeeper:latest'
    environment:
      - ALLOW_ANONYMOUS_LOGIN=yes
  nifi:
    image: apache/nifi:latest
    container_name: nifi_docker
    ports:
      - 8080:8080
      - 8081:8081
    environment:
      - NIFI_WEB_HTTP_PORT=8080
      - NIFI_CLUSTER_IS_NODE=true
      - NIFI_CLUSTER_NODE_PROTOCOL_PORT=8082
      - NIFI_ZK_CONNECT_STRING=zookeeper:2181
      - NIFI_ELECTION_MAX_WAIT=1 min
    volumes:
      - ./nifi/state:/opt/nifi/nifi-current/state
      - ./nifi/db:/opt/nifi/nifi-current/database_repository
      - ./nifi/flowfile:/opt/nifi/nifi-current/flowfile_repository
      - ./nifi/content:/opt/nifi/nifi-current/content_repository
```

The terminal output shows the following logs:

```
2-e00f-47b6-95b2-04ae49682791 (type=HEARTBEAT, length=2315 bytes) from c0156f90e94b:8080 in 7 millis
nifi_docker | 2021-05-16 09:28:50,131 INFO [Clustering Tasks Thread-2] o.a.n.c.c.ClusterProtocolHeartbeater Heartbeat created at 2021-05-16 09:28:50,106
and sent to c0156f90e94b:8082 at 2021-05-16 09:28:50,131; send took 28 millis
nifi_docker | 2021-05-16 09:28:51,659 INFO [pool-23-thread-1] o.a.n.c.r.WriteAheadFlowfileRepository Initiating checkpoint of Flowfile Repository
nifi_docker | 2021-05-16 09:28:51,660 INFO [pool-23-thread-1] o.a.n.c.r.WriteAheadFlowfileRepository Successfully checkpointed Flowfile Repository with 0
records in 0 milliseconds
nifi_docker | 2021-05-16 09:28:52,381 INFO [Heartbeat Monitor Thread-1] o.a.n.c.c.h.AbstractHeartbeatMonitor Finished processing 1 heartbeats in 51100 na
nos
nifi_docker | 2021-05-16 09:28:55,173 INFO [Process Cluster Protocol Request-6] o.a.n.c.p.impl.SocketProtocolListener Finished processing request 0d840ke
e-5fb2-4445-b9cd-a4af4249bfc (type=HEARTBEAT, length=2315 bytes) from c0156f90e94b:8080 in 5 millis
nifi_docker | 2021-05-16 09:28:55,176 INFO [Clustering Tasks Thread-2] o.a.n.c.c.ClusterProtocolHeartbeater Heartbeat created at 2021-05-16 09:28:55,162
and sent to c0156f90e94b:8082 at 2021-05-16 09:28:55,176; send took 13 millis
```

\$ docker-compose up

JupyterLab <http://localhost:8880/>

Password:
psut

Nifi :
localhost:8080/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)
1.2 MB 169 kB/s eta 0:00:01
Collecting text-unidecode==1.3
Downloading text-unidecode-1.3-py2.py3-none-any.whl (78 kB)
78 kB 2.6 MB/s eta 0:00:01
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()
```

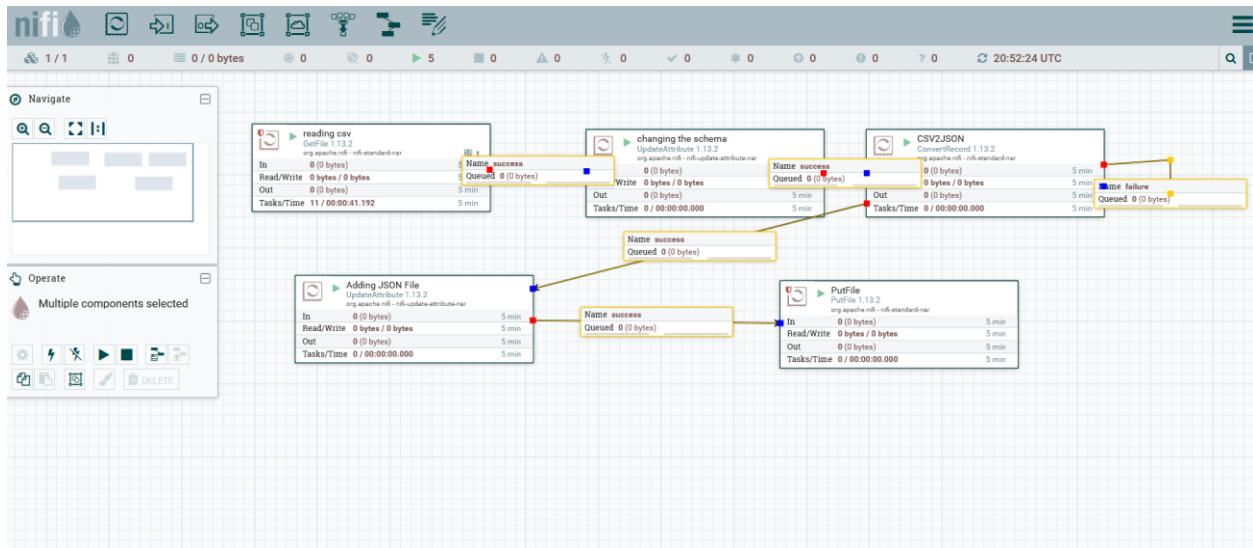
Output:

```
[['Carolyn Guzman', 70, '84751 Roberts Rest Apt. 118', 'Reidview', 'Ohio', '39170', Decimal('-60.550337'), Decimal('-19.4597015')],
 ['Darrell Anderson', 75, '5213 Richard Manor', 'Evanstown', 'Alaska', '180249', Decimal('65.775564'), Decimal('-36.9511045')],
 ['Timothy Wright', 28, '806 Melissa Hill', 'West Michael', 'Delaware', '13562', Decimal('-65.813193'), Decimal('19.7249405')],
 ['Clayton Johnson', 55, '6576 Morales Ports', 'East Donaldburgh', 'Utah', '69827', Decimal('-108.918537'), Decimal('72.608956')],
 ['Carolyn Carr', 38, '7356 Jenkins Common', 'Elizabethville', 'Indiana', '03060', Decimal('-14.891548'), Decimal('48.866408')],
 ['Antonio Alvarado', 22, '4717 Heidi Underpass Suite 499', 'East Jasonhaven', 'Mississippi', '84447', Decimal('147.055366'), Decimal('12.152132')],
 ['Jacqueline Meyer', 42, '05705 Knight Pine Suite 272', 'Port Paulside', 'Mississippi', '84352', Decimal('74.549257'), Decimal('80.0367415')],
 ['Nicole Nichols', 62, '2161 Castillo Lakes', 'Markbury', 'Georgia', '86490', Decimal('36.594606'), Decimal('-67.8874195')],
 ['Mitchell Farley', 75, '10456 James Alley Apt. 985', 'Lake Miaberg', 'North Dakota', '34225', Decimal('-176.996710'), Decimal('88.5585605')]]
```

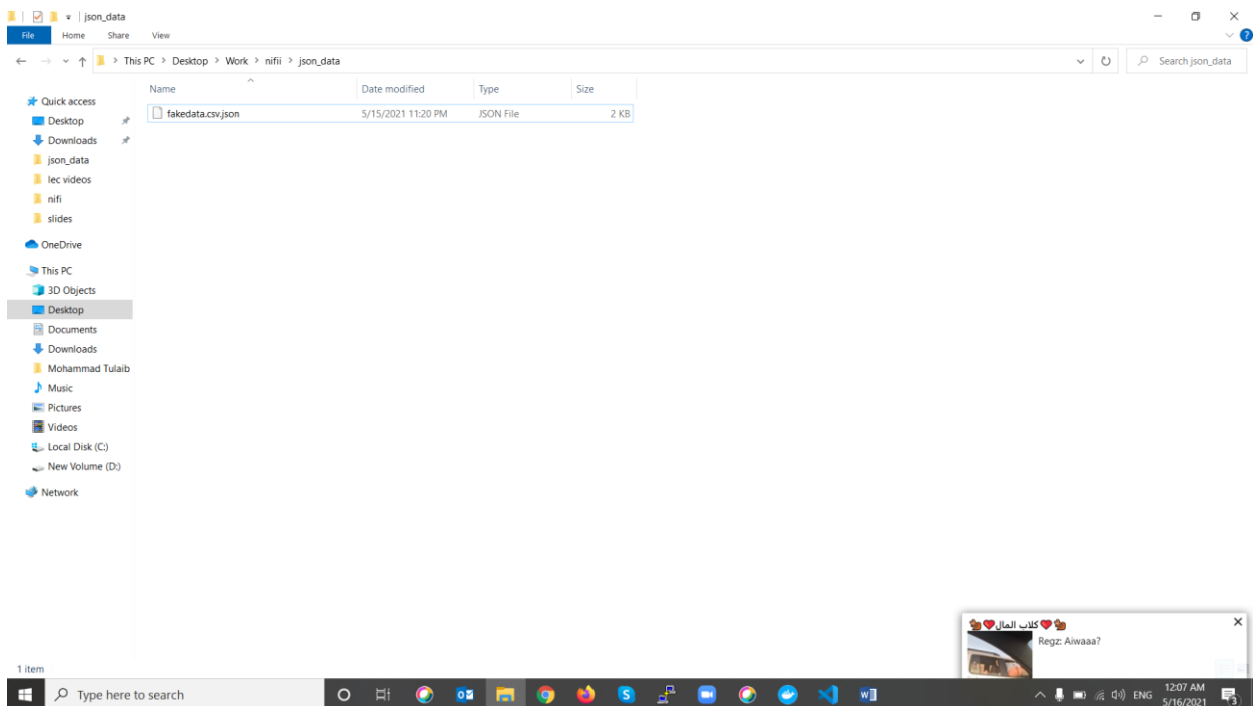
Data schema

```
{
  "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"]}
  ]
}
```

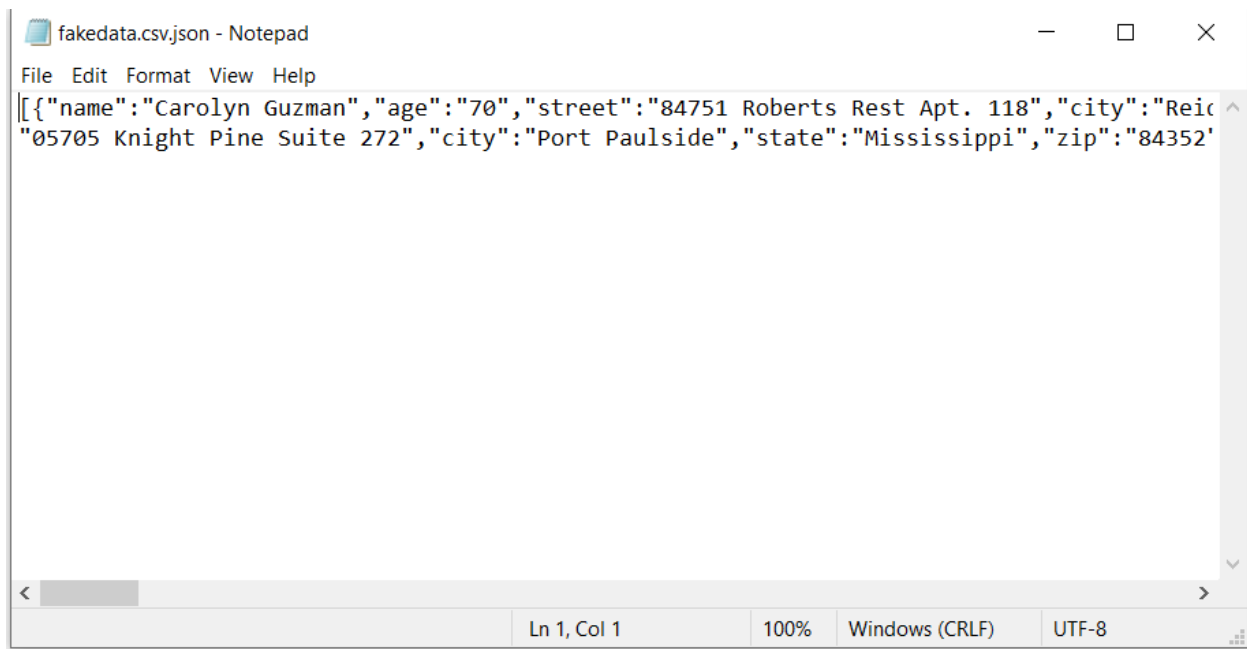
Nifi Environment



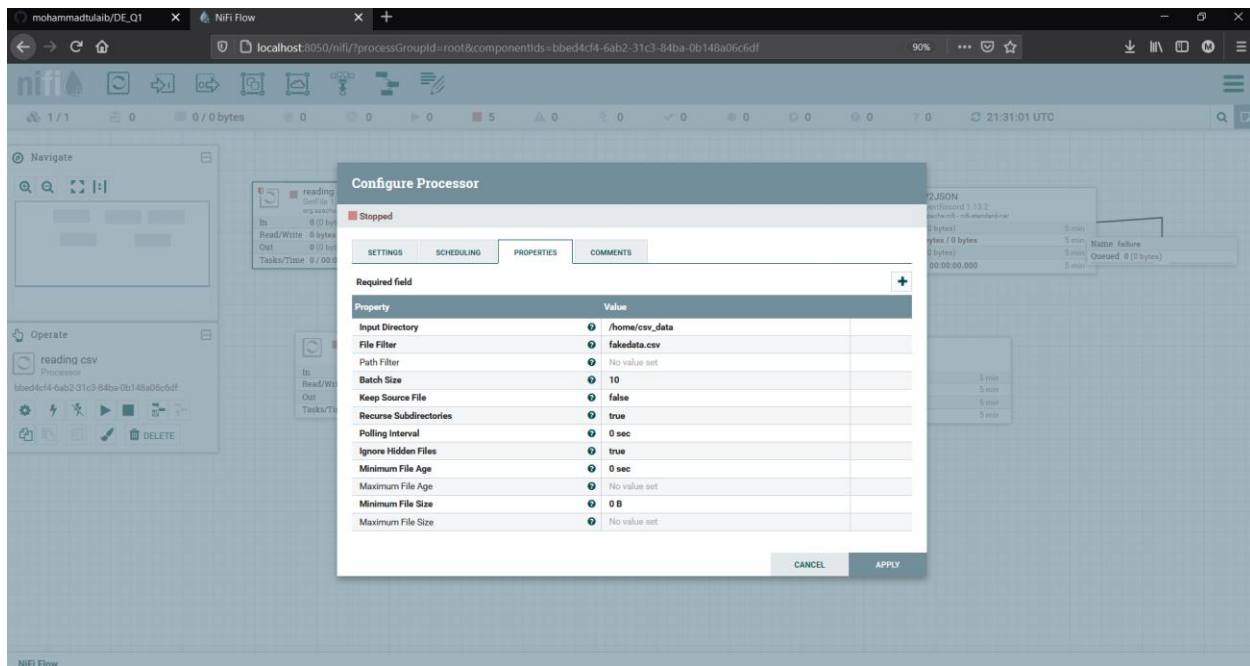
Output File



Output data



Nifi Config Screenshoots



Configure Processor

Stopped

SETTINGS

SCHEDULING

PROPERTIES

COMMENTS

Required field



Property		Value	
Delete Attributes Expression	?	No value set	
Store State	?	Do not store state	
Stateful Variables Initial Value	?	No value set	
Cache Value Lookup Cache Size	?	100	
schema.name	?	users	

ADVANCED

CANCEL

APPLY

Configuration window for the CSV2JSON Processor in the NIFI Flow interface.

The window shows the following properties:

Property	Value
Record Reader	CSVReader
Record Writer	JsonRecordSetWriter
Include Zero Record FlowFiles	true

The background shows the NIFI Flow interface with a flow diagram and a sidebar.

