

Capstone Project: Instant Health Alert System - Mid-Project Submission

Introduction:

This document outlines the code implementation for the Instant Health Alert System's mid-project submission. The system fetches patients' vital information from a MySQL database and streams it to a Kafka topic. This setup enables real-time data analysis and alerts for healthcare monitoring. This explains the file **kafka_produce_patient_vitals.py**

1. Database Connection parameters and producing the data in JSON format

```
import mysql.connector
from mysql.connector import Error
from kafka import KafkaProducer
import json
import time

HOST = 'upgraddetest.cyaie1c9bmnf.us-east-1.rds.amazonaws.com'
DATABASE = 'testdatabase'
USER = 'student'
PASSWORD = 'STUDENT123'
TOPIC='patients_vital_info'
KAFKA_BOOTSTRAP_SERVERS='localhost:9092'

producer = KafkaProducer(bootstrap_servers=KAFKA_BOOTSTRAP_SERVERS,
value_serializer=lambda x: json.dumps(x).encode('utf-8'))
```

2. Database Connection Check

This function, `check_connection`, tries to establish a connection to the MySQL database. If successful, it confirms connectivity. Otherwise, it prints an error.

```
def check_connection(host, database, user, password):  
    try:  
        connection = mysql.connector.connect(  
            host=host,  
            database=database,  
            user=user,  
            password=password  
        )  
        if connection.is_connected():  
            print("Connected to MySQL database")  
            return connection  
    except Error as e:  
        print(f"Error while connecting to MySQL: {e}")  
    return None
```

3. Data Fetching Function

The `fetch_data` function retrieves all records from the `patients_vital_info` table. If the connection is successful, it fetches and returns the data; otherwise, it handles the error and closes the connection.

```
def fetch_data(host, database, user, password):  
    connection = check_connection(host, database, user, password)  
    if connection is None:  
        print("Failed to connect to the database.")  
        return None  
  
    try:  
        cursor = connection.cursor()  
        cursor.execute('SELECT * FROM patients_vital_info')  
        result = cursor.fetchall()  
        return result  
    except Error as e:  
        print(f"Error while fetching data: {e}")  
        return None  
    finally:  
        if connection.is_connected():  
            cursor.close()  
  
    connection.close()
```

4. Kafka Producer Function

The `produce_to_kafka` function iterates over the fetched data records. Each record is serialized and sent to the Kafka topic. The function includes error handling and rate control to prevent overload.

```
def produce_to_kafka(data, topic):  
    if not data:  
        print("No data to produce.")  
        return  
    try:  
        for row in data:  
            data_dict = {  
                'customerId': row[0],  
                'heartBeat': row[1],  
                'bp': row[2],  
            }  
            print(data_dict)  
            producer.send(topic, data_dict)  
            time.sleep(1)  
        producer.flush()
```

5. Main Execution

The main section initiates the process by fetching data and passing it to the `produce_to_kafka` function to stream the information to the Kafka topic

```
if __name__ == "__main__":  
    data = fetch_data(HOST, DATABASE, USER, PASSWORD)  
    produce_to_kafka(data, TOPIC)
```

6. Setting Up Apache Kafka on an EMR Cluster

This guide provides step-by-step instructions to download, install, and configure Apache Kafka on an AWS EMR cluster. Additionally, it includes steps to create Kafka topics and install required Python libraries

6.1 Connect to your EMR cluster via SSH

[illegible]

6.2 Download Kafka using the wget command

```
hadoop@ip-172.31.5.58:~$ wget https://downloads.apache.org/kafka/3.6.2/kafka_2.12-3.6.2.tgz
--2024-11-05 07:15:56-- https://downloads.apache.org/kafka/3.6.2/kafka_2.12-3.6.2.tgz
Resolving downloads.apache.org (downloads.apache.org)... 88.99.208.237, 135.181.214.104, 2a01:4f9:3a:2c57::2, ...
Connecting to downloads.apache.org (downloads.apache.org)[88.99.208.237]:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 113986658 (109M) [application/x-gzip]
Saving to: 'kafka_2.12-3.6.2.tgz'

64% [=====] 73,135,871 22.2MB/s eta 2s
```

6.3 : Extract the Kafka Files

```

kafka_2.12-3.6.2/Libs/jetty-client-9.4.53.v20231009.jar
kafka_2.12-3.6.2/Libs/maven-artifact-3.8.8.jar
kafka_2.12-3.6.2/Libs/swagger-annotations-2.2.8.jar
kafka_2.12-3.6.2/Libs/jersey-container-servlet-core-2.39.1.jar
kafka_2.12-3.6.2/Libs/jersey-server-2.39.1.jar
kafka_2.12-3.6.2/Libs/jersey-client-2.39.1.jar
kafka_2.12-3.6.2/Libs/jersey-common-2.39.1.jar
kafka_2.12-3.6.2/Libs/jakarta.ws.rs-api-2.1.6.jar
kafka_2.12-3.6.2/Libs/hk2-locator-2.6.1.jar
kafka_2.12-3.6.2/Libs/javassist-3.29.2-GA.jar
kafka_2.12-3.6.2/Libs/javax.activation-api-1.2.0.jar
kafka_2.12-3.6.2/Libs/javax.servlet-api-3.1.0.jar
kafka_2.12-3.6.2/Libs/jetty-http-9.4.53.v20231009.jar
kafka_2.12-3.6.2/Libs/jetty-io-9.4.53.v20231009.jar
kafka_2.12-3.6.2/Libs/jetty-util-ajax-9.4.53.v20231009.jar
kafka_2.12-3.6.2/Libs/jetty-continuation-9.4.53.v20231009.jar
kafka_2.12-3.6.2/Libs/jetty-util-9.4.53.v20231009.jar
kafka_2.12-3.6.2/Libs/jsr305-3.0.2.jar
kafka_2.12-3.6.2/Libs/plexus-utils-3.3.1.jar
kafka_2.12-3.6.2/Libs/commons-lang3-3.8.1.jar
kafka_2.12-3.6.2/Libs/hk2-api-2.6.1.jar
kafka_2.12-3.6.2/Libs/hk2-utils-2.6.1.jar
kafka_2.12-3.6.2/Libs/jakarta.inject-2.6.1.jar
kafka_2.12-3.6.2/Libs/jakarta.annotation-api-1.3.5.jar
kafka_2.12-3.6.2/Libs/opsi-resource-locator-1.0.2.jar
kafka_2.12-3.6.2/Libs/jakarta.validation-api-2.0.2.jar
kafka_2.12-3.6.2/Libs/aopalliance-repackaged-2.6.1.jar
kafka_2.12-3.6.2/Libs/troldor-3.6.2.jar
kafka_2.12-3.6.2/Libs/kafka-shell-3.6.2.jar
kafka_2.12-3.6.2/Libs/jline-3.25.1.jar
kafka_2.12-3.6.2/Libs/connect-file-3.6.2.jar
kafka_2.12-3.6.2/Libs/connect-basic-auth-extension-3.6.2.jar
kafka_2.12-3.6.2/Libs/connect-mirror-3.6.2.jar
kafka_2.12-3.6.2/Libs/connect-mirror-client-3.6.2.jar
kafka_2.12-3.6.2/Libs/kafka-streams-3.6.2.jar
kafka_2.12-3.6.2/Libs/rocksdbjni-7.9.2.jar
kafka_2.12-3.6.2/Libs/kafka-streams-scala_2.12-3.6.2.jar
kafka_2.12-3.6.2/Libs/kafka-streams-test-utils-3.6.2.jar
kafka_2.12-3.6.2/Libs/kafka-streams-examples-3.6.2.jar
[hadoop@ip-172-31-5-58 ~]$

```

6.4 : Configure Kafka Server

Update the advertised listeners to use our EMR IP address

vi config/server.properties

```

# (the "License"); you may not use this file except in compliance with
# the License. You may obtain a copy of the License at
#
# http://www.apache.org/licenses/LICENSE-2.0
#
# Unless required by applicable law or agreed to in writing, software
# distributed under the License is distributed on an "AS IS" BASIS,
# WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
# See the License for the specific language governing permissions and
# limitations under the License.
#
# This configuration file is intended for use in ZK-based mode, where Apache ZooKeeper is required.
# See kafka.server.KafkaConfig for additional details and defaults
#
##### Server Basics #####
# The id of the broker. This must be set to a unique integer for each broker.
broker.id=0
##### Socket Server Settings #####
# The address the socket server listens on. If not configured, the host name will be equal to the value of
# java.net.InetAddress.getCanonicalHostName(), with PLAINTEXT listener name, and port 9092.
#
# FORMAT:
#   listeners = listener_name://host_name:port
#
# EXAMPLE:
#   listeners = PLAINTEXT://your.host.name:9092
#listeners=PLAINTEXT://:9092
#
# Listener name, hostname and port the broker will advertise to clients.
# If not set, it uses the value for "listeners".
advertised.listeners=PLAINTEXT://ec2-3-239-186-188.compute-1.amazonaws.com:9092
#
# Maps listener names to security protocols, the default is for them to be the same. See the config documentation for more details
#listener.security.protocol.map=PLAINTEXT:PLAINTEXT,SSL:SSL,SASL_PLAINTEXT:SASL_PLAINTEXT,SASL_SSL:SASL_SSL
#
# The number of threads that the server uses for receiving requests from the network and sending responses to the network
num.network.threads=3

```

6.5 Start the Zookeeper

bin/zookeeper-server-start.sh config/zookeeper.properties

```

[2024-11-05 07:19:03,499] INFO zookeeper.flushDelay = 0 ms (org.apache.zookeeper.server.ZooKeeperServer)
[2024-11-05 07:19:03,499] INFO zookeeper.maxWriteQueuePollTime = 0 ms (org.apache.zookeeper.server.ZooKeeperServer)
[2024-11-05 07:19:03,499] INFO zookeeper.maxBatchSize=1000 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-11-05 07:19:03,499] INFO zookeeper.intBufferStartingSizeBytes = 1024 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-11-05 07:19:03,500] INFO Weighed connection throttling is disabled (org.apache.zookeeper.server.BlueThrottle)
[2024-11-05 07:19:03,502] INFO minSessionTimeout set to 6000 ms (org.apache.zookeeper.server.ZooKeeperServer)
[2024-11-05 07:19:03,502] INFO maxSessionTimeout set to 60000 ms (org.apache.zookeeper.server.ZooKeeperServer)
[2024-11-05 07:19:03,502] INFO getData response cache size is initialized with value 400. (org.apache.zookeeper.server.ResponseCache)
[2024-11-05 07:19:03,503] INFO getChildren response cache size is initialized with value 400. (org.apache.zookeeper.server.ResponseCache)
[2024-11-05 07:19:03,504] INFO zookeeper.pathStats.slotCapacity = 60 (org.apache.zookeeper.server.util.RequestPathMetricsCollector)
[2024-11-05 07:19:03,504] INFO zookeeper.pathStats.slotDuration = 15 (org.apache.zookeeper.server.util.RequestPathMetricsCollector)
[2024-11-05 07:19:03,504] INFO zookeeper.pathStats.maxDepth = 6 (org.apache.zookeeper.server.util.RequestPathMetricsCollector)
[2024-11-05 07:19:03,505] INFO zookeeper.pathStats.initialDelay = 5 (org.apache.zookeeper.server.util.RequestPathMetricsCollector)
[2024-11-05 07:19:03,505] INFO zookeeper.pathStats.delay = 5 (org.apache.zookeeper.server.util.RequestPathMetricsCollector)
[2024-11-05 07:19:03,505] INFO zookeeper.pathStats.enabled = false (org.apache.zookeeper.server.util.RequestPathMetricsCollector)
[2024-11-05 07:19:03,508] INFO The max bytes for all large requests are set to 104857600 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-11-05 07:19:03,508] INFO The large request threshold is set to -1 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-11-05 07:19:03,508] INFO zookeeper.enforce.auth.enabled = false (org.apache.zookeeper.server.AuthenticationHelper)
[2024-11-05 07:19:03,508] INFO zookeeper.enforce.auth.schemes = [] (org.apache.zookeeper.server.AuthenticationHelper)
[2024-11-05 07:19:03,509] INFO Created server with tickTime 3000 ms minSessionTimeout 6000 ms maxSessionTimeout 60000 ms clientPortListenBacklog -1 datadir /tmp/zookeeper/version-2 snapdir /tmp/zookeeper/version-2 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-11-05 07:19:03,516] INFO Using org.apache.zookeeper.server.NIOServerCnxnFactory as server connection factory (org.apache.zookeeper.server.ServerCnxnFactory)
[2024-11-05 07:19:03,517] WARN maxCnxns is not configured, using default value 0. (org.apache.zookeeper.server.ServerCnxnFactory)
[2024-11-05 07:19:03,519] INFO Configuring NIO connection handler with 10s sessionless connection timeout, 1 selector thread(s), 8 worker threads, and 64 kB direct buffers. (org.apache.zookeeper.server.NIOServerCnxnFactory)
[2024-11-05 07:19:03,525] INFO Binding to port 0.0.0.0/0.0.0.0:2181 (org.apache.zookeeper.server.NIOServerCnxnFactory)
[2024-11-05 07:19:03,551] INFO Using org.apache.zookeeper.server.watch.WatchManager as watch manager (org.apache.zookeeper.server.watch.WatchManagerFactory)
[2024-11-05 07:19:03,551] INFO Using org.apache.zookeeper.server.watch.WatchManager as watch manager (org.apache.zookeeper.server.watch.WatchManagerFactory)
[2024-11-05 07:19:03,551] INFO zookeeper.snapshotSizeFactor = 0.33 (org.apache.zookeeper.server.ZKDatabase)
[2024-11-05 07:19:03,551] INFO zookeeper.commitLogCount=500 (org.apache.zookeeper.server.ZKDatabase)
[2024-11-05 07:19:03,559] INFO zookeeper.snapshot.compression.method = CHECKED (org.apache.zookeeper.server.persistence.SnapStream)
[2024-11-05 07:19:03,559] INFO Snapshotting: 0x0 to /tmp/zookeeper/version-2/snapshot.0 (org.apache.zookeeper.server.persistence.FileTxnSnapLog)
[2024-11-05 07:19:03,563] INFO Snapshot loaded in 12 ms, highest zxid is 0x0, digest is 1371985504 (org.apache.zookeeper.server.ZKDatabase)
[2024-11-05 07:19:03,565] INFO Snapshotting: 0x0 to /tmp/zookeeper/version-2/snapshot.0 (org.apache.zookeeper.server.persistence.FileTxnSnapLog)
[2024-11-05 07:19:03,566] INFO Snapshot taken in 0 ms (org.apache.zookeeper.server.ZooKeeperServer)
[2024-11-05 07:19:03,567] INFO ZooKeeper started, configuration: false (org.apache.zookeeper.server.PreRequestProcessor)
[2024-11-05 07:19:03,578] INFO zookeeper.request.throttler.shutdownTimeout = 10000 ms (org.apache.zookeeper.server.RequestThrottler)
[2024-11-05 07:19:03,601] INFO Using checkIntervalMs=60000 ms maxPerMinute=10000 ms neverUsedIntervalMs=0 (org.apache.zookeeper.server.ContainerManager)
[2024-11-05 07:19:03,602] INFO ZooKeeper audit is disabled. (org.apache.zookeeper.audit.ZKAuditProvider)

```

6.6 Start the server

bin/kafka-server-start.sh config/server.properties

```

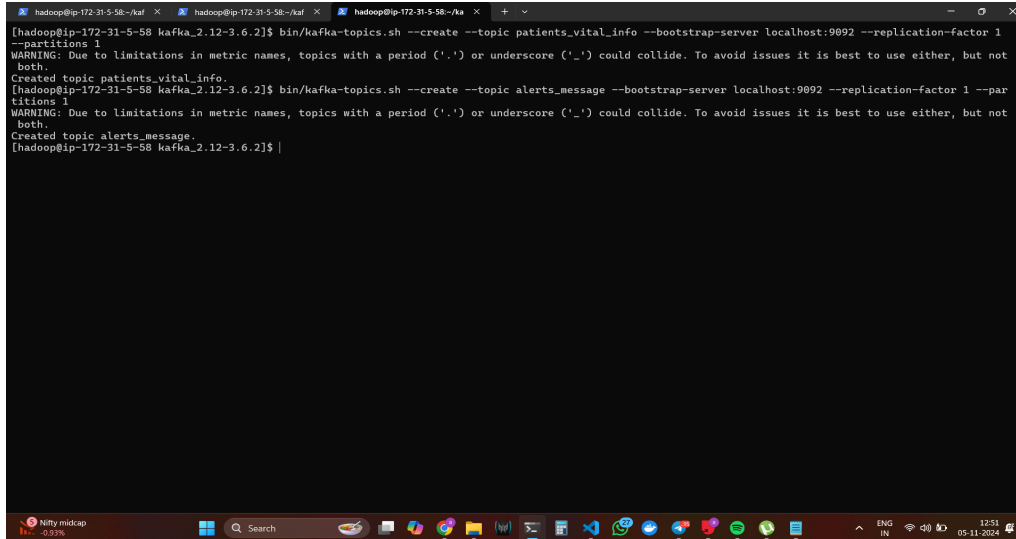
[2024-11-05 07:19:57,121] INFO [ExpirationReaper-0-Produce]: Starting (kafka.server.DelayedOperationPurgatory$ExpiredOperationReaper)
[2024-11-05 07:19:57,122] INFO [ExpirationReaper-0-Fetch]: Starting (kafka.server.DelayedOperationPurgatory$ExpiredOperationReaper)
[2024-11-05 07:19:57,122] INFO [ExpirationReaper-0-DeleteRecords]: Starting (kafka.server.DelayedOperationPurgatory$ExpiredOperationReaper)
[2024-11-05 07:19:57,125] INFO [ExpirationReaper-0-ElectLeader]: Starting (kafka.server.DelayedOperationPurgatory$ExpiredOperationReaper)
[2024-11-05 07:19:57,125] INFO [ExpirationReaper-0-RemoteFetch]: Starting (kafka.server.DelayedOperationPurgatory$ExpiredOperationReaper)
[2024-11-05 07:19:57,155] INFO [LogDirFailureHandler]: Starting (kafka.server.ReplicaManager$LogDirFailureHandler)
[2024-11-05 07:19:57,158] INFO [AddPartitionsToTxnSenderThread-0]: Starting (kafka.server.AddPartitionsToTxnManager)
[2024-11-05 07:19:57,186] INFO Creating /brokers/ids/0 (is it secure? false) (kafka.zk.KafkaZkClient)
[2024-11-05 07:19:57,234] INFO State of the created znode at /brokers/ids/0 is: 25,25,1730791197222,1730791197222,1,0,0,72057628895739904,266,0,25 (kafka.zk.KafkaZkClient)
[2024-11-05 07:19:57,235] INFO Registered broker 0 at path /brokers/ids/0 with addresses: PLAINTEXT://ec2-3-239-186-180.compute-1.amazonaws.com:9092, cxxid (broker epoch): 25 (kafka.zk.KafkaZkClient)
[2024-11-05 07:19:57,332] INFO [ExpirationReaper-0-topic]: Starting (kafka.server.DelayedOperationPurgatory$ExpiredOperationReaper)
[2024-11-05 07:19:57,340] INFO Successfully created /controller epoch with initial epoch 0 (kafka.zk.KafkaZkClient)
[2024-11-05 07:19:57,341] INFO [ExpirationReaper-0-Heartbeat]: Starting (kafka.server.DelayedOperationPurgatory$ExpiredOperationReaper)
[2024-11-05 07:19:57,342] INFO [ExpirationReaper-0-Rebalance]: Starting (kafka.server.DelayedOperationPurgatory$ExpiredOperationReaper)
[2024-11-05 07:19:57,362] INFO Feature ZK node created at path: /feature (kafka.server.FinalizedFeatureChangeListener)
[2024-11-05 07:19:57,363] INFO [GroupCoordinator 0]: Starting up. (kafka.coordinator.group.GroupCoordinator)
[2024-11-05 07:19:57,391] INFO [GroupCoordinator 0]: Startup complete. (kafka.coordinator.group.GroupCoordinator)
[2024-11-05 07:19:57,423] INFO [TransactionCoordinator id=0] Starting up. (kafka.coordinator.transaction.TransactionCoordinator)
[2024-11-05 07:19:57,431] INFO [MetadataCache brokerId=0] Updated cache from existing None to latest Features(version=3.6-IV2, finalizedFeatures={}, finalizedFeaturesEpoch=0) (kafka.server.metadata.ZkMetadataCache)
[2024-11-05 07:19:57,439] INFO [TxnMarkerSenderThread-0]: Starting (kafka.coordinator.transaction.TransactionMarkerChannelManager)
[2024-11-05 07:19:57,439] INFO [TransactionCoordinator id=0] Startup complete. (kafka.coordinator.transaction.TransactionCoordinator)
[2024-11-05 07:19:57,502] INFO [ExpirationReaper-0-AlterAcls]: Starting (kafka.server.DelayedOperationPurgatory$ExpiredOperationReaper)
[2024-11-05 07:19:57,540] INFO [Controller id=0, targetBrokerId=0] Node 0 disconnected. (org.apache.kafka.clients.NetworkClient)
[2024-11-05 07:19:57,540] WARN [Controller id=0, targetBrokerId=0] Connection to node 0 (ec2-3-239-186-180.compute-1.amazonaws.com/172.31.5.58:9092) could not be established. Broker may not be available. (org.apache.kafka.clients.NetworkClient)
[2024-11-05 07:19:57,557] INFO [Controller id=0, targetBrokerId=0] Client requested connection close from node 0 (org.apache.kafka.clients.NetworkClient)
[2024-11-05 07:19:57,567] INFO [/config/changes-event-process-thread]: Starting (kafka.common.ZkNodeChangeNotificationListener$ChangeEventProcessThread)
[2024-11-05 07:19:57,592] INFO [SocketServer listenerType=ZK_BROKER, nodeId=0] Enabling request processing. (kafka.network.SocketServer)
[2024-11-05 07:19:57,597] INFO [SocketServer listenerType=ZK_BROKER, nodeId=0] Awaiting socket connections on 0.0.0.0:9092. (kafka.network.DataPlaneAcceptor)
[2024-11-05 07:19:57,641] INFO Kafka version: 3.6.2 (org.apache.kafka.common.utils.AppInfoParser)
[2024-11-05 07:19:57,642] INFO Kafka commitId: c0deed513057c94e (org.apache.kafka.common.utils.AppInfoParser)
[2024-11-05 07:19:57,642] INFO Kafka startTimes: 1730791197627 (org.apache.kafka.common.utils.AppInfoParser)
[2024-11-05 07:19:57,644] INFO [KafkaServer id=0] started (kafka.server.KafkaServer)
[2024-11-05 07:19:57,790] INFO [zk-broker-0-to-controller-alter-partition-channel-manager]: Recorded new controller, from now on will use node ec2-3-239-186-180.compute-1.amazonaws.com:9092 (id: 0 rack: null) (kafka.server.BrokerToControllerRequestThread)
[2024-11-05 07:19:57,792] INFO [zk-broker-0-to-controller-forwarding-channel-manager]: Recorded new controller, from now on will use node ec2-3-239-186-180.compute-1.amazonaws.com:9092 (id: 0 rack: null) (kafka.server.BrokerToControllerRequestThread)

```

6.7 Create Kafka Topics

```
bin/kafka-topics.sh --create --topic patients_vital_info --bootstrap-server localhost:9092  
--replication-factor 1 --partitions 1
```

```
bin/kafka-topics.sh --create --topic alerts_message --bootstrap-server localhost:9092  
--replication-factor 1 --partitions 1
```



```
hadoop@ip-172-31-5-58:~$ bin/kafka-topics.sh --create --topic patients_vital_info --bootstrap-server localhost:9092 --replication-factor 1 --partitions 1
WARNING: Due to limitations in metric names, topics with a period ('.') or underscore ('_') could collide. To avoid issues it is best to use either, but not both.
Created topic patients_vital_info.
hadoop@ip-172-31-5-58:~$ bin/kafka-topics.sh --create --topic alerts_message --bootstrap-server localhost:9092 --replication-factor 1 --partitions 1
WARNING: Due to limitations in metric names, topics with a period ('.') or underscore ('_') could collide. To avoid issues it is best to use either, but not both.
Created topic alerts_message.
hadoop@ip-172-31-5-58:~$
```

7. Install the required dependencies

`pip install mysql-connector-python kafka-python`

```
pip install mysql-connector-python
pip install kafka-python
```

7.1. Run the python producer script

`python kafka_produce_patient_vitals.py`

This python script reads data from RDS per second and simulates IoT devices to push patient vital data in the JSON format into the Kafka Queue.

Screenshot of JSON vital information written to the console (per second)

```
Connected to MySQL database
{'customerId': 1, 'heartBeat': 74, 'bp': 202}
{'customerId': 2, 'heartBeat': 68, 'bp': 173}
{'customerId': 3, 'heartBeat': 71, 'bp': 152}
{'customerId': 4, 'heartBeat': 72, 'bp': 166}
{'customerId': 5, 'heartBeat': 68, 'bp': 171}
{'customerId': 1, 'heartBeat': 70, 'bp': 189}
{'customerId': 2, 'heartBeat': 72, 'bp': 173}
{'customerId': 3, 'heartBeat': 68, 'bp': 178}
{'customerId': 4, 'heartBeat': 71, 'bp': 152}
{'customerId': 5, 'heartBeat': 73, 'bp': 166}
{'customerId': 1, 'heartBeat': 74, 'bp': 185}
{'customerId': 2, 'heartBeat': 67, 'bp': 177}
{'customerId': 3, 'heartBeat': 66, 'bp': 158}
{'customerId': 4, 'heartBeat': 71, 'bp': 177}
{'customerId': 5, 'heartBeat': 66, 'bp': 155}
{'customerId': 1, 'heartBeat': 71, 'bp': 220}
```