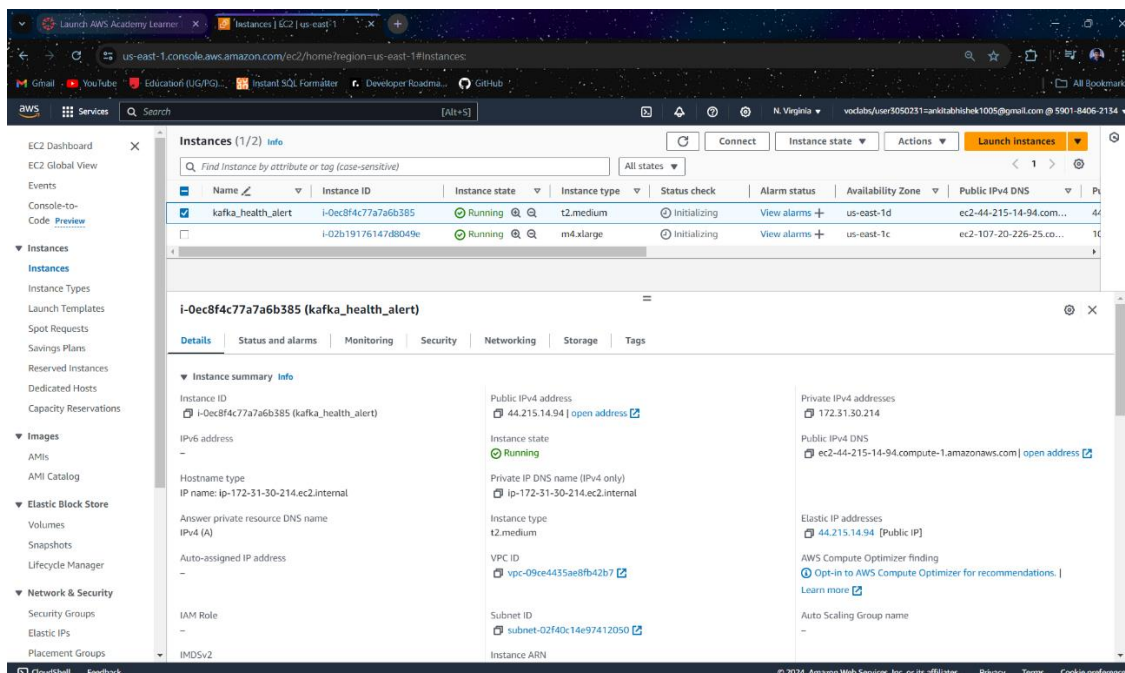


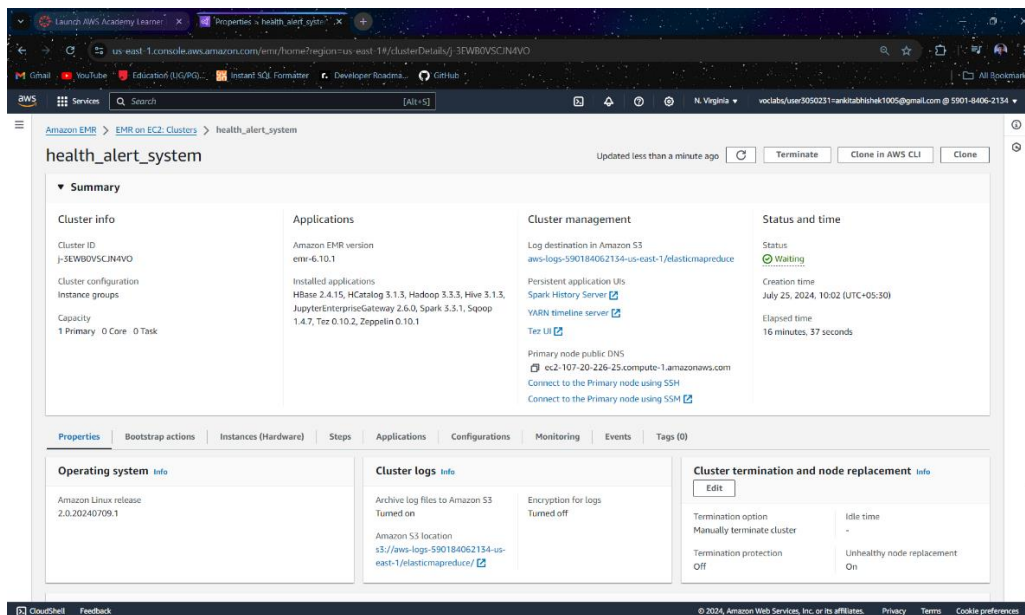
CODE LOGIC

STEP1: CREATION OF KAFKA CLUSTER AND EMR CLUSTER

1. Create a kafka cluster with the help of ec2 instance. (Referred with PDF attached in the Apache Kafka modules). Make the required configuration needed to run kafka instance. Kafka is pre-installed on below mentioned ec2 cluster with the selection of **ami-06c41d8b5a6ddd3c2** while creating **Amazon Machine Image** as pdf within modules.



2. Create an EMR instance with required below mentioned libraries (Referred with PDF attached in the modules).
Spark 3.3.1, Sqoop 1.4.7, HBase 2.4.15, HCatalog 3.1.3, Hadoop 3.3.3, Hive 3.1.3, JupyterEnterpriseGateway 2.6.0, Zeppelin 0.10.1



INSTALLING REQUIRED PACKAGES ON KAFKA CLUSTER

Sudo pip3 install kafka-python
Sudo pip3 install mysql-connector
Sudo pip3 install boto3

```
ec2-user@ip-172-31-30-214:~
[ec2-user@ip-172-31-30-214 ~]$ sudo pip3 install kafka-python
WARNING: Running pip install with root privileges is generally not a good idea.
Try 'pip3 install --user' instead.
Collecting kafka-python
  Downloading kafka-python-2.0.2-py2.py3-none-any.whl (246 kB)
    |████████████████████| 246 kB 36.2 MB/s
Installing collected packages: kafka-python
Successfully installed kafka-python-2.0.2
[ec2-user@ip-172-31-30-214 ~]$ sudo pip3 install mysql-connector
WARNING: Running pip install with root privileges is generally not a good idea.
Try 'pip3 install --user' instead.
Collecting mysql-connector
  Downloading mysql-connector-2.2.9.tar.gz (11.9 MB)
    |████████████████████| 11.9 MB 69 kB/s
Using legacy 'setup.py install' for mysql-connector, since package 'wheel' is not installed.
Installing collected packages: mysql-connector
  Running setup.py install for mysql-connector ... done
Successfully installed mysql-connector-2.2.9
[ec2-user@ip-172-31-30-214 ~]$ sudo pip3 install boto3
WARNING: Running pip install with root privileges is generally not a good idea.
Try 'pip3 install --user' instead.
Collecting boto3
  Downloading boto3-1.33.13-py3-none-any.whl (139 kB)
    |████████████████████| 139 kB 13.2 MB/s
```

```
ec2-user@ip-172-31-30-214:~  
Try `pip3 install --user` instead.  
Collecting boto3  
  Downloading boto3-1.33.13-py3-none-any.whl (139 kB)  
    |████████████████████████████████████████| 139 kB 13.2 MB/s  
Collecting s3transfer<0.9.0,>=0.8.2  
  Downloading s3transfer-0.8.2-py3-none-any.whl (82 kB)  
    |████████████████████████████████████████| 82 kB 122 kB/s  
Collecting jmespath<2.0.0,>=0.7.1  
  Downloading jmespath-1.0.1-py3-none-any.whl (20 kB)  
Collecting botocore<1.34.0,>=1.33.13  
  Downloading botocore-1.33.13-py3-none-any.whl (11.8 MB)  
    |████████████████████████████████████████| 11.8 MB 35 kB/s  
Collecting python-dateutil<3.0.0,>=2.1  
  Downloading python_dateutil-2.9.0.post0-py2.py3-none-any.whl (229 kB)  
    |████████████████████████████████████████| 229 kB 59.4 MB/s  
Collecting urllib3<1.27,>=1.25.4; python_version < "3.10"  
  Downloading urllib3-1.26.19-py2.py3-none-any.whl (143 kB)  
    |████████████████████████████████████████| 143 kB 63.5 MB/s  
Collecting six>=1.5  
  Downloading six-1.16.0-py2.py3-none-any.whl (11 kB)  
Installing collected packages: six, python-dateutil, urllib3, jmespath, botocore, s3transfer, boto3  
Successfully installed boto3-1.33.13 botocore-1.33.13 jmespath-1.0.1 python-dateutil-2.9.0.post0 s3transfer-0.8.2 six-1.16.0 urllib3-1.26.19  
[ec2-user@ip-172-31-30-214 ~]$
```

STATEMENT FOR STARTING KAFKA SERVER

1. STARTING ZOOKEEPER SERVER:

Inside cd downloads/kafka 2.12-2.3.0 run

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

```
ec2-user@ip-172-31-30-214:~/downloads/kafka_2.12-2.3.0
```

```
[ec2-user@ip-172-31-30-214 kafka_2.12-2.3.0]$ bin/zookeeper-server-start.sh config/zookeeper.properties
[2024-07-25 04:53:01,268] INFO Reading configuration from: config/zookeeper.properties (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-07-25 04:53:01,457] INFO autopurge.snapRetainCount set to 3 (org.apache.zookeeper.server.DataDirCleanupManager)
[2024-07-25 04:53:01,457] INFO autopurge.purgeInterval set to 0 (org.apache.zookeeper.server.DataDirCleanupManager)
[2024-07-25 04:53:01,458] INFO Purge task is not scheduled. (org.apache.zookeeper.server.DataDirCleanupManager)
[2024-07-25 04:53:01,458] WARN Either no config or no quorum defined in config, running in standalone mode (org.apache.zookeeper.server.quorum.QuorumPeerMain)
[2024-07-25 04:53:01,560] INFO Reading configuration from: config/zookeeper.properties (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-07-25 04:53:01,560] INFO Starting server (org.apache.zookeeper.server.ZooKeeperServerMain)
[2024-07-25 04:53:01,683] INFO Server environment:zookeeper.version=3.4.14-4c25d480e66aadd371de8bd2fd8da255ac140bcf, built on 03/06/2019 16:18 GMT (org.apache.zookeeper.server.ZooKeeperServer)
[2024-07-25 04:53:01,683] INFO Server environment:host.name=ip-172-31-30-214.ec2.internal (org.apache.zookeeper.server.ZooKeeperServer)
[2024-07-25 04:53:01,683] INFO Server environment:java.version=1.8.0_161 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-07-25 04:53:01,684] INFO Server environment:java.vendor=Oracle Corporation (org.apache.zookeeper.server.ZooKeeperServer)
```

```
ec2-user@ip-172-31-30-214:~/downloads/kafka_2.12-2.3.0
[2024-07-25 04:53:01,685] INFO Server environment:java.compiler=<NA> (org.apache.zookeeper.server.ZooKeeperServer)
[2024-07-25 04:53:01,685] INFO Server environment:os.name=Linux (org.apache.zookeeper.server.ZooKeeperServer)
[2024-07-25 04:53:01,685] INFO Server environment:os.arch=amd64 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-07-25 04:53:01,685] INFO Server environment:os.version=4.14.193-149.317.amzn2.x86_64 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-07-25 04:53:01,685] INFO Server environment:user.name=ec2-user (org.apache.zookeeper.server.ZooKeeperServer)
[2024-07-25 04:53:01,685] INFO Server environment:user.home=/home/ec2-user (org.apache.zookeeper.server.ZooKeeperServer)
[2024-07-25 04:53:01,685] INFO Server environment:user.dir=/home/ec2-user/downloads/kafka_2.12-2.3.0 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-07-25 04:53:02,135] INFO tickTime set to 3000 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-07-25 04:53:02,135] INFO minSessionTimeout set to -1 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-07-25 04:53:02,136] INFO maxSessionTimeout set to -1 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-07-25 04:53:02,760] INFO Using org.apache.zookeeper.server.NIOServerCnxnFactory as server connection factory (org.apache.zookeeper.server.ServerCnxnFactory)
[2024-07-25 04:53:03,150] INFO binding to port 0.0.0.0/0.0.0.0:2181 (org.apache.zookeeper.server.NIOServerCnxnFactory)
```

2. STARTING KAFKA SERVER:

Into another putty Session of kafka cluster inside cd downloads/kafka_2.12-2.3.0 run **bin/kafka-server-start.sh config/server.properties**

```
ec2-user@ip-172-31-30-214:~/downloads/kafka_2.12-2.3.0
[ec2-user@ip-172-31-30-214 kafka_2.12-2.3.0]$ bin/kafka-server-start.sh config/server.properties
[2024-07-25 04:53:11,706] INFO Registered kafka:type=kafka.Log4jController MBean (kafka.utils.Log4jControllerRegistration$)
[2024-07-25 04:53:19,457] INFO Registered signal handlers for TERM, INT, HUP (org.apache.kafka.common.utils.LoggingSignalHandler)
[2024-07-25 04:53:19,458] INFO starting (kafka.server.KafkaServer)
[2024-07-25 04:53:19,460] INFO Connecting to zookeeper on localhost:2181 (kafka.server.KafkaServer)
[2024-07-25 04:53:19,862] INFO [ZooKeeperClient Kafka server] Initializing a new session to localhost:2181. (kafka.zookeeper.ZooKeeperClient)
[2024-07-25 04:53:19,877] INFO Client environment:zookeeper.version=3.4.14-4c25d480e66aadd371de8bd2fd8da255ac140bcf, built on 03/06/2019 16:18 GMT (org.apache.zookeeper.ZooKeeper)
[2024-07-25 04:53:19,877] INFO Client environment:host.name=ip-172-31-30-214.ec2.internal (org.apache.zookeeper.ZooKeeper)
[2024-07-25 04:53:19,877] INFO Client environment:java.version=1.8.0_161 (org.apache.zookeeper.ZooKeeper)
[2024-07-25 04:53:19,877] INFO Client environment:java.vendor=Oracle Corporation (org.apache.zookeeper.ZooKeeper)
[2024-07-25 04:53:19,877] INFO Client environment:java.home=/usr/java/jdk1.8.0_161/jre (org.apache.zookeeper.ZooKeeper)
[2024-07-25 04:53:19,877] INFO Client environment:java.class.path=/home/ec2-user/downloads/kafka_2.12-2.3.0/bin/../libs/activation-1.1.1.jar:/home/ec2-user/downloads/kafka_2.12-2.3.0/bin/../libs/aopalliance-repackaged-2.5.0.jar:/home/ec2-user
```

```
ec2-user@ip-172-31-30-214:~/downloads/kafka_2.12-2.3.0
ordinator.group.GroupCoordinator)
[2024-07-25 04:53:31,523] INFO [GroupMetadataManager brokerId=0] Removed 0 expired offsets in 63 milliseconds. (kafka.coordinator.group.GroupMetadataManager)
[2024-07-25 04:53:31,552] INFO [ProducerId Manager 0]: Acquired new producerId block (brokerId:0,blockStartProducerId:0,blockEndProducerId:999) by writing to Zk with path version 1 (kafka.coordinator.transaction.ProducerIdManager)
[2024-07-25 04:53:31,615] INFO [TransactionCoordinator id=0] Starting up. (kafka.coordinator.transaction.TransactionCoordinator)
[2024-07-25 04:53:31,623] INFO [TransactionCoordinator id=0] Startup complete. (kafka.coordinator.transaction.TransactionCoordinator)
[2024-07-25 04:53:31,629] INFO [Transaction Marker Channel Manager 0]: Starting (kafka.coordinator.transaction.TransactionMarkerChannelManager)
[2024-07-25 04:53:31,853] INFO [/config/changes-event-process-thread]: Starting (kafka.common.ZkNodeChangeNotificationListener$ChangeEventProcessThread)
[2024-07-25 04:53:31,876] INFO [SocketServer brokerId=0] Started data-plane processors for 1 acceptors (kafka.network.SocketServer)
[2024-07-25 04:53:31,920] INFO Kafka version: 2.3.0 (org.apache.kafka.common.utils.AppInfoParser)
[2024-07-25 04:53:31,922] INFO Kafka commitId: fclaaa116b661c8a (org.apache.kafka.common.utils.AppInfoParser)
[2024-07-25 04:53:31,924] INFO Kafka startTimeMs: 1721883211877 (org.apache.kafka.common.utils.AppInfoParser)
[2024-07-25 04:53:31,926] INFO [KafkaServer id=0] started (kafka.server.KafkaServer)
```

STATEMENT TO CREATE TOPICS

To create topic in kafka server, the command used is

bin/kafka-topics.sh --create --bootstrap-server localhost:9092 --replication-factor 1 --partitions 1 --topic Patients-Vital-Info

STATEMENT TO LIST TOPICS

To list the created topic inside cd downloads/kafka_2.12-2.3.0, the command used is

bin/kafka-topics.sh --list --bootstrap-server localhost:9092

```
ec2-user@ip-172-31-30-214:~/downloads/kafka_2.12-2.3.0
[ec2-user@ip-172-31-30-214 ~]$ cd downloads/kafka_2.12-2.3.0
[ec2-user@ip-172-31-30-214 kafka_2.12-2.3.0]$ bin/kafka-topics.sh --list --bootstrap-server localhost:9092

[ec2-user@ip-172-31-30-214 kafka_2.12-2.3.0]$ bin/kafka-topics.sh --create --bootstrap-server localhost:9092 --replication-factor 1 --partitions 1 --topic Patients-Vital-Info
[ec2-user@ip-172-31-30-214 kafka_2.12-2.3.0]$ bin/kafka-topics.sh --list --bootstrap-server localhost:9092
Patients-Vital-Info
[ec2-user@ip-172-31-30-214 kafka_2.12-2.3.0]$
```


STEP 2: EXECUTING PRODUCER APPLICATION AND CONSUMER APPLICATION:

Producer application which is file named as **kafka_produce_patient_vitals.py** is built on the **python language** which will consume data residing on rds with below mentioned credentials:

Hostname = "upgraddetest.cyaiehc9bmnf.us-east-1.rds.amazonaws.com"

username = "student"

password = "STUDENT123"

dbname = "testdatabase".

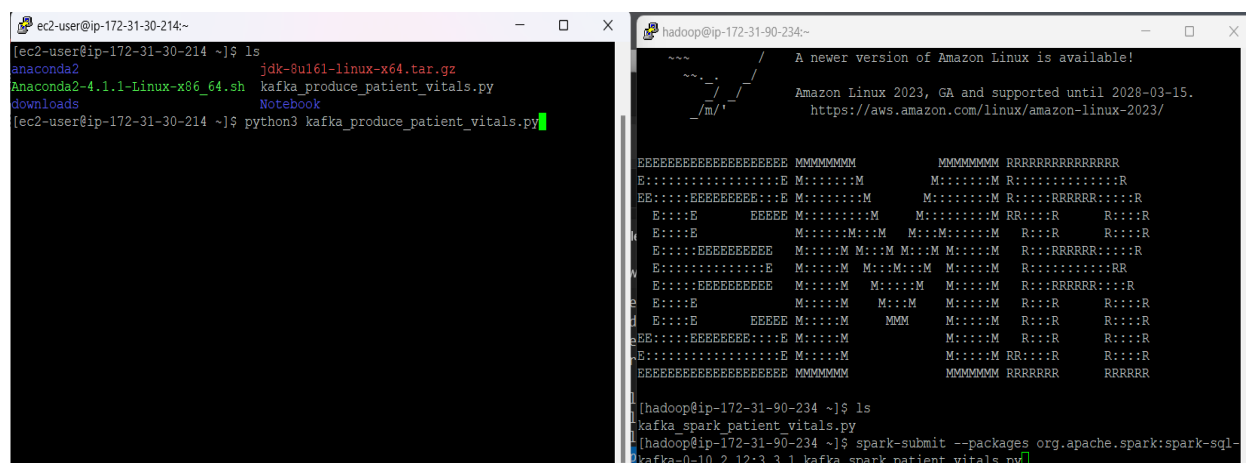
Consumer Application which is file named as **kafka_spark_patient_vitals.py** is built on the **Apache PySpark** language which will consume data being produced with the help of above mentioned producer application

NOTE: Run the producer application on ec2 Kafka cluster after starting the consumer application on EMR cluster created with Spark, Hive and another libraries

STATEMENT FOR EXECUTING PRODUCER APPLICATION AND CONSUMER APPLICATION

Spark Submitting Job to Consume Message from The Topic Patients-Vital-Info And Stored To HDFS Location

For Producer application: **python3 kafka_produce_patients_vitals.py**



The image shows two terminal windows. The left window is an EC2 instance terminal with IP 172-31-30-214. It shows the user installing Anaconda2, downloading the Java JDK, and then running the producer application `python3 kafka_produce_patient_vitals.py`. The right window is a Hadoop terminal with IP 172-31-90-234. It shows a message about a newer version of Amazon Linux, followed by a large ASCII art graphic. Below the graphic, it shows the user running `ls` and then submitting a Spark job with `spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.12:3.3.1 kafka_spark_patient_vitals.py`.

```

ec2-user@ip-172-31-30-214:~$ ls
anaconda2                                jdk-8u161-linux-x64.tar.gz
Anaconda2-4.1.1-Linux-x86_64.sh          kafka_produce_patient_vitals.py
downloads                                Notebook
ec2-user@ip-172-31-30-214:~$ python3 kafka_produce_patient_vitals.py
{'customerId': 1, 'heartBeat': 74, 'bp': 202}
{'customerId': 2, 'heartBeat': 68, 'bp': 173}
{'customerId': 3, 'heartBeat': 71, 'bp': 152}

```

```

hadoop@ip-172-31-90-234:~$
-234.ec2.internal:46147
24/07/25 05:16:41 INFO BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy
24/07/25 05:16:41 INFO BlockManager: external shuffle service port = 7337
24/07/25 05:16:41 INFO BlockManagerMaster: Registering BlockManager BlockManagerId(driver, ip-172-31-90-234.ec2.internal, 46147, None)
24/07/25 05:16:41 INFO BlockManagerMasterEndpoint: Registering block manager ip-172-31-90-234.ec2.internal:46147 with 912.3 MiB RAM, BlockManagerId(driver, ip-172-31-90-234.ec2.internal, 46147, None)
24/07/25 05:16:41 INFO BlockManagerMaster: Registered BlockManager BlockManagerId(driver, ip-172-31-90-234.ec2.internal, 46147, None)
24/07/25 05:16:41 INFO BlockManager: Initialized BlockManager: BlockManagerId(driver, ip-172-31-90-234.ec2.internal, 46147, None)
24/07/25 05:16:42 INFO SingleEventLogFileWriter: Logging events to hdfs:/var/log/spark/apps/local-1721884601542.inprogress
-----
Batch: 0
-----
+-----+-----+---+-----+
|customerId|heartBeat|bp|message_time|
+-----+-----+---+-----+

```

For Consumer Application: **spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.12:3.3.1 kafka_spark_patient_vitals.py**

```

hadoop@ip-172-31-90-129:~$
-129.ec2.internal:45885
24/07/21 10:19:50 INFO BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy
24/07/21 10:19:50 INFO BlockManager: external shuffle service port = 7337
24/07/21 10:19:50 INFO BlockManagerMaster: Registering BlockManager BlockManagerId(driver, ip-172-31-90-129.ec2.internal, 45885, None)
24/07/21 10:19:50 INFO BlockManagerMasterEndpoint: Registering block manager ip-172-31-90-129.ec2.internal:45885 with 912.3 MiB RAM, BlockManagerId(driver, ip-172-31-90-129.ec2.internal, 45885, None)
24/07/21 10:19:50 INFO BlockManagerMaster: Registered BlockManager BlockManagerId(driver, ip-172-31-90-129.ec2.internal, 45885, None)
24/07/21 10:19:50 INFO BlockManager: Initialized BlockManager: BlockManagerId(driver, ip-172-31-90-129.ec2.internal, 45885, None)
24/07/21 10:19:51 INFO SingleEventLogFileWriter: Logging events to hdfs:/var/log/spark/apps/local-1721557189930.inprogress
-----
Batch: 0
-----
+-----+-----+---+-----+
|customerId|heartBeat|bp|message_time|
+-----+-----+---+-----+

```

After 30 minutes when all 1800 data being streamed and saved to Parquet file of the required HDFS location

```
ec2-user@ip-172-31-30-214:~$ cat /tmp/health-alert-patients-vital-info/
{'customerId': 2, 'heartBeat': 72, 'bp': 163}
{'customerId': 3, 'heartBeat': 69, 'bp': 164}
{'customerId': 4, 'heartBeat': 73, 'bp': 168}
{'customerId': 5, 'heartBeat': 72, 'bp': 172}
{'customerId': 1, 'heartBeat': 73, 'bp': 209}
{'customerId': 2, 'heartBeat': 71, 'bp': 174}
{'customerId': 3, 'heartBeat': 70, 'bp': 175}
{'customerId': 4, 'heartBeat': 68, 'bp': 157}
{'customerId': 5, 'heartBeat': 72, 'bp': 169}
{'customerId': 1, 'heartBeat': 74, 'bp': 220}
{'customerId': 2, 'heartBeat': 70, 'bp': 153}
{'customerId': 3, 'heartBeat': 67, 'bp': 154}
{'customerId': 4, 'heartBeat': 70, 'bp': 174}
{'customerId': 5, 'heartBeat': 67, 'bp': 177}
{'customerId': 1, 'heartBeat': 76, 'bp': 167}
{'customerId': 2, 'heartBeat': 89, 'bp': 174}
{'customerId': 3, 'heartBeat': 71, 'bp': 155}
{'customerId': 4, 'heartBeat': 73, 'bp': 175}
{'customerId': 5, 'heartBeat': 69, 'bp': 161}
{'customerId': 1, 'heartBeat': 75, 'bp': 207}
{'customerId': 2, 'heartBeat': 66, 'bp': 159}
{'customerId': 3, 'heartBeat': 71, 'bp': 153}
{'customerId': 4, 'heartBeat': 72, 'bp': 153}
{'customerId': 5, 'heartBeat': 66, 'bp': 171}

hadoop@ip-172-31-90-234:~$ cat /tmp/health-alert-patients-vital-info/
customerid|heartBeat|bp|message_time
-----+-----+-----+-----+
15      | 67      | 177 | 2024-07-25 05:47:10.002 |
11      | 76      | 167 | 2024-07-25 05:47:10.002 |
12      | 89      | 174 | 2024-07-25 05:47:10.002 |
13      | 71      | 155 | 2024-07-25 05:47:10.002 |
14      | 73      | 175 | 2024-07-25 05:47:10.002 |
15      | 69      | 161 | 2024-07-25 05:47:10.002 |
11      | 75      | 207 | 2024-07-25 05:47:10.002 |
12      | 66      | 159 | 2024-07-25 05:47:10.002 |
13      | 71      | 153 | 2024-07-25 05:47:10.002 |
14      | 72      | 153 | 2024-07-25 05:47:10.002 |
-----+-----+-----+-----+
Batch: 182
-----+-----+-----+-----+
customerid|heartBeat|bp|message_time
-----+-----+-----+-----+
15      | 66      | 171 | 2024-07-25 05:47:20.004 |
-----+-----+-----+-----+
```

STATEMENT TO CHECK DATA STORED IN HDFS LOCATION:

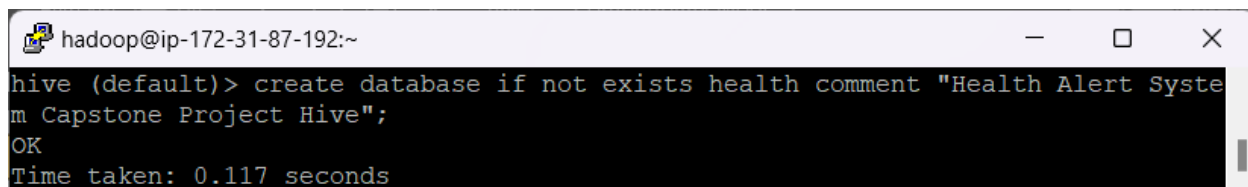
`hadoop fs -ls /user/hadoop/health-alert/patients-vital-info/`

```
hadoop@ip-172-31-90-234:~$ hadoop fs -ls /user/hadoop/health-alert/patients-vital-info/
Found 184 items
drwxr-xr-x   - hadoop hdfsadmin group      0 2024-07-25 05:47 /user/hadoop/health-alert/patients-vital-info/_spark_metadata
-rw-r--r--   1 hadoop hdfsadmin group    1382 2024-07-25 05:46 /user/hadoop/health-alert/patients-vital-info/part-00000-00761bef-8719-4441-b8d5-1c649069b313-c0000.snappy.parquet
-rw-r--r--   1 hadoop hdfsadmin group    1383 2024-07-25 05:28 /user/hadoop/health-alert/patients-vital-info/part-00000-00c9b395-8bf9-4dee-a6e7-4e3f654ef849-c0000.snappy.parquet
-rw-r--r--   1 hadoop hdfsadmin group    1383 2024-07-25 05:22 /user/hadoop/health-alert/patients-vital-info/part-00000-01acd8c4-3e3f-475f-bd25-808b98628ae4-c0000.snappy.parquet
-rw-r--r--   1 hadoop hdfsadmin group    1355 2024-07-25 05:27 /user/hadoop/health-alert/patients-vital-info/part-00000-02cb7a57-6c7a-4c09-8e38-33de79ebae8e-c0000.snappy.parquet
-rw-r--r--   1 hadoop hdfsadmin group    1372 2024-07-25 05:17 /user/hadoop/health-alert/patients-vital-info/part-00000-04053378-5b8f-413d-a74e-c854890a365e-c0000.snappy.parquet
-rw-r--r--   1 hadoop hdfsadmin group    1387 2024-07-25 05:19 /user/hadoop/health-alert/patients-vital-info/part-00000-044b3d0e-03c2-4adf-8ea3-e9fc5530d49d-c0000.snappy.parquet
-rw-r--r--   1 hadoop hdfsadmin group    1374 2024-07-25 05:26 /user/hadoop/health-alert/patients-vital-info/part-00000-0490cc09-00d6-42c2-82bf-bcc32c9e997d-c0000.snappy.parquet
```

So till now task 1 and task 2 was accomplished and patients-vital-info was stored in hdfs from rds through kafka.

STEP 3: Build an external hive table on HDFS and view data**STATEMENTS TO CREATE DATABASE FOR PATIENTS' VITAL INFORMATION**

create database if not exists health comment "Health Alert System Capstone Project ` Database";

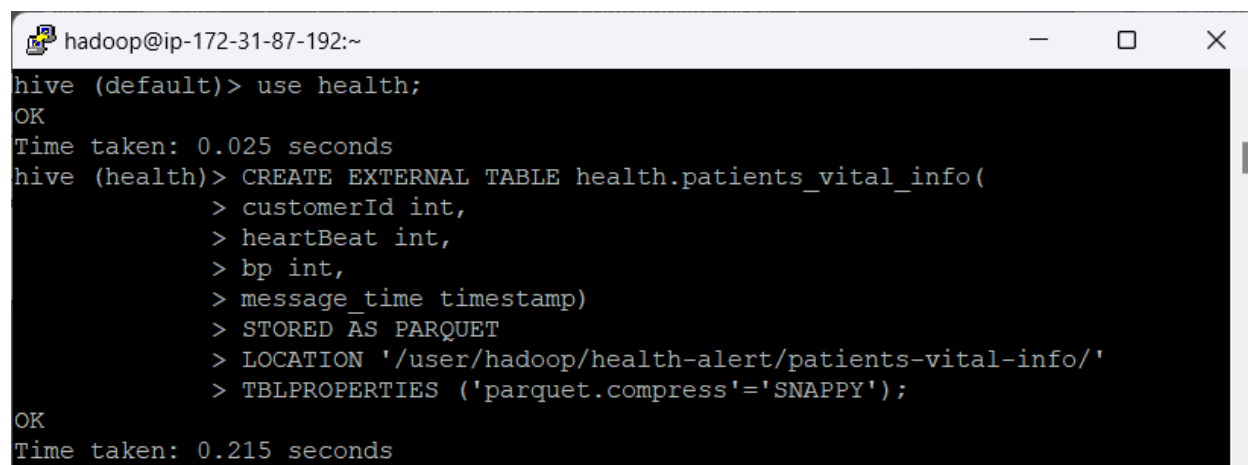


```
hadoop@ip-172-31-87-192:~  
hive (default)> create database if not exists health comment "Health Alert System Capstone Project Hive";  
OK  
Time taken: 0.117 seconds
```

STATEMENTS TO CREATE EXTERNAL HIVE TABLE FOR PATIENTS' VITAL INFORMATION

use health;

```
CREATE EXTERNAL TABLE health.patients_vital_info(  
  customerId int,  
  heartBeat int,  
  bp int,  
  message_time timestamp)  
STORED AS PARQUET  
LOCATION '/user/hadoop/health-alert/patients-vital-info/'  
TBLPROPERTIES ('parquet.compress'='SNAPPY');
```



```
hadoop@ip-172-31-87-192:~  
hive (default)> use health;  
OK  
Time taken: 0.025 seconds  
hive (health)> CREATE EXTERNAL TABLE health.patients_vital_info(  
  > customerId int,  
  > heartBeat int,  
  > bp int,  
  > message_time timestamp)  
  > STORED AS PARQUET  
  > LOCATION '/user/hadoop/health-alert/patients-vital-info/'  
  > TBLPROPERTIES ('parquet.compress'='SNAPPY');  
OK  
Time taken: 0.215 seconds
```

STATEMENTS TO VIEW PATIENTS' VITAL INFORMATION HIVE TABLE:

```
select * from health.patients_vital_info limit;
```

```
hadoop@ip-172-31-87-192:~
hive (health)> select * from health.patients_vital_info limit;
OK
limit.customerid      limit.heartbeat limit.bp      limit.message_time
1      70      189      2024-07-29 05:37:00.007
2      72      173      2024-07-29 05:37:00.007
3      68      178      2024-07-29 05:37:00.007
4      71      152      2024-07-29 05:37:00.007
5      73      166      2024-07-29 05:37:00.007
1      74      185      2024-07-29 05:37:00.007
2      67      177      2024-07-29 05:37:00.007
3      66      158      2024-07-29 05:37:00.007
4      71      177      2024-07-29 05:37:00.007
5      66      155      2024-07-29 05:37:00.007
5      73      173      2024-07-29 05:49:30.003
1      77      203      2024-07-29 05:49:30.003
2      67      165      2024-07-29 05:49:30.003
3      71      166      2024-07-29 05:49:30.003
4      66      158      2024-07-29 05:49:30.003
5      71      152      2024-07-29 05:49:30.003
1      78      177      2024-07-29 05:49:30.003
2      71      172      2024-07-29 05:49:30.003
3      70      151      2024-07-29 05:49:30.003
4      68      156      2024-07-29 05:49:30.003
1      77      184      2024-07-29 05:43:10.003
2      66      162      2024-07-29 05:43:10.003
3      68      159      2024-07-29 05:43:10.003
2      72      166      2024-07-29 05:38:40.004
3      66      158      2024-07-29 05:38:40.004
4      71      176      2024-07-29 05:38:40.004
5      67      168      2024-07-29 05:38:40.004
5      66      174      2024-07-29 05:54:10.003
1      75      194      2024-07-29 05:54:10.003
2      67      152      2024-07-29 05:54:10.003
3      72      165      2024-07-29 05:54:10.003
4      72      163      2024-07-29 05:54:10.003
5      68      172      2024-07-29 05:54:10.003
1      74      189      2024-07-29 05:54:10.003
2      68      161      2024-07-29 05:54:10.003
3      67      170      2024-07-29 05:54:10.003
4      68      161      2024-07-29 05:54:10.003
4      70      169      2024-07-29 05:59:10.003
5      73      175      2024-07-29 05:59:10.003
1      77      207      2024-07-29 05:59:10.003
2      69      158      2024-07-29 05:59:10.003
3      68      178      2024-07-29 05:59:10.003
4      68      172      2024-07-29 05:59:10.003
5      69      154      2024-07-29 05:59:10.003
1      71      202      2024-07-29 05:59:10.003
2      71      179      2024-07-29 05:59:10.003
3      70      156      2024-07-29 05:59:10.003
Time taken: 1.712 seconds, Fetched: 1800 row(s)
hive (health)>
```

STEP 4: Created hbase table with 3 families (attribute, limit, alert) and inserted 12 records into hbase table:

STATEMENTS TO CREATE HBASE THRESHOLD TABLE

```
create 'threshold_ref_hbase', 'attribute', 'limit', 'alert'
```

STATEMENTS TO VIEW HBASE THRESHOLD TABLE ABOVE MENTIONED 3 FAMILIES:

```
describe 'threshold_ref_hbase'
```

```
hadoop@ip-172-31-87-192:~
[hadoop@ip-172-31-87-192 ~]$ sudo hbase shell
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/hbase/lib/client-facing-thirdparty/slf4j-reload4j-1.7.33.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Reload4jLoggerFactory]
HBase Shell
Use "help" to get list of supported commands.
Use "exit" to quit this interactive shell.
For Reference, please visit: http://hbase.apache.org/2.0/book.html#shell
Version 2.4.15-amzn-0.1, rUnknown, Fri Jun 23 16:31:13 UTC 2023
Took 0.0032 seconds
hbase:001:0> create 'threshold_ref_hbase', 'attribute', 'limit', 'alert'
Created table threshold_ref_hbase
Took 1.2594 seconds
=> Hbase::Table - threshold_ref_hbase
hbase:002:0> describe 'threshold_ref_hbase'
Table threshold_ref_hbase is ENABLED
threshold_ref_hbase
COLUMN FAMILIES DESCRIPTION
{NAME => 'alert', BLOOMFILTER => 'ROW', IN MEMORY => 'false', VERSIONS => '1', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', COMPRESSION => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}

{NAME => 'attribute', BLOOMFILTER => 'ROW', IN MEMORY => 'false', VERSIONS => '1', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', COMPRESSION => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}

{NAME => 'limit', BLOOMFILTER => 'ROW', IN MEMORY => 'false', VERSIONS => '1', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', COMPRESSION => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}

3 row(s)
Quota is disabled
Took 0.1750 seconds
hbase:003:0> █
```

STATEMENTS TO INSERT THRESHOLD PARAMETERS INTO HBASE THRESHOLD TABLE:

```
put 'threshold_ref_hbase', '1', 'attribute:attribute', 'heartBeat'
put 'threshold_ref_hbase', '1', 'limit:low_age_limit', '0'
put 'threshold_ref_hbase', '1', 'limit:high_age_limit', '40'
put 'threshold_ref_hbase', '1', 'limit:low_range_value', '0'
put 'threshold_ref_hbase', '1', 'limit:high_range_value', '69'
put 'threshold_ref_hbase', '1', 'alert:alert_flag', '1'
put 'threshold_ref_hbase', '1', 'alert:alert_message', 'Low Heart Rate than Normal'
```

```
put 'threshold_ref_hbase', '2', 'attribute:attribute', 'heartBeat'
put 'threshold_ref_hbase', '2', 'limit:low_age_limit', '0'
put 'threshold_ref_hbase', '2', 'limit:high_age_limit', '40'
put 'threshold_ref_hbase', '2', 'limit:low_range_value', '70'
put 'threshold_ref_hbase', '2', 'limit:high_range_value', '78'
put 'threshold_ref_hbase', '2', 'alert:alert_flag', '0'
put 'threshold_ref_hbase', '2', 'alert:alert_message', 'Normal'
```

```
put 'threshold_ref_hbase', '3', 'attribute:attribute', 'heartBeat'
put 'threshold_ref_hbase', '3', 'limit:low_age_limit', '0'
put 'threshold_ref_hbase', '3', 'limit:high_age_limit', '40'
put 'threshold_ref_hbase', '3', 'limit:low_range_value', '79'
put 'threshold_ref_hbase', '3', 'limit:high_range_value', '9999'
put 'threshold_ref_hbase', '3', 'alert:alert_flag', '1'
put 'threshold_ref_hbase', '3', 'alert:alert_message', 'Higher Heart Rate than Normal'
```

```
put 'threshold_ref_hbase', '4', 'attribute:attribute', 'bp'
put 'threshold_ref_hbase', '4', 'limit:low_age_limit', '0'
put 'threshold_ref_hbase', '4', 'limit:high_age_limit', '40'
put 'threshold_ref_hbase', '4', 'limit:low_range_value', '0'
put 'threshold_ref_hbase', '4', 'limit:high_range_value', '160'
put 'threshold_ref_hbase', '4', 'alert:alert_flag', '1'
put 'threshold_ref_hbase', '4', 'alert:alert_message', 'Low BP than Normal'
```

```
put 'threshold_ref_hbase', '5', 'attribute:attribute', 'bp'
put 'threshold_ref_hbase', '5', 'limit:low_age_limit', '0'
put 'threshold_ref_hbase', '5', 'limit:high_age_limit', '40'
put 'threshold_ref_hbase', '5', 'limit:low_range_value', '161'
put 'threshold_ref_hbase', '5', 'limit:high_range_value', '220'
put 'threshold_ref_hbase', '5', 'alert:alert_flag', '0'
put 'threshold_ref_hbase', '5', 'alert:alert_message', 'Normal'
```

```
put 'threshold_ref_hbase', '6', 'attribute:attribute', 'bp'  
put 'threshold_ref_hbase', '6', 'limit:low_age_limit', '0'  
put 'threshold_ref_hbase', '6', 'limit:high_age_limit', '40'  
put 'threshold_ref_hbase', '6', 'limit:low_range_value', '221'  
put 'threshold_ref_hbase', '6', 'limit:high_range_value', '9999'  
put 'threshold_ref_hbase', '6', 'alert:alert_flag', '1'  
put 'threshold_ref_hbase', '6', 'alert:alert_message', 'Higher BP than Normal'
```

```
put 'threshold_ref_hbase', '7', 'attribute:attribute', 'heartBeat'  
put 'threshold_ref_hbase', '7', 'limit:low_age_limit', '41'  
put 'threshold_ref_hbase', '7', 'limit:high_age_limit', '100'  
put 'threshold_ref_hbase', '7', 'limit:low_range_value', '0'  
put 'threshold_ref_hbase', '7', 'limit:high_range_value', '65'  
put 'threshold_ref_hbase', '7', 'alert:alert_flag', '1'  
put 'threshold_ref_hbase', '7', 'alert:alert_message', 'Low Heart Rate than Normal'
```

```
put 'threshold_ref_hbase', '8', 'attribute:attribute', 'heartBeat'  
put 'threshold_ref_hbase', '8', 'limit:low_age_limit', '41'  
put 'threshold_ref_hbase', '8', 'limit:high_age_limit', '100'  
put 'threshold_ref_hbase', '8', 'limit:low_range_value', '66'  
put 'threshold_ref_hbase', '8', 'limit:high_range_value', '73'  
put 'threshold_ref_hbase', '8', 'alert:alert_flag', '0'  
put 'threshold_ref_hbase', '8', 'alert:alert_message', 'Normal'
```

```
put 'threshold_ref_hbase', '9', 'attribute:attribute', 'heartBeat'  
put 'threshold_ref_hbase', '9', 'limit:low_age_limit', '41'  
put 'threshold_ref_hbase', '9', 'limit:high_age_limit', '100'  
put 'threshold_ref_hbase', '9', 'limit:low_range_value', '74'  
put 'threshold_ref_hbase', '9', 'limit:high_range_value', '9999'  
put 'threshold_ref_hbase', '9', 'alert:alert_flag', '1'  
put 'threshold_ref_hbase', '9', 'alert:alert_message', 'Higher Heart Rate than Normal'
```

```
put 'threshold_ref_hbase', '10', 'attribute:attribute', 'bp'  
put 'threshold_ref_hbase', '10', 'limit:low_age_limit', '41'  
put 'threshold_ref_hbase', '10', 'limit:high_age_limit', '100'  
put 'threshold_ref_hbase', '10', 'limit:low_range_value', '0'  
put 'threshold_ref_hbase', '10', 'limit:high_range_value', '150'  
put 'threshold_ref_hbase', '10', 'alert:alert_flag', '1'  
put 'threshold_ref_hbase', '10', 'alert:alert_message', 'Low BP than Normal'
```

```
put 'threshold_ref_hbase', '11', 'attribute:attribute', 'bp'  
put 'threshold_ref_hbase', '11', 'limit:low_age_limit', '41'  
put 'threshold_ref_hbase', '11', 'limit:high_age_limit', '100'
```



```
put 'threshold_ref_hbase', '11', 'limit:low_range_value', '151'
put 'threshold_ref_hbase', '11', 'limit:high_range_value', '180'
put 'threshold_ref_hbase', '11', 'alert:alert_flag', '0'
put 'threshold_ref_hbase', '11', 'alert:alert_message', 'Normal'
```

```
put 'threshold_ref_hbase', '12', 'attribute:attribute', 'bp'
put 'threshold_ref_hbase', '12', 'limit:low_age_limit', '41'
put 'threshold_ref_hbase', '12', 'limit:high_age_limit', '100'
put 'threshold_ref_hbase', '12', 'limit:low_range_value', '181'hive
put 'threshold_ref_hbase', '12', 'limit:high_range_value', '9999'
put 'threshold_ref_hbase', '12', 'alert:alert_flag', '1'
put 'threshold_ref_hbase', '12', 'alert:alert_message', 'Higher BP than Normal'
```

```
hadoop@ip-172-31-87-192:~
hbase:001:0> put 'threshold_ref_hbase', '1', 'attribute:attribute', 'heartBeat'
Took 0.9445 seconds
hbase:002:0> put 'threshold_ref_hbase', '1', 'limit:low_age_limit', '0'
Took 0.0076 seconds
hbase:003:0> put 'threshold_ref_hbase', '1', 'limit:high_age_limit', '40'
Took 0.0041 seconds
hbase:004:0> put 'threshold_ref_hbase', '1', 'limit:low_range_value', '0'
Took 0.0083 seconds
hbase:005:0> put 'threshold_ref_hbase', '1', 'limit:high_range_value', '69'
Took 0.0069 seconds
hbase:006:0> put 'threshold_ref_hbase', '1', 'alert:alert_flag', '1'
Took 0.0082 seconds
hbase:007:0> put 'threshold_ref_hbase', '1', 'alert:alert_message', 'Low Heart Rate than Normal'
Took 0.0100 seconds
hbase:008:0> put 'threshold_ref_hbase', '2', 'attribute:attribute', 'heartBeat'
Took 0.0096 seconds
hbase:009:0> put 'threshold_ref_hbase', '2', 'limit:low_age_limit', '0'
Took 0.0075 seconds
hbase:010:0> put 'threshold_ref_hbase', '2', 'limit:high_age_limit', '40'
Took 0.0041 seconds
hbase:011:0> put 'threshold_ref_hbase', '2', 'limit:low_range_value', '70'
Took 0.0054 seconds
hbase:012:0> put 'threshold_ref_hbase', '2', 'limit:high_range_value', '78'
Took 0.0080 seconds
hbase:013:0> put 'threshold_ref_hbase', '2', 'alert:alert_flag', '0'
Took 0.0072 seconds
hbase:014:0> put 'threshold_ref_hbase', '2', 'alert:alert_message', 'Normal'
Took 0.0050 seconds
hbase:015:0> put 'threshold_ref_hbase', '3', 'attribute:attribute', 'heartBeat'
Took 0.0044 seconds
hbase:016:0> put 'threshold_ref_hbase', '3', 'limit:low_age_limit', '0'
Took 0.0053 seconds
hbase:017:0> put 'threshold_ref_hbase', '3', 'limit:high_age_limit', '40'
Took 0.0099 seconds
hbase:018:0> put 'threshold_ref_hbase', '3', 'limit:low_range_value', '79'
Took 0.0096 seconds
hbase:019:0> put 'threshold_ref_hbase', '3', 'limit:high_range_value', '9999'
Took 0.0092 seconds
hbase:020:0> put 'threshold_ref_hbase', '3', 'alert:alert_flag', '1'
Took 0.0162 seconds
hbase:021:0> put 'threshold_ref_hbase', '3', 'alert:alert_message', 'Higher Heart Rate than Normal'
Took 0.0057 seconds
```

```
hadoop@ip-172-31-87-192:~
hbase:068:0> put 'threshold_ref_hbase', '10', 'limit:high_range_value', '150'
Took 0.0068 seconds
hbase:069:0> put 'threshold_ref_hbase', '10', 'alert:alert_flag', '1'
Took 0.0047 seconds
hbase:070:0> put 'threshold_ref_hbase', '10', 'alert:alert_message', 'Low BP than Normal'
Took 0.0041 seconds
hbase:071:0> put 'threshold_ref_hbase', '11', 'attribute:attribute', 'bp'
Took 0.0055 seconds
hbase:072:0> put 'threshold_ref_hbase', '11', 'limit:low_age_limit', '41'
Took 0.0049 seconds
hbase:073:0> put 'threshold_ref_hbase', '11', 'limit:high_age_limit', '100'
Took 0.0060 seconds
hbase:074:0> put 'threshold_ref_hbase', '11', 'limit:low_range_value', '151'
Took 0.0043 seconds
hbase:075:0> put 'threshold_ref_hbase', '11', 'limit:high_range_value', '180'
Took 0.0084 seconds
hbase:076:0> put 'threshold_ref_hbase', '11', 'alert:alert_flag', '0'
Took 0.0047 seconds
hbase:077:0> put 'threshold_ref_hbase', '11', 'alert:alert_message', 'Normal'
Took 0.0055 seconds
hbase:078:0> put 'threshold_ref_hbase', '12', 'attribute:attribute', 'bp'
Took 0.0054 seconds
hbase:079:0> put 'threshold_ref_hbase', '12', 'limit:low_age_limit', '41'
Took 0.0054 seconds
hbase:080:0> put 'threshold_ref_hbase', '12', 'limit:high_age_limit', '100'
Took 0.0041 seconds
hbase:081:0> put 'threshold_ref_hbase', '12', 'limit:low_range_value', '181'
Took 0.0048 seconds
hbase:082:0> put 'threshold_ref_hbase', '12', 'limit:high_range_value', '9999'
Took 0.0041 seconds
hbase:083:0> put 'threshold_ref_hbase', '12', 'alert:alert_flag', '1'
Took 0.0052 seconds
hbase:084:0> put 'threshold_ref_hbase', '12', 'alert:alert_message', 'Higher BP than Normal'
hbase:085:0> quit
hbase:086:0> ^C
hbase:086:0> put 'threshold_ref_hbase', '12', 'alert:alert_message', 'Higher BP than Normal'
hbase:087:0> ^C
hbase:087:0> put 'threshold_ref_hbase', '12', 'alert:alert_message', 'Higher BP than Normal'
Took 0.0177 seconds
hbase:088:0> █
```

STATEMENTS TO VIEW HBASE THRESHOLD TABLE ABOVE MENTIONED AFTER DATA INSERTION:

```
scan 'threshold_ref_hbase'
```

```
hadoop@ip-172-31-87-192:~
hbase:088:0> scan 'threshold_ref_hbase'
ROW      COLUMN+CELL
1        column=alert:alert_flag, timestamp=2024-07-29T05:45:27.008
, value=1
1        column=alert:alert_message, timestamp=2024-07-29T05:45:32.
111, value=Low Heart Rate than Normal
1        column=attribute:attribute, timestamp=2024-07-29T05:45:01.
331, value=heartBeat
1        column=limit:high_age_limit, timestamp=2024-07-29T05:45:10
.898, value=40
1        column=limit:high_range_value, timestamp=2024-07-29T05:45:
21.236, value=69
1        column=limit:low_age_limit, timestamp=2024-07-29T05:45:06.
484, value=0
1        column=limit:low_range_value, timestamp=2024-07-29T05:45:1
4.592, value=0
10       column=alert:alert_flag, timestamp=2024-07-29T05:52:19.102
, value=1
10       column=alert:alert_message, timestamp=2024-07-29T05:52:22.
798, value=Low BP than Normal
10       column=attribute:attribute, timestamp=2024-07-29T05:51:42.
614, value=bp
10       column=limit:high_age_limit, timestamp=2024-07-29T05:51:53
.020, value=100
10       column=limit:high_range_value, timestamp=2024-07-29T05:52:
07.010, value=150
10       column=limit:low_age_limit, timestamp=2024-07-29T05:51:48.
268, value=41
10       column=limit:low_range_value, timestamp=2024-07-29T05:51:5
8.277, value=0
11       column=alert:alert_flag, timestamp=2024-07-29T05:52:51.239
, value=0
11       column=alert:alert_message, timestamp=2024-07-29T05:52:57.
202, value=Normal
11       column=attribute:attribute, timestamp=2024-07-29T05:52:28.
900, value=bp
11       column=limit:high_age_limit, timestamp=2024-07-29T05:52:37
.886, value=100
11       column=limit:high_range_value, timestamp=2024-07-29T05:52:
47.286, value=180
11       column=limit:low_age_limit, timestamp=2024-07-29T05:52:33.
750, value=41
11       column=limit:low_range_value, timestamp=2024-07-29T05:52:4
2.849, value=151
```

STEP 5: STATEMENTS TO CREATE EXTERNAL HIVE TABLE FOR THRESHOLD TABLE PAREMETERS:

```
CREATE EXTERNAL TABLE IF NOT EXISTS threshold_ref_hive(  
    ref_id INT,  
    attribute VARCHAR(20),  
    low_age_limit INT,  
    high_age_limit INT,  
    low_range_value INT,  
    high_range_value INT,  
    alert_flag INT,  
    alert_message VARCHAR(255))  
STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'  
WITH SERDEPROPERTIES (  
    "hbase.columns.mapping" = ":key,  
    attribute:attribute,  
    limit:low_age_limit,  
    limit:high_age_limit,  
    limit:low_range_value,  
    limit:high_range_value,  
    alert:alert_flag,  
    alert:alert_message")  
TBLPROPERTIES ("hbase.table.name" = "threshold_ref_hbase");
```

```
hadoop@ip-172-31-81-61:~  
[hadoop@ip-172-31-81-61 ~]$ hive  
Hive Session ID = 5c12e684-32c8-4638-9c84-72f3f1724dfb  
  
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.properties Async: true  
hive> show tables;  
OK  
Time taken: 0.964 seconds  
hive> CREATE EXTERNAL TABLE IF NOT EXISTS threshold_ref_hive(  
    >     ref_id INT,  
    >     attribute VARCHAR(20),  
    >     low_age_limit INT,  
    >     high_age_limit INT,  
    >     low_range_value INT,  
    >     high_range_value INT,  
    >     alert_flag INT,  
    >     alert_message VARCHAR(255))  
    > STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'  
    > WITH SERDEPROPERTIES (  
    >     "hbase.columns.mapping" = ":key,  
    >     attribute:attribute,  
    >     limit:low_age_limit,  
    >     limit:high_age_limit,  
    >     limit:low_range_value,  
    >     limit:high_range_value,  
    >     alert:alert_flag,  
    >     alert:alert_message")  
    > TBLPROPERTIES ("hbase.table.name" = "threshold_ref_hbase");  
OK  
Time taken: 1.615 seconds
```

STATEMENT FOR VIEWING HEADER OF EACH COLUMNS:

```
set hive.cli.print.header=true;
```

STATEMENTS TO VIEW THRESHOLD DATA

```
SELECT * FROM threshold_ref_hive
ORDER BY ref_id ASC;
```

```
hadoop@ip-172-31-81-61:~$
> ref_id INT,
> attribute VARCHAR(20),
> low_age_limit INT,
> high_age_limit INT,
> low_range_value INT,
> high_range_value INT,
> alert_flag INT,
> alert_message VARCHAR(255))
> STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
> WITH SERDEPROPERTIES (
>   "hbase.columns.mapping" = ":key,
>   attribute:attribute,
>   limit:low_age_limit,
>   limit:high_age_limit,
>   limit:low_range_value,
>   limit:high_range_value,
>   alert:alert_flag,
>   alert:alert_message")
> TBLPROPERTIES ("hbase.table.name" = "threshold_ref_hbase");
OK
Time taken: 1.615 seconds
hive> set hive.cli.print.header = true
> ;
hive> SELECT * FROM threshold_ref_hive
> ORDER BY ref_id ASC;
Query ID = hadoop_20240726074954_04a243e3-af33-47d1-9cc9-850690f98caf
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1721974991030_0006)
```

VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	1	1	0	0	0	0
Reducer 2	container	SUCCEEDED	1	1	0	0	0	0

```
VERTICES: 02/02 [=====] 100% ELAPSED TIME: 0.53 s
OK
threshold_ref_hive.ref_id      threshold_ref_hive.attribute  threshold_ref_hive.low_age_limit  threshold_ref_hive.high_age_limit  threshold_ref_hive.low_range_value  threshold_ref_hive.high_range_value  threshold_ref_hive.alert_flag  threshold_ref_hive.alert_message
1      heartBeat      0      40      0      69      1      Low Heart Rate than Normal
2      heartBeat      0      40      70      78      0      Normal
3      heartBeat      0      40      79      9999      1      Higher Heart Rate than Normal
4      bp      0      40      0      160      1      Low BP than Normal
5      bp      0      40      161      220      0      Normal
6      bp      0      40      221      9999      1      Higher BP than Normal
7      heartBeat      41      100      0      65      1      Low Heart Rate than Normal
8      heartBeat      41      100      66      73      0      Normal
9      heartBeat      41      100      74      9999      1      Higher Heart Rate than Normal
10     bp      41      100      0      150      1      Low BP than Normal
11     bp      41      100      151      180      0      Normal
12     bp      41      100      181      9999      1      Higher BP than Normal
Time taken: 13.136 seconds, Fetched: 12 row(s)
hive>
```

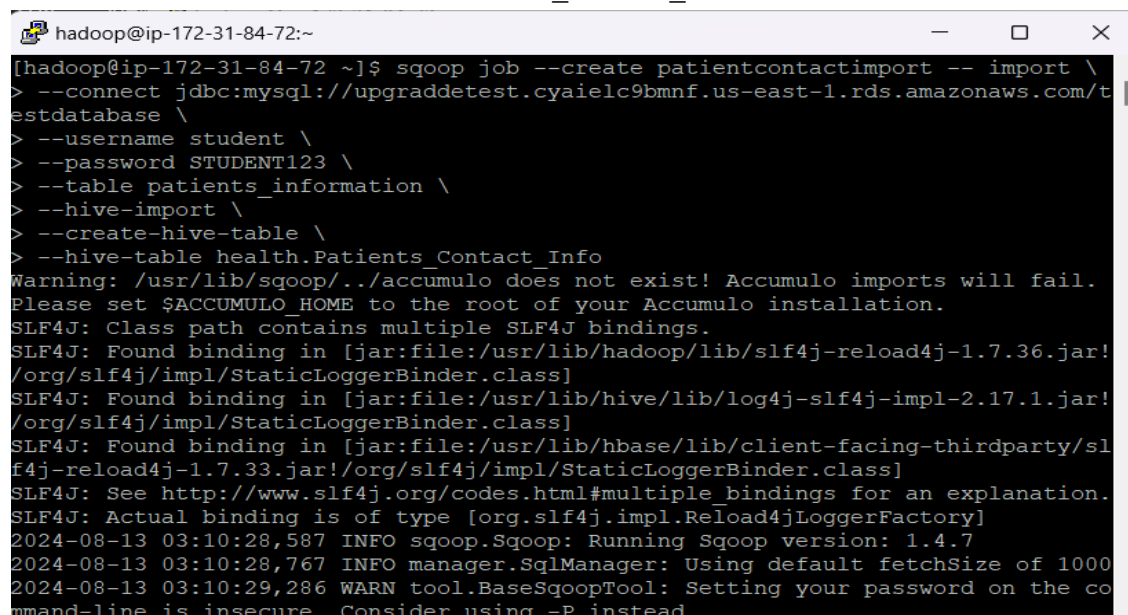
So till now task 2 was accomplished and we have patients_vital_info stored in hdfs and threshold_ref_hive table storing the threshold parameters.

STEP 6: Extract patient info using sqoop into hive table.

SQOOP COMMAND

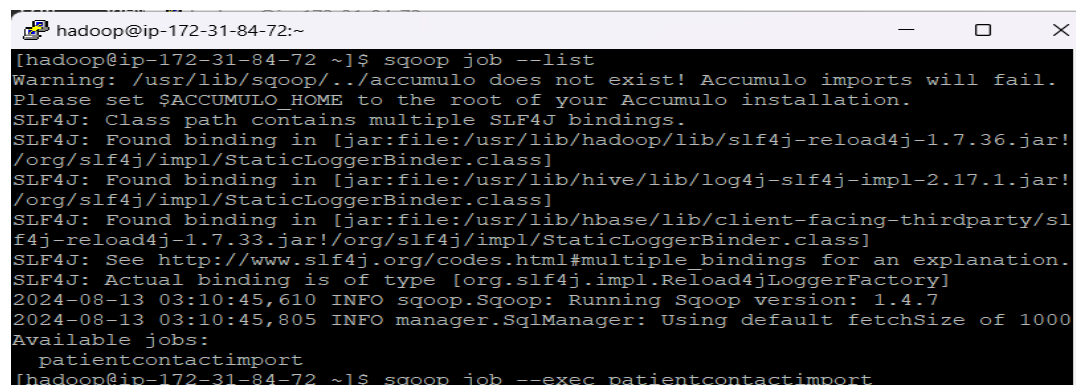
SQOOP COMMAND FOR 'PATIENTCONTACTIMPORT' JOB CREATION:

```
sqoop job --create patientcontactimport -- import \
--connect jdbc:mysql://upgradtest.cyaieic9bmnf.us-east-
1.rds.amazonaws.com/testdatabase \
--username student \
--password STUDENT123 \
--table patients_information \
--hive-import \
--create-hive-table \
--hive-table health.Patients_Contact_Info
```



```
hadoop@ip-172-31-84-72:~
[hadoop@ip-172-31-84-72 ~]$ sqoop job --create patientcontactimport -- import \
> --connect jdbc:mysql://upgradtest.cyaieic9bmnf.us-east-1.rds.amazonaws.com/t
estdatabase \
> --username student \
> --password STUDENT123 \
> --table patients_information \
> --hive-import \
> --create-hive-table \
> --hive-table health.Patients_Contact_Info
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-reload4j-1.7.36.jar!
/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/hive/lib/log4j-slf4j-impl-2.17.1.jar!
/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/hbase/lib/client-facing-thirdparty/sl
f4j-reload4j-1.7.33.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Reload4jLoggerFactory]
2024-08-13 03:10:28,587 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
2024-08-13 03:10:28,767 INFO manager.SqlManager: Using default fetchSize of 1000
2024-08-13 03:10:29,286 WARN tool.BaseSqoopTool: Setting your password on the co
mmand-line is insecure. Consider using -P instead.
```

sqoop job --list



```
hadoop@ip-172-31-84-72:~
[hadoop@ip-172-31-84-72 ~]$ sqoop job --list
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-reload4j-1.7.36.jar!
/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/hive/lib/log4j-slf4j-impl-2.17.1.jar!
/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/hbase/lib/client-facing-thirdparty/sl
f4j-reload4j-1.7.33.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Reload4jLoggerFactory]
2024-08-13 03:10:45,610 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
2024-08-13 03:10:45,805 INFO manager.SqlManager: Using default fetchSize of 1000
Available jobs:
  patientcontactimport
[hadoop@ip-172-31-84-72 ~]$ sqoop job --exec patientcontactimport
```

SQOOP COMMAND FOR 'PATIENTCONTACTIMPORT' JOB EXECUTION:

sqoop job --exec patientcontactimport

```
hadoop@ip-172-31-84-72:~
[hadoop@ip-172-31-84-72 ~]$ sqoop job --exec patientcontactimport
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/hive/lib/log4j-slf4j-impl-2.17.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/hbase/lib/client-facing-thirdparty/slf4j-reload4j-1.7.33.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Reload4jLoggerFactory]
2024-08-13 03:11:09,917 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
2024-08-13 03:11:10,114 INFO manager.SqlManager: Using default fetchSize of 1000
Enter password:
2024-08-13 03:11:18,464 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
2024-08-13 03:11:18,464 INFO tool.CodeGenTool: Beginning code generation
Loading class 'com.mysql.jdbc.Driver'. This is deprecated. The new driver class is 'com.mysql.cj.jdbc.Driver'. The driver is automatically registered via the SPI and manual loading of the driver class is generally unnecessary.
2024-08-13 03:11:18,620 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `patients_information` AS t LIMIT 1
2024-08-13 03:11:18,665 INFO manager.SqlManager: Executing SQL statement: SELECT
```

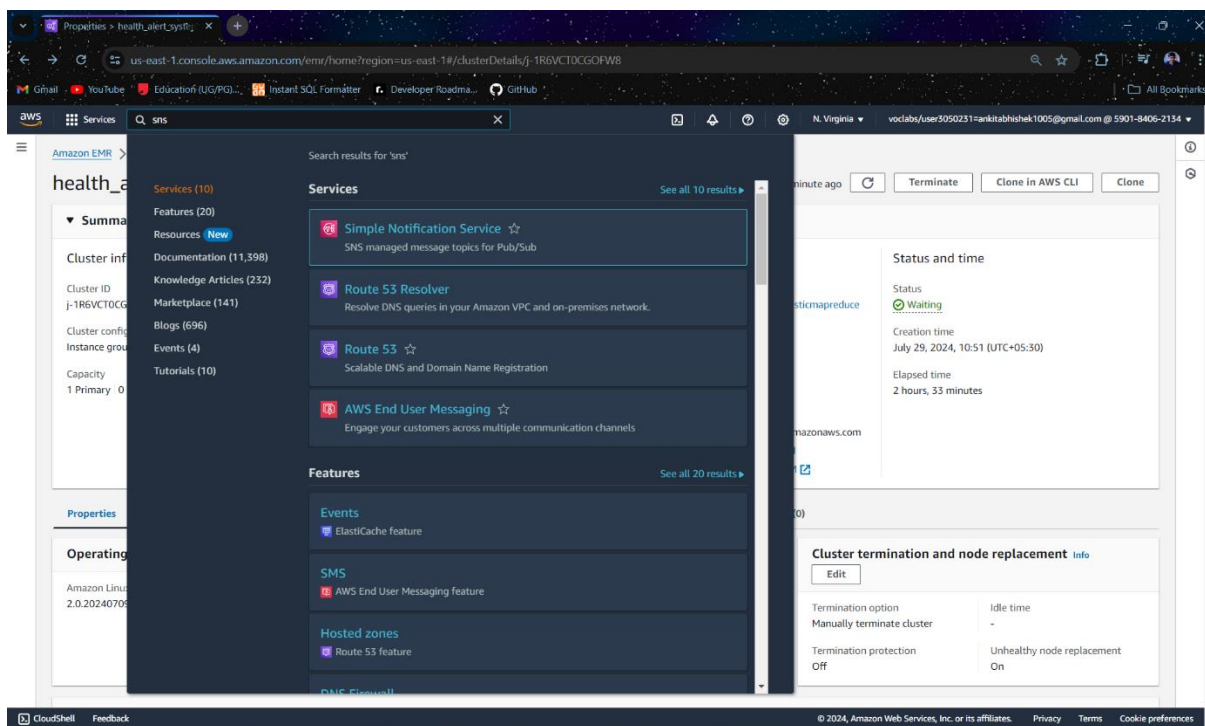
SNAPSHOT OF DATA COLLECTED:

select * from health.patients_contact_info;

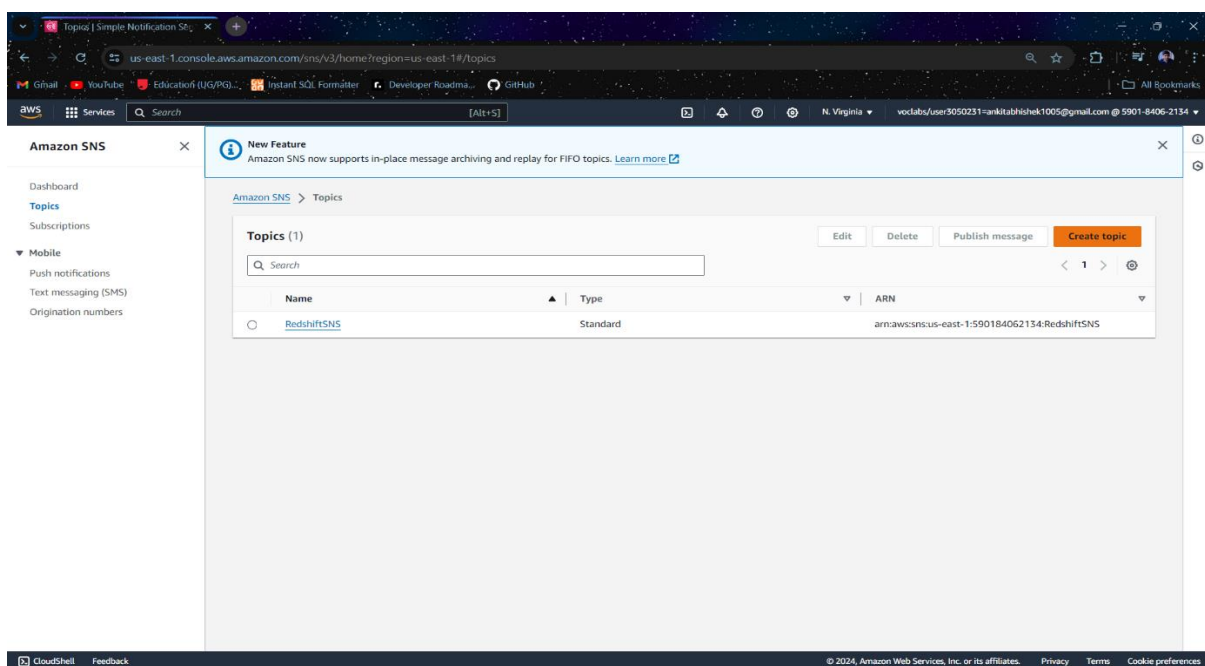
```
hadoop@ip-172-31-87-192:~
hive> select * from health.patients_contact_info;
OK
patients_contact_info.patientid patients_contact_info.patientname patients
_contact_info.patientaddress patients_contact_info.phone_number patients
_contact_info.admitted_ward patients_contact_info.age patients_contact
_info.other_details
1 Alex S XDC test Address 8982739282 1 23 NULL
2 Sammy A New Building Address 2382739282 2 45 NULL
3 Karan C Aws Address 8923739282 3 56 NULL
4 Dara M India Address 2182739282 4 67 NULL
5 Pam ABC test Address 4982739282 5 72 NULL
Time taken: 2.16 seconds, Fetched: 5 row(s)
```

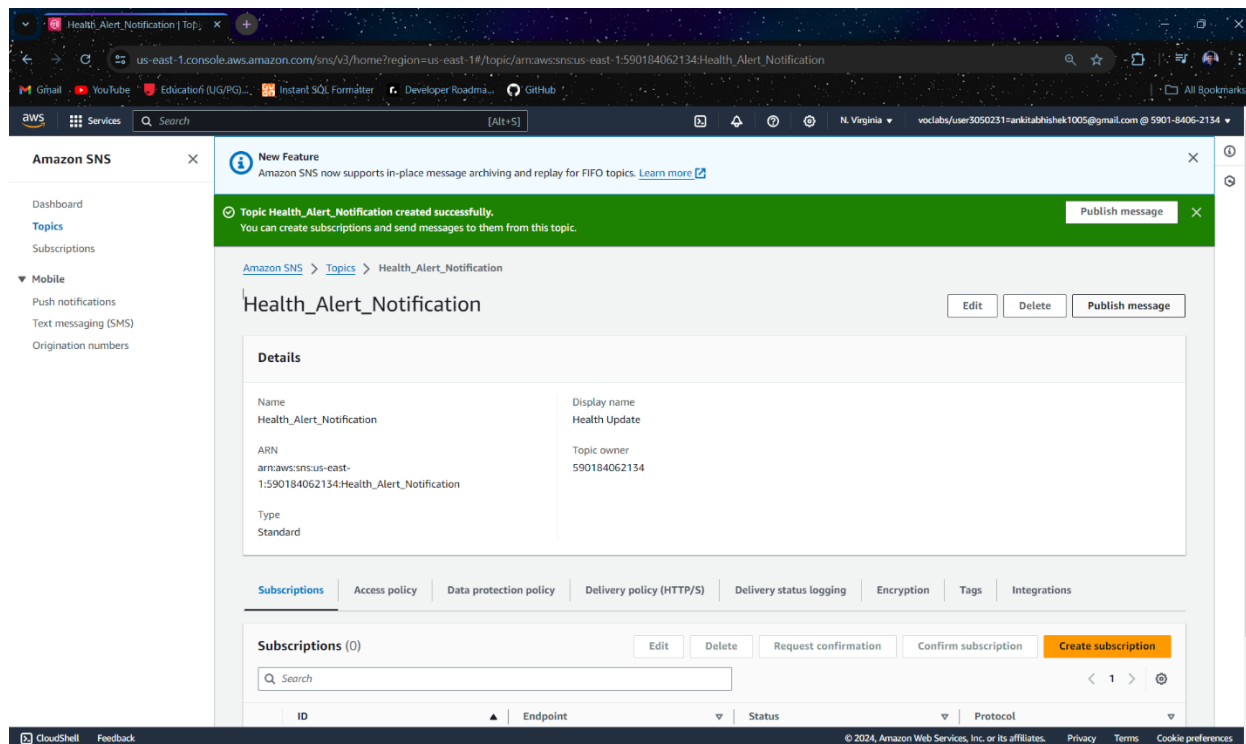
STEP 7: SNS configuration to the assessment email ID

STEP 1: SEARCH FOR AMAZON SNS (SIMPLE NOTIFICATION SERVICE):



STEP 2: CREATE TOPIC 'HEALTH_ALERT_NOTIFICATION':

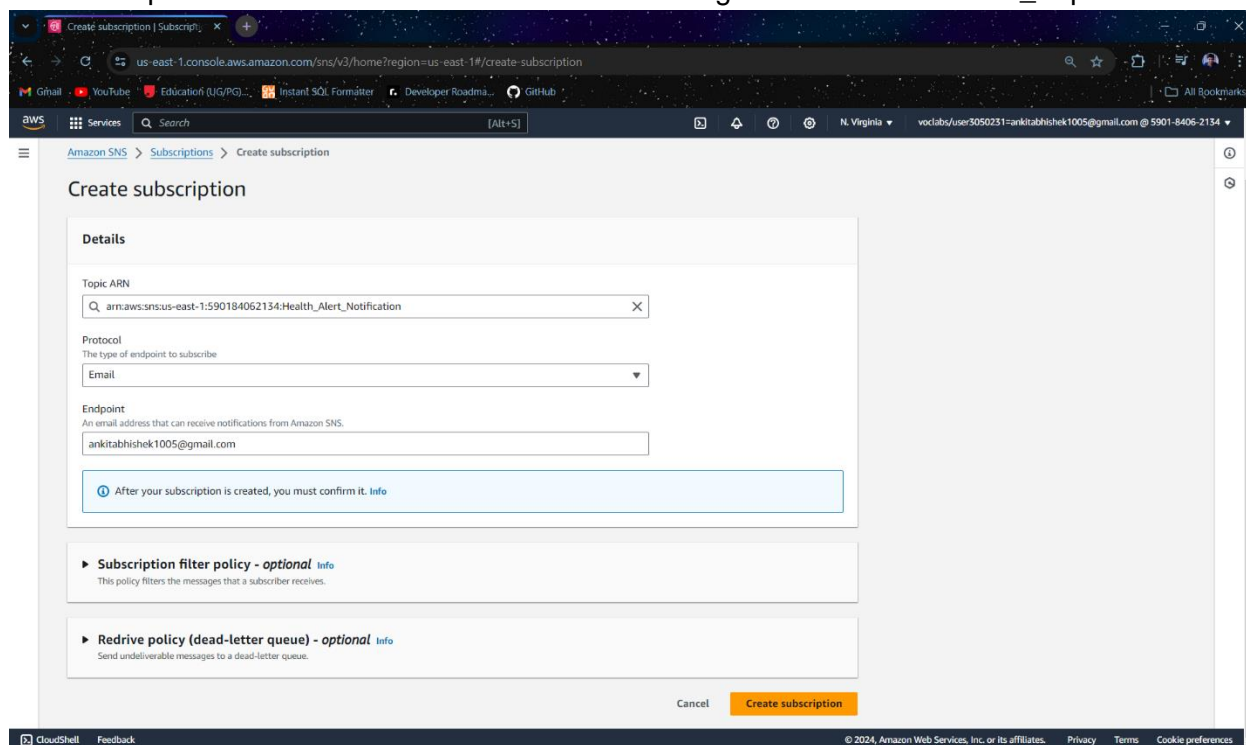


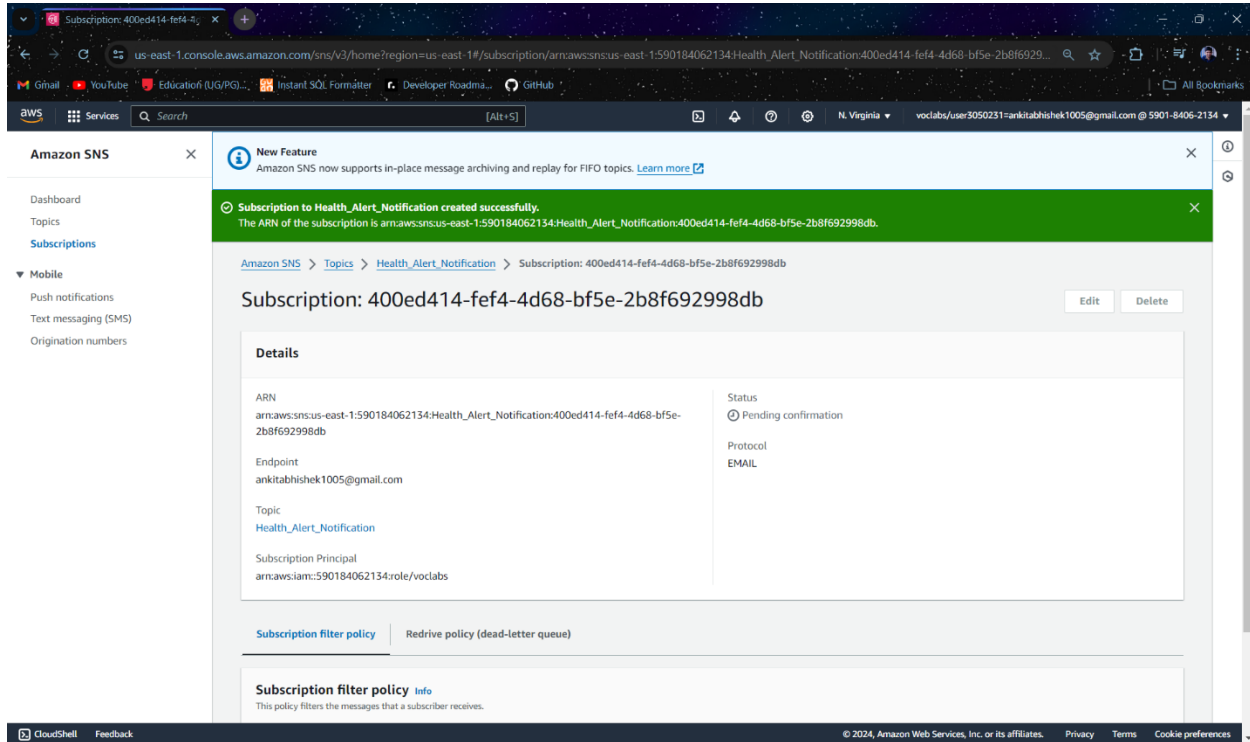


STEP 3: NOW CREATE SUBSCRIPTION FOR COLLECTING HEALTH_ALERTS:

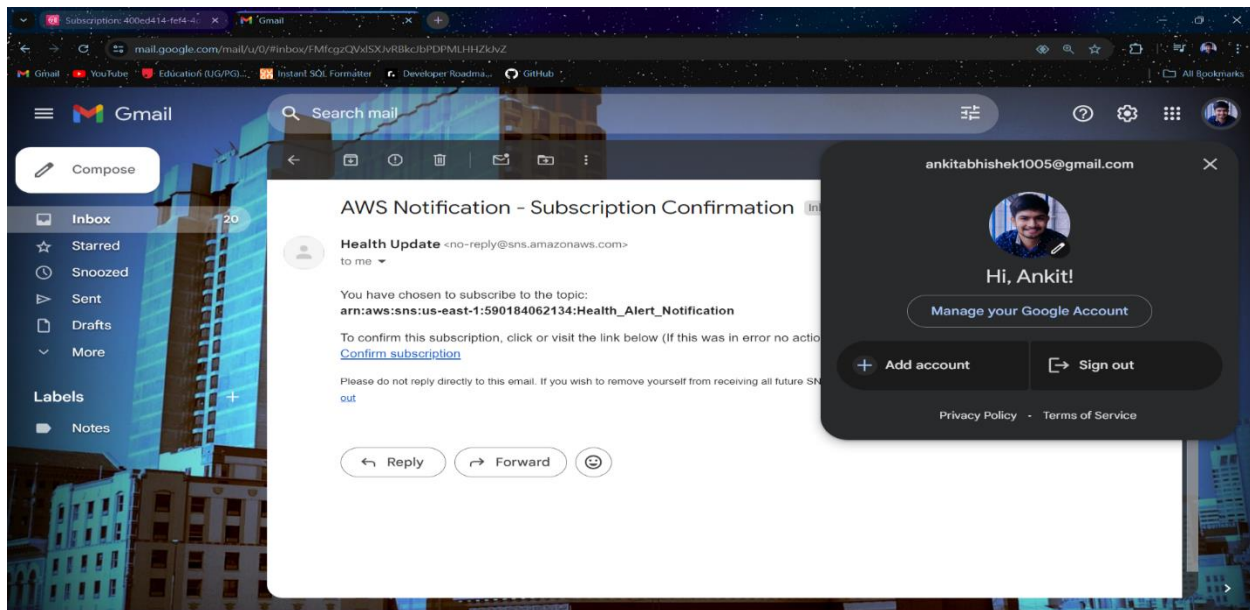
Select Protocol: Email

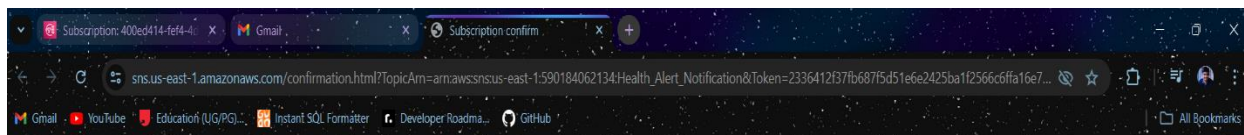
Inside Endpoint: Provided email id where we want to get alert from the Kafka_Topic





STEP 4: CONFIRM FROM THE EMAIL ID WHERE SUBSCRIPTION CONFIRMATION EMAIL ARRIVED:





Simple Notification Service

Subscription confirmed!

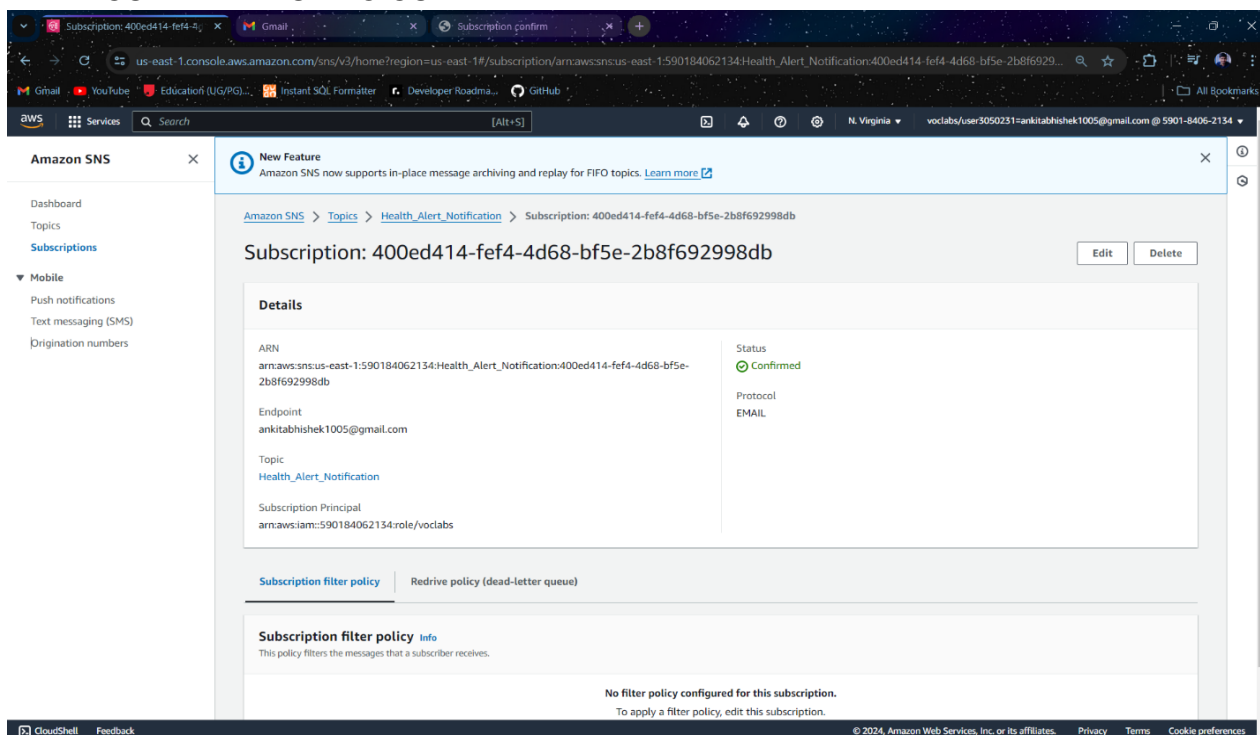
You have successfully subscribed.

Your subscription's id is:

arn:aws:sns:us-east-1:590184062134:Health_Alert_Notification:400ed414-fef4-4d68-bf5e-2b8f692998db

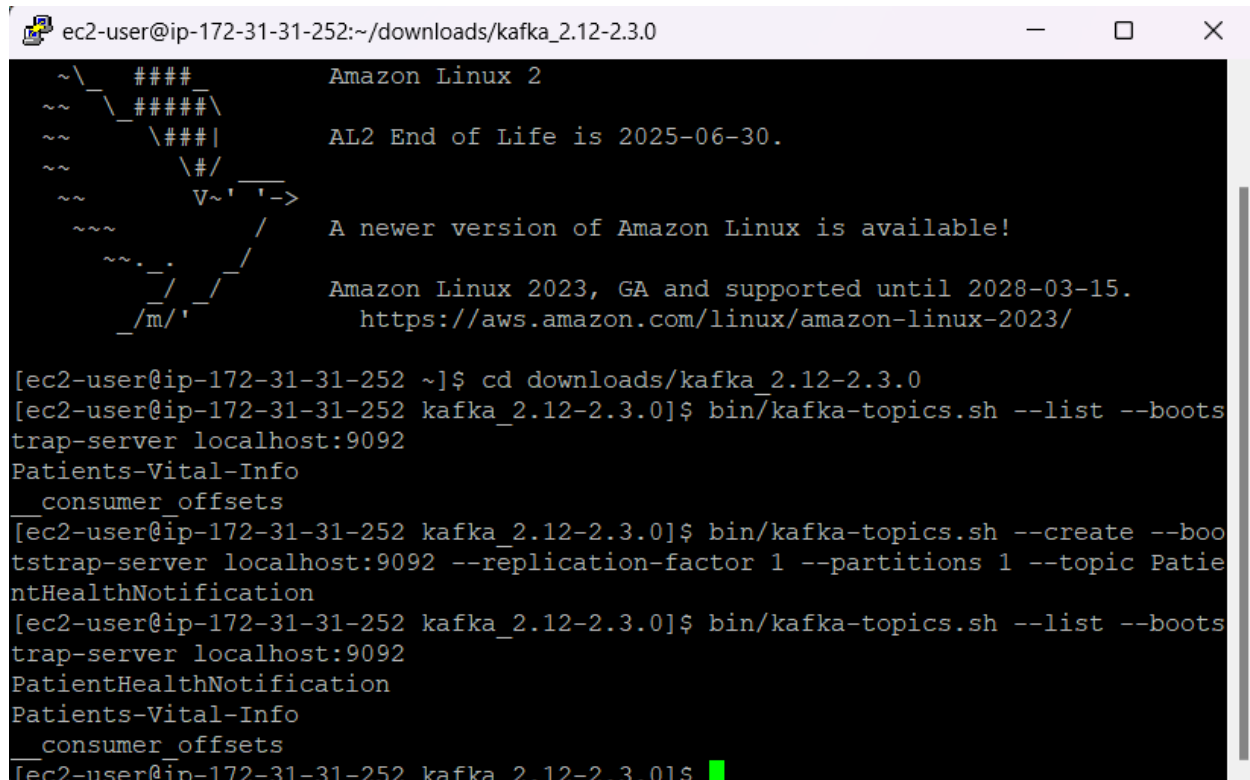
If it was not your intention to subscribe, [click here to unsubscribe](#).

STEP 5: AS STATUS OF SUBSCRIPTION IS SWITCHED FROM PENDING CONFIRMATION TO CONFIRMED:



STEP 8: Creation of PatientHealthNotification topic for receiving the alerts to the KafkaQueue from where SNS will consume

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



```
ec2-user@ip-172-31-31-252:~/downloads/kafka_2.12-2.3.0
~\ ##### Amazon Linux 2
~~ \ #####\
~~ \###| AL2 End of Life is 2025-06-30.
~~ \#/
~~ V~' '->
~~~
~~~. / A newer version of Amazon Linux is available!
~~ / Amazon Linux 2023, GA and supported until 2028-03-15.
_/m/' https://aws.amazon.com/linux/amazon-linux-2023/

[ec2-user@ip-172-31-31-252 ~]$ cd downloads/kafka_2.12-2.3.0
[ec2-user@ip-172-31-31-252 kafka_2.12-2.3.0]$ bin/kafka-topics.sh --list --bootstrap-server localhost:9092
Patients-Vital-Info
__consumer_offsets
[ec2-user@ip-172-31-31-252 kafka_2.12-2.3.0]$ bin/kafka-topics.sh --create --bootstrap-server localhost:9092 --replication-factor 1 --partitions 1 --topic PatientHealthNotification
[ec2-user@ip-172-31-31-252 kafka_2.12-2.3.0]$ bin/kafka-topics.sh --list --bootstrap-server localhost:9092
PatientHealthNotification
Patients-Vital-Info
__consumer_offsets
[ec2-user@ip-172-31-31-252 kafka_2.12-2.3.0]$
```

STEP 9: EXECUTION OF PRODUCER SAPRK APPLICATION AND CONSUMER APPLICATION

Execution of Spark streaming application to read data from HDFS and compare it with hbase and transfer the output data to PatientHealthNotification topic:

```
spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.12:3.3.1
kafka_spark_generate_alerts.py
```

Execution of Consumer application to receive email regarding generated alerts received on PatientHealthNotification topic

```
python3 kafka_consume_alerts.py
```

STEP 10: Soon we were able to receive Alerts for first user as soon as Higher BP than normal was caught up by the spark streaming application as mentioned in below screenshot:

