

Cluster: 5 30 1

- Spark 2.4.5
- Hive 2.3.6
- HBase 1.4.13
- Sqoop 1.4.7
- HCatalog 2.3.6
- Kafka 2.3.0 needs to be downloaded separately

Installing Kafka And Other Packages:

```
wget https://de-mysql-connector.s3.amazonaws.com/mysql-con
```

```
sudo cp ./mysql-connector-java-8.0.25/mysql-connector-java-8.0.25.jar /usr/lib/sqoop/lib/
sudo pip3 install kafka-python
sudo pip3 install mysql-connector
sudo pip3 install boto3
```

Import Required Libraries From Hive and Hbase To Access Through Spark:

```
sudo cp /usr/lib/hive/lib/hive-hbase-handler* /usr/lib/spark/jars/
sudo cp /usr/lib/hbase/lib/hbase-common* /usr/lib/spark/jars/
sudo cp /usr/lib/hbase/lib/hbase-client* /usr/lib/spark/jars/
sudo cp /usr/lib/hbase/lib/hbase-server* /usr/lib/spark/jars/
sudo cp /usr/lib/hbase/lib/hbase-protocol* /usr/lib/spark/jars/
sudo cp /usr/lib/hbase/lib/htrace-core* /usr/lib/spark/jars/
sudo cp /usr/lib/hbase/lib/hbase-metrics* /usr/lib/spark/jars/
sudo cp /usr/lib/hbase/lib/metrics-core-2.2.0* /usr/lib/spark/jars/
```

Starting Kafka Server:

3.0/config/server.properties

Kafka topics:

```
calhost:9092 -- rep
```

```
/home/hadoop/kafka_2.12-2.3.0/bin/kafka-topics.sh --list --bootstrap-server localhost:9092
```

Screenshots

```
100% |████████████████████████████████████████| 143kB 4.2M/s
Collecting boto3<1.30.0,>=1.29.95 (from boto3)
  Downloading https://files.pythonhosted.org/packages/
100% |████████████████████████████████████████| 10.5MB 10M/s
Requirement already satisfied: inetspath<2.0.0,>=0.7.1
```

[illegible]

Scanned with CamScanner

```

[2023-02-21 06:35:52,481] INFO [kafka.server.election.enable = false]
zkcooper:connect = localhost:2181
zkcooper.connection.timeout.ms = 6000
zkcooper.max.in.flight.requests = 10
zkcooper.session.timeout.ms = 6000
zkcooper.startval = false
zkcooper.sync.time.ms = 2000
(kafka.network.SocketServer)
[2023-02-21 06:35:52,486] INFO [ThreadedChannelReceiver-Petch]: Starting (kafka.server.ClientQuotaManager$ThreadedChannelReceiver)
[2023-02-21 06:35:52,491] INFO [ThreadedChannelReceiver-Petch]: Starting (kafka.server.ClientQuotaManager$ThreadedChannelReceiver)
[2023-02-21 06:35:52,492] INFO [ThreadedChannelReceiver-Request]: Starting (kafka.server.ClientQuotaManager$ThreadedChannelReceiver)
[2023-02-21 06:35:52,504] INFO log directory /tmp/kafka-logs was found, creating it. (kafka.log.LogManager)
[2023-02-21 06:35:52,507] INFO loading logs (kafka.log.LogManager)
[2023-02-21 06:35:52,520] INFO Logs loading complete in 8 ms. (kafka.log.LogManager)
[2023-02-21 06:35:52,521] INFO Starting log cleanup with a period of 300000 ms. (kafka.log.LogManager)
[2023-02-21 06:35:52,548] INFO The broker will flush with a default period of 822337203684575007 ms. (kafka.log.LogManager)
[2023-02-21 06:35:53,001] INFO Waiting socket connections on 0.0.0.0:9092. (kafka.network.Acceptor)
[2023-02-21 06:35:53,550] INFO [BlockCacheBrokerId=0] Created data-plane acceptor and processors for endpoint : Endpoint(m1,9092,ListNameName(PAINTEXT),PAINTEXT) (kafka.network.SocketServer)
[2023-02-21 06:35:53,720] INFO [BlockCacheBrokerId=0] Started 1 acceptor threads for data-plane (kafka.network.SocketServer)
[2023-02-21 06:35:53,429] INFO [ExpirationBaper-0-Produce]: Starting (kafka.server.DelayedOperationPurgatory$DelayedOperationBaper)
[2023-02-21 06:35:53,429] INFO [ExpirationBaper-0-Petch]: Starting (kafka.server.DelayedOperationPurgatory$DelayedOperationBaper)
[2023-02-21 06:35:53,430] INFO [ExpirationBaper-0-BaleteSeconds]: Starting (kafka.server.DelayedOperationPurgatory$DelayedOperationBaper)
[2023-02-21 06:35:53,431] INFO [ExpirationBaper-0-DeleteRecords]: Starting (kafka.server.DelayedOperationPurgatory$DelayedOperationBaper)
[2023-02-21 06:35:53,449] INFO [ExpirationValueAndHeader]: Starting (kafka.server.ReplicaCoordinatorLogInitialHeaderHandler)
[2023-02-21 06:35:53,450] INFO [ExpirationValueAndHeader]: Starting (kafka.server.ReplicaCoordinatorLogInitialHeaderHandler)
[2023-02-21 06:35:53,719] INFO State of the created zknode at /brokers/dto/1 is: 9x,9,16793055370,167930553710,1.0,705761633574918,228,9,x
(kafka.kc.KafkaKcClient)
[2023-02-21 06:35:53,720] INFO Registered broker 0 at path /brokers/dto/1 with address: ArrayBuffer(Endpoint(ip=172-31-88-146.ec2.internal,9092,ListNameName(PAINTEXT),PAINTEXT)), czxid=
broker epoch= 9 (kafka.kc.KafkaKcClient)
[2023-02-21 06:35:53,722] WARN No meta.properties file under dir /tmp/kafka-logs/user_data/kafka.server.RocketMetaDataCheckpoint)
[2023-02-21 06:35:53,826] INFO [ExpirationBaper-0-Topic]: Starting (kafka.server.DelayedOperationPurgatory$DelayedOperationBaper)
[2023-02-21 06:35:53,827] INFO [ExpirationBaper-0-GroupMetadata]: Starting (kafka.server.DelayedOperationPurgatory$DelayedOperationBaper)
[2023-02-21 06:35:53,833] INFO [ExpirationBaper-0-Rebalance]: Starting (kafka.server.DelayedOperationPurgatory$DelayedOperationBaper)
[2023-02-21 06:35:53,834] INFO [GroupCoordinator-0]: Starting up. (kafka.coordinator.group.GroupCoordinator)
[2023-02-21 06:35:53,872] INFO [GroupCoordinator-0]: Starting up. (kafka.coordinator.group.GroupCoordinator)
[2023-02-21 06:35:53,880] INFO Successfully created controller scope with initial epoch 0 (kafka.kc.KafkaKcClient)
[2023-02-21 06:35:53,880] INFO [GroupMetadataManager-knowledge] Removed 0 expired offsets in 8 milliseconds. (kafka.coordinator.group.GroupMetadataManager)
[2023-02-21 06:35:53,897] INFO [FollowerManager-0]: Acquired new producer block (brokerId=0,blockStart=0,producerId=0,blockEnd=0,controller=999) by writing to zk with path version 1 (kafka.c
CoordinatorManager)
[2023-02-21 06:35:53,897] INFO [TransactionCoordinator-0] Starting up. (kafka.coordinator.transaction.TransactionCoordinator)
[2023-02-21 06:35:53,900] INFO [TransactionCoordinator-0] Starting (kafka.coordinator.transaction.TransactionCoordinator)
[2023-02-21 06:35:53,941] INFO [TransactionCoordinator-0] Startup complete. (kafka.coordinator.transaction.TransactionCoordinator)
[2023-02-21 06:35:54,069] INFO [ConfigChange-event-process-thread]: Starting (kafka.com.common.config.change-notification.listener.ConfigChangeEventProcessThread)
[2023-02-21 06:35:54,080] INFO [BlockCacheBrokerId=0] Started delay-plane acceptors for 1 acceptor (kafka.network.SocketServer)
[2023-02-21 06:35:54,041] INFO Kafka version: 2.9.0 (org.apache.kafka.common.utils.AppInfoParser)
[2023-02-21 06:35:54,041] INFO Kafka commitId=fcbab18dcfc (org.apache.kafka.common.utils.AppInfoParser)
[2023-02-21 06:35:54,041] INFO Kafka version: 2.9.0 (org.apache.kafka.common.utils.AppInfoParser)
[2023-02-21 06:35:54,041] INFO Kafka commitId=fcbab18dcfc (org.apache.kafka.common.utils.AppInfoParser)
[2023-02-21 06:35:54,050] INFO [KafkaServer id=0] started (kafka.server.KafkaServer)

```

[illegible]

```
: 4, 'heartBeat': 67, 'bp': 153} +-----+-----+-----+-----+
: 5, 'heartBeat': 67, 'bp': 178} |customerId|heartBeat|bp |message_time
: 1, 'heartBeat': 75, 'bp': 182} +-----+-----+-----+-----+
```

```

('customerid': 2, 'heartBeat': 67, 'bp': 170)
('customerid': 3, 'heartBeat': 72, 'bp': 170)
('customerid': 4, 'heartBeat': 67, 'bp': 173)
('customerid': 5, 'heartBeat': 67, 'bp': 170)
('customerid': 1, 'heartBeat': 74, 'bp': 172)
('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)
]

Batch: 151
[customerid|heartBeat|bp |message_time |
+-----+
12 | 170 | 153 | 2023-03-21 07:18:40.001 |
13 | 167 | 154 | 2023-03-21 07:18:40.001 |
14 | 170 | 174 | 2023-03-21 07:18:40.001 |
15 | 167 | 177 | 2023-03-21 07:18:40.001 |
16 | 167 | 167 | 2023-03-21 07:18:40.001 |
12 | 189 | 174 | 2023-03-21 07:18:40.001 |
13 | 171 | 155 | 2023-03-21 07:18:40.001 |
14 | 173 | 175 | 2023-03-21 07:18:40.001 |
15 | 169 | 161 | 2023-03-21 07:18:40.001 |
16 | 175 | 207 | 2023-03-21 07:18:40.001 |
+-----+

Batch: 152
[customerid|heartBeat|bp |message_time |
+-----+
12 | 166 | 159 | 2023-03-21 07:18:50.001 |
13 | 171 | 153 | 2023-03-21 07:18:50.001 |
14 | 172 | 153 | 2023-03-21 07:18:50.001 |
15 | 166 | 171 | 2023-03-21 07:18:50.001 |
+-----+
]

[hadoop@ip-172-31-88-146 ~]$

```

```

xr-xr-x    - hadoop hadoop      0 2023-03-21 07:18 /user/ha
-r--r--r-- 1 hadoop hadoop     1294 2023-03-21 07:16 /user/ha
-rw-rw-rw- 1 hadoop hadoop     162e12b5fd3f-c000.snappy.parquet
-r--r--r-- 1 hadoop hadoop     1282 2023-03-21 06:55 /user/ha

```

57-15eb5c1a4d3c-c000, snappy.parquet	1282	2023-03-21	07:18	/user/hadoop/health-alert/patients-vital-info/part-00000-021719344-539e-41de-83
57-rw-r--r- 1 hadoop hadoop				
57-02194855c65c3-c000, snappy.parquet	1294	2023-03-21	07:04	/user/hadoop/health-alert/patients-vital-info/part-00000-0295b359-b870-404a-ab
57-rw-r--r- 1 hadoop hadoop				
a3-d581abe3a12c-c000, snappy.parquet	1312	2023-03-21	07:15	/user/hadoop/health-alert/patients-vital-info/part-00000-031227f7-91a7-4cd5-9b
57-rw-r--r- 1 hadoop hadoop				
57-02194855c65c3-c000, snappy.parquet	1312	2023-03-21	07:08	/user/hadoop/health-alert/patients-vital-info/part-00000-0606a1b4d7-80af-4e47-97-93
57-rw-r--r- 1 hadoop hadoop				
4e-e877743d32d9-c000, snappy.parquet	1282	2023-03-21	07:05	/user/hadoop/health-alert/patients-vital-info/part-00000-0707ac71-7271-4b94-a6
57-rw-r--r- 1 hadoop hadoop				
57-02194855c65c3-c000, snappy.parquet	1268	2023-03-21	07:03	/user/hadoop/health-alert/patients-vital-info/part-00000-072879749-026b-41f0-9c
57-rw-r--r- 1 hadoop hadoop				
bl-112405553be2-c000, snappy.parquet	1294	2023-03-21	07:02	/user/hadoop/health-alert/patients-vital-info/part-00000-0a025b6d-4308-4132-83
57-rw-r--r- 1 hadoop hadoop				
57-02194855c65c3-c000, snappy.parquet	1282	2023-03-21	07:06	/user/hadoop/health-alert/patients-vital-info/part-00000-0efb7ab0-2256-42ab-8f
57-rw-r--r- 1 hadoop hadoop				
b4-788f497ad1f9-c000, snappy.parquet	1308	2023-03-21	07:04	/user/hadoop/health-alert/patients-vital-info/part-00000-0f93b3bb-075e-4e2b-ad
57-rw-r--r- 1 hadoop hadoop				
3c-dba7e46350af-c000, snappy.parquet	1289	2023-03-21	07:02	/user/hadoop/health-alert/patients-vital-info/part-00000-10f24b16-85b2-4537-b5
57-rw-r--r- 1 hadoop hadoop				
0e-0bae514893d-c000, snappy.parquet	1291	2023-03-21	07:17	/user/hadoop/health-alert/patients-vital-info/part-00000-131d14d4-0a14-4228-9e
57-rw-r--r- 1 hadoop hadoop				
57-02194855c65c3-c000, snappy.parquet	1265	2023-03-21	07:13	/user/hadoop/health-alert/patients-vital-info/part-00000-133d1d46-9e65-4af1-95
57-rw-r--r- 1 hadoop hadoop				

```
3ca446f02e0e-c000.snappy.parquet
-r--r-- 1 hadoop hadoop 1286 2023-03-21 07:14 /user/hadoop/
bc93f7461f24-c000.snappy.parquet
-r--r-- 1 hadoop hadoop 1312 2023-03-21 06:59 /user/hadoop/
```

```
Using '-cat' To Read One Of The File
```

```
hadoop@ip-172-31-88-146:~$ hadoop hadoopfs ls -l /user/hadoop/health-alert/patients-vital-info/part-00000-fd66228d-eel2-4618-9d83-551add52e482-c000.snappy.parquet  
[hadoop@ip-172-31-88-146 ~]$ hadoop fs -cat /user/hadoop/health-alert/patients-vital-info/part-00000-0052cb92-31eb-4dae-b557-15ebc51a4a8d-c000.snappy.parquet  
{"id": "42FDF#(,rL",  
  "name": "LIC(IC4[Pss\\x[.H[  
  "age": 25,  
  "sex": "M",  
  "height": 175,  
  "weight": 70,  
  "blood_pressure": "120/80",  
  "heart_rate_bpm": 75,  
  "temperature_c": 37.5,  
  "oxygen_saturation_percent": 98,  
  "diagnosis": "Heart Disease",  
  "treatment_plan": "Medication and lifestyle changes",  
  "last_updated": "2023-03-21T10:30:00Z"}  
}
```