



# APACHE KAFKA LABS

Bin Jiang

03/18/2017

# Configuring Kafka

- Running Kafka

The screenshot displays the Ambari web console interface. On the left is a sidebar with a list of services: HDFS, YARN, MapReduce2, Tez, Hive, HBase, Pig, Sqoop, Oozie, ZooKeeper, Falcon, Storm, Flume, Ambari Infra, Atlas, Kafka, and Knox. The 'Kafka' service is selected and highlighted. The main content area shows the 'Summary' tab for the 'Kafka Broker' service. The status is 'Stopped' with a red warning icon and a green 'No alerts' badge. Below the summary, the 'Metrics' section contains five panels: 'Broker Topics', 'Active Controller Count', 'Controller Status', 'Replica MaxLag', and an unlabeled panel, all displaying 'No Data Available'. A 'Service Actions' dropdown menu is open on the right, listing options: Start, Stop, Restart All, Restart Kafka Brokers, Run Service Check, Turn On Maintenance Mode, and Delete Service.

Summary    Configs    Service Actions ▾

**Summary**

[Kafka Broker](#) ⚠ Stopped **No alerts**

**Metrics**

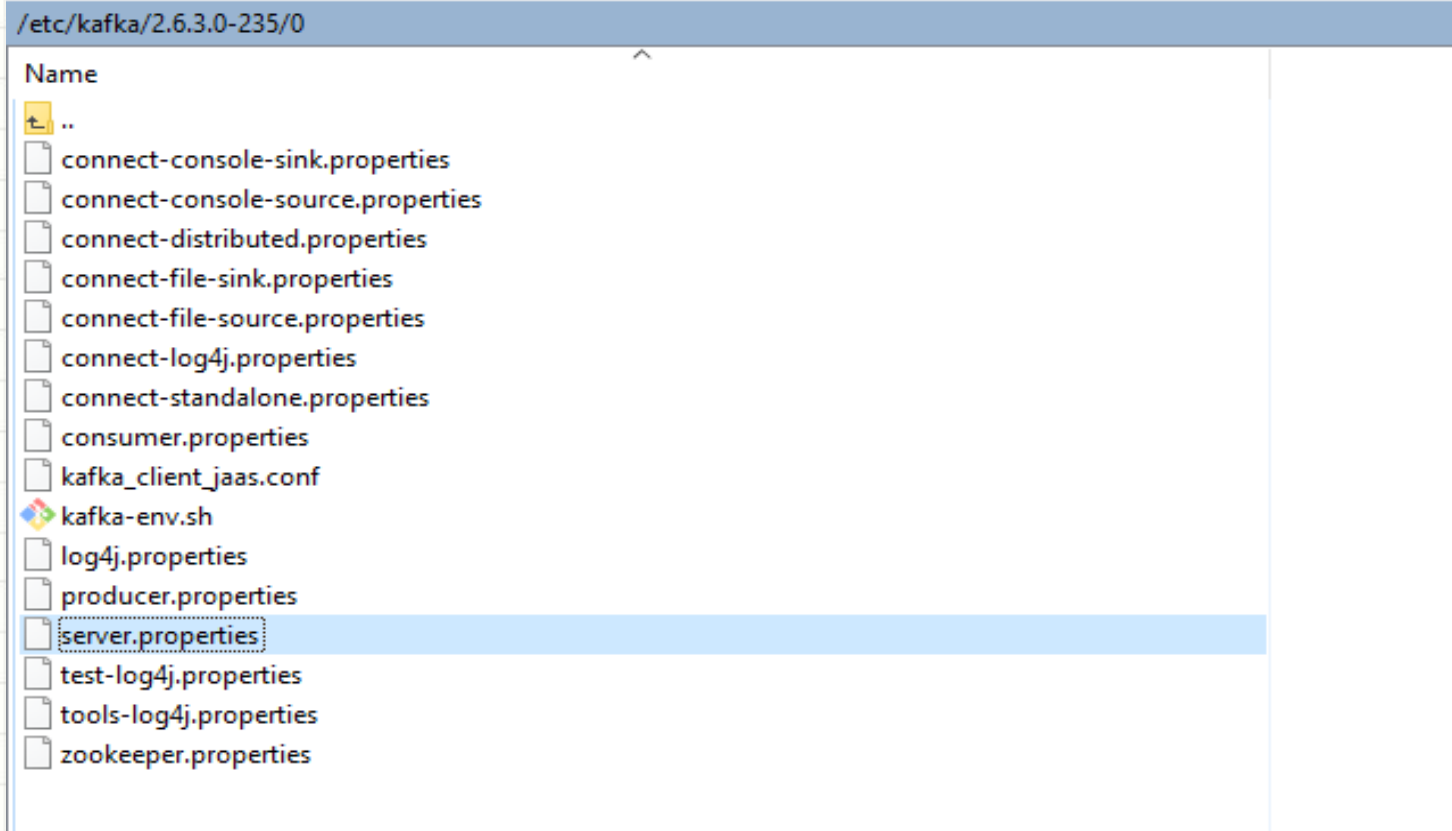
Broker Topics	Active Controller Count	Controller Status	Replica MaxLag	
No Data Available	No Data Available	No Data Available	No Data Available	No Data Available

Service Actions ▾

- ▶ Start
- Stop
- 🔄 Restart All
- 🔄 Restart Kafka Brokers
- 🔍 Run Service Check
- 🛑 Turn On Maintenance Mode
- ✖ Delete Service

# Configuring Kafka

- **Configuring Kafka brokers**





# Configuring Kafka

- **Configuring Kafka brokers**

```
external.kafka.metrics.include.prefix=kafka.network.RequestMetrics.ResponseQueue
fetch.purgatory.purge.interval.requests=10000
kafka.ganglia.metrics.group=kafka
kafka.ganglia.metrics.host=localhost
kafka.ganglia.metrics.port=8671
kafka.ganglia.metrics.reporter.enabled=false
kafka.timeline.metrics.host=
kafka.timeline.metrics.hosts=
kafka.timeline.metrics.maxRowCacheSize=10000
kafka.timeline.metrics.port=
kafka.timeline.metrics.protocol=
kafka.timeline.metrics.reporter.enabled=false
kafka.timeline.metrics.reporter.sendInterval=5900
kafka.timeline.metrics.truststore.password=
kafka.timeline.metrics.truststore.path=
kafka.timeline.metrics.truststore.type=
leader.imbalance.check.interval.seconds=300
leader.imbalance.per.broker.percentage=10
listeners=PLAINTEXT://sandbox-hdp.hortonworks.com:6667
log.cleanup.interval.mins=10
log.dirs=/kafka-logs
log.index.interval.bytes=4096
log.index.size.max.bytes=10485760
log.retention.bytes=-1
log.retention.hours=168
```

# Configuring Kafka


- **Configuring Kafka brokers**

 **V4** 

admin authored on Fri, Nov 10, 2017 09:58

Discard


Save

 Kafka Broker



Kafka Broker host

sandbox-hdp.hortonworks.com



zookeeper.connect




log.dirs



log.roll.hours

log.retention.hours

listeners

# Configuring Kafka

- Configuring Kafka topics

```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-topics.sh --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic class9Topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
Created topic "class9Topic".
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-topics.sh --list --zookeeper localhost:2181 class9Topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
ATLAS_ENTITIES
ATLAS_HOOK
__confluent.support.metrics
__consumer_offsets
__schemas
class9Topic
demoTopic
mysql-jdbc-demo_products
mysql-jdbc-demo_products1
nifi_websocket
partition-topic
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-topics.sh --describe --zookeeper localhost:2181 --topic class9Topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
Topic: class9Topic    PartitionCount:1    ReplicationFactor:1    Configs:
    Topic: class9Topic    Partition: 0    Leader: 1001    Replicas: 1001    Isr: 1001
[root@sandbox-hdp ~]#
```

# Configuring Kafka

- **Configuring Kafka topics**

```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-topics.sh --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic replicatedClass9Topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
Created topic "replicatedClass9Topic".
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-topics.sh --describe --zookeeper localhost:2181 --topic replicatedClass9Topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
Topic:replicatedClass9Topic    PartitionCount:1      ReplicationFactor:1    Configs:
      Topic: replicatedClass9Topic    Partition: 0    Leader: 1001    Replicas: 1001    Isr: 1001
[root@sandbox-hdp ~]#
```



# Configuring Kafka

- Creating a message console producer

```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-console-producer.sh --broker-list sandbox-hdp.hortonworks.com:6667 --topic class9Topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
Class 9 Big Data Streaminn Processing
Kafka is one of the distributed message systems
█
```

```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-console-producer.sh --broker-list sandbox-hdp.hortonworks.com:6667 --topic class9Topic < /root/TrainingOnHDP/dataset/message.txt
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
[root@sandbox-hdp ~]# █
```

```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-console-consumer.sh --topic class9Topic --bootstrap-server sandbox-hdp.hortonworks.com:6667 --from-beginning
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
Class 9 Big Data Streaminn Processing
Kafka is one of the distributed message systems
Class 10 Spark ETL
Class 6 ETL using NiFi
█
```



# Configuring Kafka

- **Creating a message console producer**

```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-console-producer.sh --broker-list sandbox-hdp.hortonworks.com:6667 --topic class9Topic < /root/TrainingOnHDP/dataset/message.txt
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
[root@sandbox-hdp ~]#
```

```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-console-consumer.sh --topic class9Topic --bootstrap-server sandbox-hdp.hortonworks.com:6667 --max-messages 1
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
Class 10 Spark ETL
Processed a total of 1 messages
[root@sandbox-hdp ~]#
```

# Configuring Kafka

- Creating a message console consumer

```
[root@sandbox-hdp dataset]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-console-producer.sh --broker-list sandbox-hdp.hortonworks.com:6667 --topic class9Topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
Class 8 OLAP over Hadoop
Class 7 NoSQL over Hadoop
Class 7 NoSQL over Hadoop
```

```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-console-consumer.sh --topic class9Topic --bootstrap-server sandbox-hdp.hortonworks.com:6667 --new-consumer --consumer-property group.id=consumerGroup
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
Class 8 OLAP over Hadoop
^CProcessed a total of 1 messages
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-console-consumer.sh --topic class9Topic --bootstrap-server sandbox-hdp.hortonworks.com:6667 --new-consumer --consumer-property group.id=consumerGroup
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
Class 7 NoSQL over Hadoop
Class 7 NoSQL over Hadoop
```

# Configuring Kafka




























- Configuring the broker settings

The screenshot displays the Ambari web console interface for configuring Kafka. On the left is a sidebar with a list of services: HDFS, YARN, MapReduce2, Tez, Hive, HBase, Pig, Sqoop, Oozie, ZooKeeper, Falcon, Storm, Flume, Ambari Infra, Atlas, Kafka (selected), Knox, Ranger, and Spark2. The main content area is titled 'Summary' and 'Configs'. Below this, there's a 'Group' dropdown set to 'Default (1)' and a 'Manage Config Groups' link. A 'Filter...' input field is also present. A horizontal scrollable list of configuration versions (V1 to V5) is shown, with V5 being the active version. Below this, a status bar indicates 'V5 admin authored on Thu, Jan 18, 2018 20:58' with 'Discard' and 'Save' buttons. The 'Kafka Broker' configuration section is expanded, showing the following settings:

Property	Value	Actions
Kafka Broker host	sandbox-hdp.hortonworks.com	
zookeeper.connect	sandbox-hdp.hortonworks.com:2181	Lock, Refresh, Copy
log.dirs	/kafka-logs	Lock, Refresh, Copy
log.roll.hours	168	Lock, Refresh, Copy
log.retention.hours	168	Lock, Refresh, Copy
listeners	PLAINTEXT://localhost:6667	Lock, Refresh, Copy

# Configuring Kafka

- **Configuring threads and performance**

message.max.bytes	<input type="text" value="1000000"/>			
min.insync.replicas	<input type="text" value="1"/>			
num.io.threads	<input type="text" value="8"/>			
num.network.threads	<input type="text" value="3"/>			
num.partitions	<input type="text" value="1"/>			
num.recovery.threads. per.data.dir	<input type="text" value="1"/>			
num.replica.fetchers	<input type="text" value="1"/>			
offset.metadata.max. bytes	<input type="text" value="4096"/>			
offsets.commit.required. acks	<input type="text" value="-1"/>			

























# Configuring Kafka

- **Configuring the log settings**

kafka.timeline.metrics.truststore.path	<input type="text" value="{{metric_truststore_path}}"/>			
kafka.timeline.metrics.truststore.type	<input type="text" value="{{metric_truststore_type}}"/>			
leader.imbalance.check.interval.seconds	<input type="text" value="300"/>			
leader.imbalance.per.broker.percentage	<input type="text" value="10"/>			
log.cleanup.interval.mins	<input type="text" value="10"/>			
log.index.interval.bytes	<input type="text" value="4096"/>			
log.index.size.max.bytes	<input type="text" value="10485760"/>			
log.retention.bytes	<input type="text" value="-1"/>			
log.segment.bytes	<input type="text" value="1073741824"/>			
message.max.bytes	<input type="text" value="1000000"/>			
min.insync.replicas	<input type="text" value="1"/>			
num.io.threads	<input type="text" value="8"/>			


# Configuring Kafka

- **Configuring the replica settings**

replica.fetch.min.bytes	<input type="text" value="1"/>			
replica.fetch.wait.max.ms	<input type="text" value="500"/>			
replica.high.watermark.checkpoint.interval.ms	<input type="text" value="5000"/>			
replica.lag.max.messages	<input type="text" value="4000"/>			
replica.lag.time.max.ms	<input type="text" value="10000"/>			
replica.socket.receive.buffer.bytes	<input type="text" value="65536"/>			
replica.socket.timeout.ms	<input type="text" value="30000"/>			
socket.receive.buffer.bytes	<input type="text" value="102400"/>			

# Configuring Kafka

- **Configuring the Zookeeper settings**

zookeeper.connection.timeout.ms	<input type="text" value="25000"/>	  
zookeeper.session.timeout.ms	<input type="text" value="30000"/>	  
zookeeper.sync.time.ms	<input type="text" value="2000"/>	  



# Configuring Kafka

- **Configuring other miscellaneous parameters**

offsets.commit.required.acks	<input type="text" value="-1"/>			
offsets.commit.timeout.ms	<input type="text" value="5000"/>			
offsets.load.buffer.size	<input type="text" value="5242880"/>			
offsets.retention.check.interval.ms	<input type="text" value="600000"/>			
offsets.retention.minutes	<input type="text" value="86400000"/>			
offsets.topic.compression.codec	<input type="text" value="0"/>			
offsets.topic.num.partitions	<input type="text" value="50"/>			
offsets.topic.replication.factor	<input type="text" value="3"/>			
offsets.topic.segment.bytes	<input type="text" value="104857600"/>			
port	<input type="text" value="6667"/>			

# Message Validation

```
2. localhost (root)
[root@sandbox-hdp dataset]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-topics.sh --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic source-topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
Created topic "source-topic".
[root@sandbox-hdp dataset]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-topics.sh --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic good-topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
Created topic "good-topic".
[root@sandbox-hdp dataset]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-topics.sh --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic bad-topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
Created topic "bad-topic".
[root@sandbox-hdp dataset]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-topics.sh --list --zookeeper localhost:2181
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
ATLAS_ENTITIES
ATLAS_HOOK
__confluent.support.metrics
__consumer_offsets
__schemas
bad-topic
class9Topic
demoTopic
good-topic
mysql-jdbc-demo_products
mysql-jdbc-demo_products1
nifi_websocket
partition-topic
replicatedClass9Topic
source-topic
[root@sandbox-hdp dataset]#
```

# Message Validation

```
[root@sandbox-hdp dataset]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-console-producer.sh --broker-list sandbox-hdp.hortonworks.com:6667 --topic source-topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
{"event": "CUSTOMER_SEES_BTCPRI", "customer": {"id": "86689427", "name": "Edward S.", "ipAddress": "95.31.18.119"}, "currency": {"name": "bitcoin", "price": "USD"}, "timestamp": "2017-07-03T12:00:35Z"}
{"event": "CUSTOMER_SEES_BTCPRI", "customer": {"id": "18313440", "name": "Julian A.", "ipAddress": "185.86.151.11"}, "currency": {"name": "bitcoin", "price": "USD"}, "timestamp": "2017-07-04T15:00:35Z"}
{"event": "CUSTOMER_SEES_BTCPRI", "customer": {"id": "56886468", "name": "Lindsay M.", "ipAddress": "186.46.129.15"}, "currency": {"name": "bitcoin", "price": "USD"}, "timestamp": "2017-07-11T19:00:35Z"}
This is the bad message
```

```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-console-consumer.sh --bootstrap-server sandbox-hdp.hortonworks.com:6667 --topic good-topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
{"event": "CUSTOMER_SEES_BTCPRI", "customer": {"id": "86689427", "name": "Edward S.", "ipAddress": "95.31.18.119"}, "currency": {"name": "bitcoin", "price": "USD"}, "timestamp": "2017-07-03T12:00:35Z"}
{"event": "CUSTOMER_SEES_BTCPRI", "customer": {"id": "18313440", "name": "Julian A.", "ipAddress": "185.86.151.11"}, "currency": {"name": "bitcoin", "price": "USD"}, "timestamp": "2017-07-04T15:00:35Z"}
{"event": "CUSTOMER_SEES_BTCPRI", "customer": {"id": "56886468", "name": "Lindsay M.", "ipAddress": "186.46.129.15"}, "currency": {"name": "bitcoin", "price": "USD"}, "timestamp": "2017-07-11T19:00:35Z"}
```

# Message Validation

```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-console-consumer.sh --bootstrap-server sandbox-hdp.hortonworks.com:6667 --from-beginning --topic bad-topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
{"error": "JsonParseException: Unrecognized token 'This': was expecting ('true', 'false' or 'null')
at [Source: This is the bad message; line: 1, column: 5]"}

```

```
[root@sandbox-hdp ~]# java -cp /root/TrainingOnHDP/StreamingApplicationOnKafka/target/StreamingApplicationOnKafka-1.0-SNAPSHOT-jar-with-dependencies.jar ca.training.bigdata.kafka.validation.ProcessingApp sandbox-hdp.hortonworks.com:6667 consumerGroup source-topic good-topic bad-topic
log4j:WARN No appenders could be found for logger (org.apache.kafka.clients.consumer.ConsumerConfig).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.

```

# Message Enrichment

- **Download a free copy of the MaxMind GeoIP database**



```
[root@sandbox-hdp ~]# cd /root/TrainingOnHDP/dataset
[root@sandbox-hdp dataset]# wget "http://geolite.maxmind.com/download/geoip/database/GeoLiteCity.dat.gz"
--2018-01-20 02:38:34-- http://geolite.maxmind.com/download/geoip/database/GeoLiteCity.dat.gz
Resolving geolite.maxmind.com... 104.16.37.47, 104.16.38.47, 2400:cb00:2048:1::6810:262f, ...
Connecting to geolite.maxmind.com|104.16.37.47|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 11283440 (11M) [application/octet-stream]
Saving to: "GeoLiteCity.dat.gz"

100%[=====>] 11,283,440  458K/s  in 17s

2018-01-20 02:38:51 (666 KB/s) - "GeoLiteCity.dat.gz" saved [11283440/11283440]

[root@sandbox-hdp dataset]# gunzip GeoLiteCity.dat.gz
[root@sandbox-hdp dataset]#
```



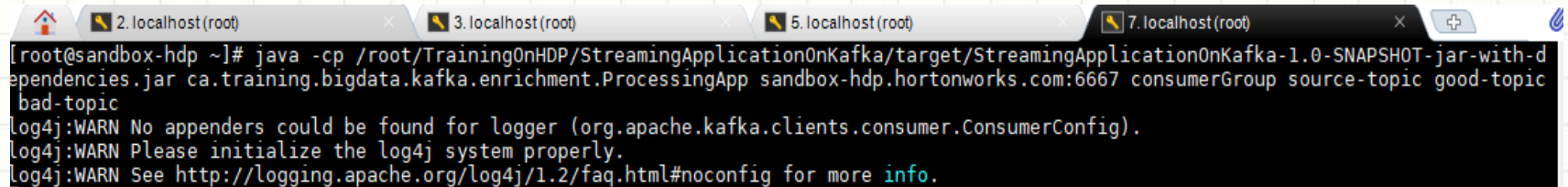
# Message Enrichment

```
2. localhost (root) 3. localhost (root) 5. localhost (root) 7. localhost (root)
[root@sandbox-hdp dataset]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-console-producer.sh --broker-list sandbox-hdp.hortonworks.com:6667 --topic good-topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
{"event": "CUSTOMER_SEES_BTCPPRICE", "customer": {"id": "86689427", "name": "Edward S.", "ipAddress": "95.31.18.119"}, "currency": {"name": "bitcoin", "price": "USD"}, "timestamp": "2017-07-03T12:00:35Z"}
{"event": "CUSTOMER_SEES_BTCPPRICE", "customer": {"id": "18313440", "name": "Julian A.", "ipAddress": "185.86.151.11"}, "currency": {"name": "bitcoin", "price": "USD"}, "timestamp": "2017-07-04T15:00:35Z"}
{"event": "CUSTOMER_SEES_BTCPPRICE", "customer": {"id": "56886468", "name": "Lindsay M.", "ipAddress": "186.46.129.15"}, "currency": {"name": "bitcoin", "price": "USD"}, "timestamp": "2017-07-11T19:00:35Z"}
```

```
2. localhost (root) 3. localhost (root) 5. localhost (root) 7. localhost (root)
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-console-consumer.sh --bootstrap-server sandbox-hdp.hortonworks.com:6667 --topic good-topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
{"event": "CUSTOMER_SEES_BTCPPRICE", "customer": {"id": "86689427", "name": "Edward S.", "ipAddress": "95.31.18.119", "country": "Russian Federation", "city": "Moscow"}, "currency": {"name": "bitcoin", "price": "USD", "rate": 8.5233983E-5}, "timestamp": "2017-07-03T12:00:35Z"}
{"event": "CUSTOMER_SEES_BTCPPRICE", "customer": {"id": "18313440", "name": "Julian A.", "ipAddress": "185.86.151.11", "country": "United Kingdom", "city": "London"}, "currency": {"name": "bitcoin", "price": "USD", "rate": 8.5233983E-5}, "timestamp": "2017-07-04T15:00:35Z"}
{"event": "CUSTOMER_SEES_BTCPPRICE", "customer": {"id": "56886468", "name": "Lindsay M.", "ipAddress": "186.46.129.15", "country": "Ecuador", "city": "Quito"}, "currency": {"name": "bitcoin", "price": "USD", "rate": 8.5233983E-5}, "timestamp": "2017-07-11T19:00:35Z"}
```

# Message Enrichment

```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-console-consumer.sh --bootstrap-server sandbox-hdp.hortonworks.com:6667 --topic bad-topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
```

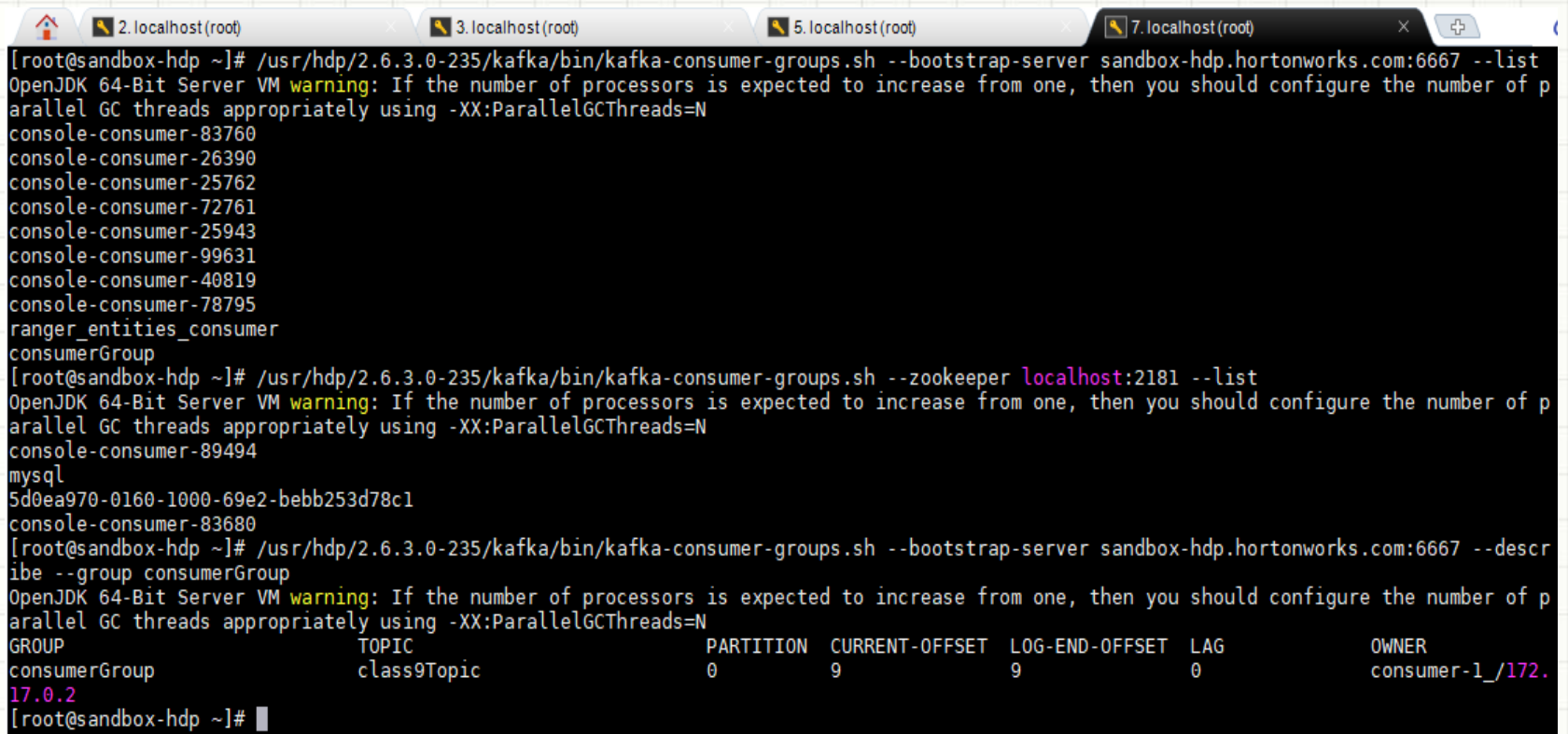


```
2. localhost (root) 3. localhost (root) 5. localhost (root) 7. localhost (root)
[root@sandbox-hdp ~]# java -cp /root/TrainingOnHDP/StreamingApplicationOnKafka/target/StreamingApplicationOnKafka-1.0-SNAPSHOT-jar-with-dependencies.jar ca.training.bigdata.kafka.enrichment.ProcessingApp sandbox-hdp.hortonworks.com:6667 consumerGroup source-topic good-topic bad-topic
log4j:WARN No appenders could be found for logger (org.apache.kafka.clients.consumer.ConsumerConfig).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.
```



# Managing Kafka

- Managing consumer groups

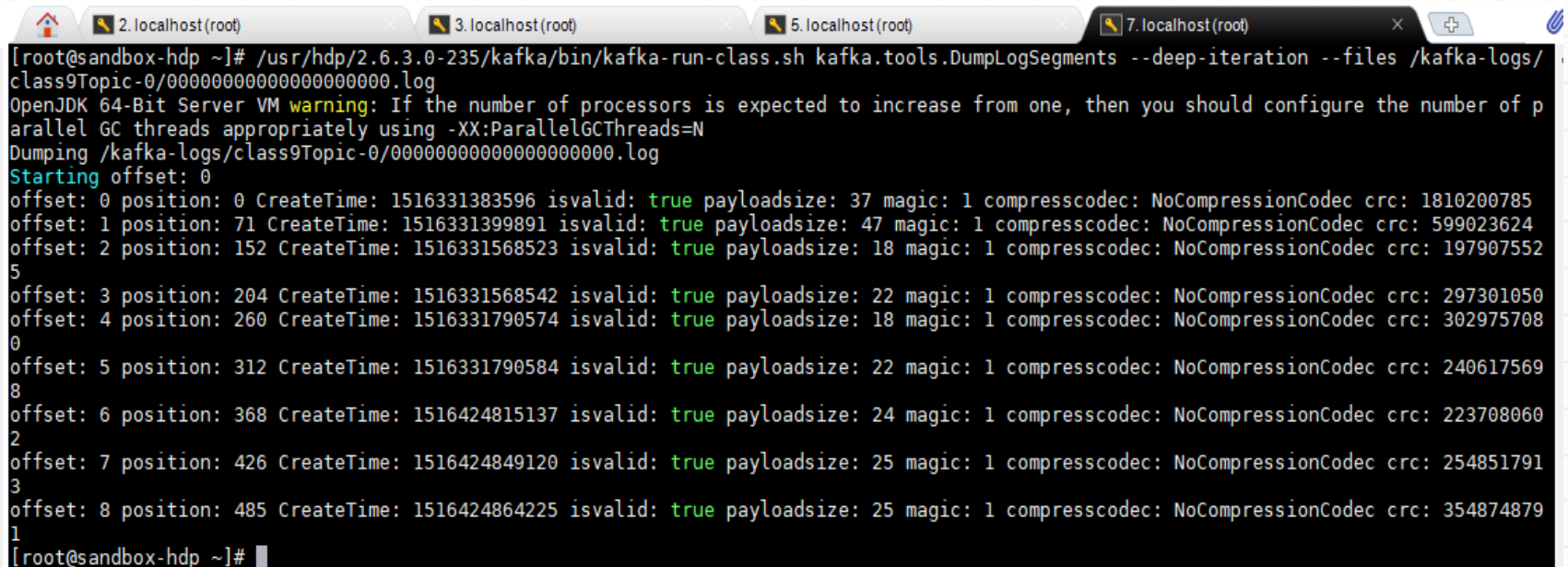


The screenshot shows a terminal window with four tabs: '2. localhost (root)', '3. localhost (root)', '5. localhost (root)', and '7. localhost (root)'. The terminal content is as follows:

```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-consumer-groups.sh --bootstrap-server sandbox-hdp.hortonworks.com:6667 --list
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
console-consumer-83760
console-consumer-26390
console-consumer-25762
console-consumer-72761
console-consumer-25943
console-consumer-99631
console-consumer-40819
console-consumer-78795
ranger_entities_consumer
consumerGroup
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-consumer-groups.sh --zookeeper localhost:2181 --list
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
console-consumer-89494
mysql
5d0ea970-0160-1000-69e2-bebb253d78c1
console-consumer-83680
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-consumer-groups.sh --bootstrap-server sandbox-hdp.hortonworks.com:6667 --describe --group consumerGroup
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
GROUP                                TOPIC                                PARTITION  CURRENT-OFFSET  LOG-END-OFFSET  LAG              OWNER
consumerGroup                        class9Topic                        0           9                9                0              consumer-1_/172.17.0.2
[root@sandbox-hdp ~]#
```

# Managing Kafka

- Dumping log segments

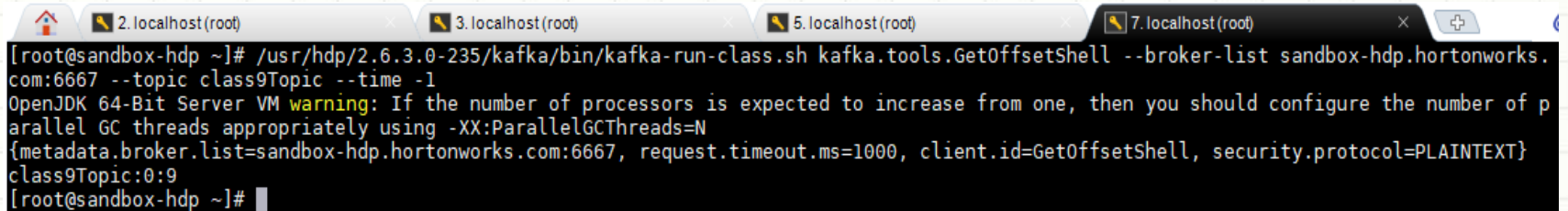


A terminal window with four tabs labeled '2. localhost (root)', '3. localhost (root)', '5. localhost (root)', and '7. localhost (root)'. The active tab shows the following command and output:

```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-run-class.sh kafka.tools.DumpLogSegments --deep-iteration --files /kafka-logs/class9Topic-0/00000000000000000000.log
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
Dumping /kafka-logs/class9Topic-0/00000000000000000000.log
Starting offset: 0
offset: 0 position: 0 CreateTime: 1516331383596 invalid: true payloadsize: 37 magic: 1 compresscodec: NoCompressionCodec crc: 1810200785
offset: 1 position: 71 CreateTime: 1516331399891 invalid: true payloadsize: 47 magic: 1 compresscodec: NoCompressionCodec crc: 599023624
offset: 2 position: 152 CreateTime: 1516331568523 invalid: true payloadsize: 18 magic: 1 compresscodec: NoCompressionCodec crc: 197907552
5
offset: 3 position: 204 CreateTime: 1516331568542 invalid: true payloadsize: 22 magic: 1 compresscodec: NoCompressionCodec crc: 297301050
offset: 4 position: 260 CreateTime: 1516331790574 invalid: true payloadsize: 18 magic: 1 compresscodec: NoCompressionCodec crc: 302975708
0
offset: 5 position: 312 CreateTime: 1516331790584 invalid: true payloadsize: 22 magic: 1 compresscodec: NoCompressionCodec crc: 240617569
8
offset: 6 position: 368 CreateTime: 1516424815137 invalid: true payloadsize: 24 magic: 1 compresscodec: NoCompressionCodec crc: 223708060
2
offset: 7 position: 426 CreateTime: 1516424849120 invalid: true payloadsize: 25 magic: 1 compresscodec: NoCompressionCodec crc: 254851791
3
offset: 8 position: 485 CreateTime: 1516424864225 invalid: true payloadsize: 25 magic: 1 compresscodec: NoCompressionCodec crc: 354874879
1
[root@sandbox-hdp ~]#
```

# Managing Kafka

- Using the GetOffsetShell

A terminal window with four tabs labeled '2. localhost (root)', '3. localhost (root)', '5. localhost (root)', and '7. localhost (root)'. The active tab shows the execution of the command `/usr/hdp/2.6.3.0-235/kafka/bin/kafka-run-class.sh kafka.tools.GetOffsetShell --broker-list sandbox-hdp.hortonworks.com:6667 --topic class9Topic --time -1`. The output includes a warning about the number of processors and GC threads, followed by the command's arguments and the topic name.

```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-run-class.sh kafka.tools.GetOffsetShell --broker-list sandbox-hdp.hortonworks.com:6667 --topic class9Topic --time -1
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
{metadata.broker.list=sandbox-hdp.hortonworks.com:6667, request.timeout.ms=1000, client.id=GetOffsetShell, security.protocol=PLAINTEXT}
class9Topic:0:9
[root@sandbox-hdp ~]#
```

# Managing Kafka

- Using the JMX tool

```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-run-class.sh kafka.tools.JmxTool --jmx-url service:jmx:rmi:///jndi/rmi:///15500/jmxrmi
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
"time","JMIImplementation:type=MBeanServerDelegate:ImplementationName","JMIImplementation:type=MBeanServerDelegate:ImplementationVendor","JMIImplementation:type=MBeanServerDelegate:ImplementationVersion","JMIImplementation:type=MBeanServerDelegate:MBeanServerId","JMIImplementation:type=MBeanServerDelegate:SpecificationName","JMIImplementation:type=MBeanServerDelegate:SpecificationVendor","JMIImplementation:type=MBeanServerDelegate:SpecificationVersion","com.sun.management:type=HotSpotDiagnostic:DiagnosticOptions","com.sun.management:type=HotSpotDiagnostic:ObjectName","java.lang:type=ClassLoading:LoadedClassCount","java.lang:type=ClassLoading:ObjectName","java.lang:type=ClassLoading:TotalLoadedClassCount","java.lang:type=ClassLoading:UnloadedClassCount","java.lang:type=ClassLoading:Verbose","java.lang:type=Compilation:CompilationTimeMonitoringSupported","java.lang:type=Compilation:Name","java.lang:type=Compilation:ObjectName","java.lang:type=Compilation:TotalCompilationTime","java.lang:type=GarbageCollector,name=G1 Old Generation:CollectionCount","java.lang:type=GarbageCollector,name=G1 Old Generation:CollectionTime","java.lang:type=GarbageCollector,name=G1 Old Generation:LastGcInfo","java.lang:type=GarbageCollector,name=G1 Old Generation:MemoryPoolNames","java.lang:type=GarbageCollector,name=G1 Old Generation:Name","java.lang:type=GarbageCollector,name=G1 Old Generation:ObjectName","java.lang:type=GarbageCollector,name=G1 Old Generation:Valid","java.lang:type=GarbageCollector,name=G1 Young Generation:CollectionCount","java.lang:type=GarbageCollector,name=G1 Young Generation:CollectionTime","java.lang:type=GarbageCollector,name=G1 Young Generation:LastGcInfo","java.lang:type=GarbageCollector,name=G1 Young Generation:MemoryPoolNames","java.lang:type=GarbageCollector,name=G1 Young Generation:Name","java.lang:type=GarbageCollector,name=G1 Young Generation:ObjectName","java.lang:type=GarbageCollector,name=G1 Young Generation:Valid","java.lang:type=Memory:HeapMemoryUsage","java.lang:type=Memory:NonHeapMemoryUsage","java.lang:type=Memory:ObjectName","java.lang:type=Memory:ObjectPendingFinalizationCount","java.lang:type=Memory:Verbose","java.lang:type=MemoryManager,name=CodeCacheManager:MemoryPoolNames","java.lang:type=MemoryManager,name=CodeCacheManager:Name","java.lang:type=MemoryManager,name=CodeCacheManager:ObjectName","java.lang:type=MemoryManager,name=CodeCacheManager:Valid","java.lang:type=MemoryManager,name=Metaspace Manager:Memo
```

# Operating Kafka

- **Adding or removing topics**

controlled.shutdown.retry.backoff.ms	5000	🔒	✓	↺
controller.message.queue.size	10	🔒	+	↺
controller.socket.timeout.ms	30000	🔒	+	↺
default.replication.factor	1	🔒	+	↺
delete.topic.enable	true	🔒	+	↺
external.kafka.metrics.exclude.prefix	kafka.network.RequestMetrics,kafka.server.DelayedOperationPurgatory,kafka.server.Brok	🔒	+	↺
external.kafka.metrics.include.prefix	kafka.network.RequestMetrics.ResponseQueueTimeMs.request.OffsetCommit.98percenti	🔒	+	↺



# Operating Kafka

- Adding or removing topics

```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-topics.sh --create --zookeeper localhost:2181 --topic test-topic --partitions 5 --replication-factor 1
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
Created topic "test-topic".
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-topics.sh --describe --zookeeper localhost:2181 --topic test-topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
Topic:test-topic      PartitionCount:5      ReplicationFactor:1   Configs:
Topic: test-topic      Partition: 0          Leader: 1001           Replicas: 1001        Isr: 1001
Topic: test-topic      Partition: 1          Leader: 1001           Replicas: 1001        Isr: 1001
Topic: test-topic      Partition: 2          Leader: 1001           Replicas: 1001        Isr: 1001
Topic: test-topic      Partition: 3          Leader: 1001           Replicas: 1001        Isr: 1001
Topic: test-topic      Partition: 4          Leader: 1001           Replicas: 1001        Isr: 1001
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-topics.sh --delete --zookeeper localhost:2181 --topic test-topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
Topic test-topic is marked for deletion.
Note: This will have no impact if delete.topic.enable is not set to true.
[root@sandbox-hdp ~]#
```

# Operating Kafka

- **Modifying message topics**

```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-topics.sh --create --zookeeper localhost:2181 --topic test1-topic --partitions 1 --replication-factor 1
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
Created topic "test1-topic".
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-topics.sh --zookeeper localhost:2181 --alter --topic test1-topic --partitions 5 --config delete.retention.ms=10000 --delete-config retention.ms
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
WARNING: Altering topic configuration from this script has been deprecated and may be removed in future releases.
        Going forward, please use kafka-configs.sh for this functionality
Updated config for topic "test1-topic".
WARNING: If partitions are increased for a topic that has a key, the partition logic or ordering of the messages will be affected
Adding partitions succeeded!
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-configs.sh --zookeeper localhost:2181 --entity-type topics --entity-name test-topic --alter --add-config retention.ms=1000
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
Updated config for entity: topic 'test-topic'.
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-configs.sh --zookeeper localhost:2181 --entity-type topics --entity-name test-topic --alter --delete-config retention.ms
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
Updated config for entity: topic 'test-topic'.
[root@sandbox-hdp ~]#
```



# Operating Kafka

- Implementing a graceful shutdown

auto.create.topics.enable	<input type="text" value="true"/>			
auto.leader.rebalance.enable	<input type="text" value="true"/>			
compression.type	<input type="text" value="producer"/>			
controlled.shutdown.enable	<input type="text" value="true"/>			
controlled.shutdown.max.retries	<input type="text" value="3"/>			
controlled.shutdown.retry.backoff.ms	<input type="text" value="5000"/>			
controller.message.queue.size	<input type="text" value="10"/>			
controller.socket.timeout	<input type="text" value="30000"/>			

# Operating Kafka

- **Balancing leadership**

auto.create.topics.enable true

auto.leader.rebalance.  
enable true

compression.type producer

controlled.shutdown.  
enable true

controlled.shutdown.max.  
retries 3

controlled.shutdown.retry.  
backoff.ms 5000

controller.message.  
queue.size 10

controller.socket.timeout. 30000

# Operating Kafka

- **Balancing leadership**

```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-preferred-replica-election.sh --zookeeper localhost:2181
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
Successfully started preferred replica election for partitions Set([__consumer_offsets,32], [test1-topic,2], [__consumer_offsets,16], [partition-topic,1], [__consumer_offsets,49], [__consumer_offsets,44], [test-topic,4], [__consumer_offsets,28], [ATLAS_ENTITIES,0], [test-topic,1], [ATLAS_HOOK,0], [__consumer_offsets,17], [__consumer_offsets,23], [__consumer_offsets,7], [test-topic,0], [class9Topic,0], [__consumer_offsets,4], [partition-topic,2], [good-topic,0], [__consumer_offsets,29], [mysql-jdbc-demo_products,0], [__consumer_offsets,35], [schemas,0], [__consumer_offsets,3], [__consumer_offsets,24], [__consumer_offsets,41], [__consumer_offsets,0], [demoTopic,0], [__consumer_offsets,38], [mysql-jdbc-demo_products1,0], [__consumer_offsets,13], [__consumer_offsets,8], [__consumer_offsets,5], [__consumer_offsets,39], [partition-topic,0], [__consumer_offsets,36], [__consumer_offsets,40], [confluent.support.metrics,0], [__consumer_offsets,45], [__consumer_offsets,15], [__consumer_offsets,33], [source-topic,0], [__consumer_offsets,37], [__consumer_offsets,21], [__consumer_offsets,6], [test-topic,2], [test-topic,3], [test1-topic,0], [__consumer_offsets,11], [__consumer_offsets,20], [__consumer_offsets,47], [__consumer_offsets,2], [__consumer_offsets,27], [__consumer_offsets,34], [__consumer_offsets,9], [__consumer_offsets,22], [test1-topic,4], [__consumer_offsets,42], [test1-topic,3], [__consumer_offsets,14], [__consumer_offsets,25], [__consumer_offsets,10], [__consumer_offsets,48], [__consumer_offsets,31], [__consumer_offsets,18], [__consumer_offsets,19], [bad-topic,0], [__consumer_offsets,12], [test1-topic,1], [__consumer_offsets,46], [__consumer_offsets,43], [__consumer_offsets,1], [nifi_websocket,0], [__consumer_offsets,26], [replicatedClass9Topic,0], [__consumer_offsets,30])
[root@sandbox-hdp ~]#
```

# Operating Kafka

- Checking the consumer position

```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-consumer-groups.sh --bootstrap-server sandbox-hdp.hortonworks.com:6667 --describe --group consumerGroup
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
GROUP          TOPIC          PARTITION  CURRENT-OFFSET  LOG-END-OFFSET  LAG          OWNER
consumerGroup  class9Topic    0          9               9               0            consumer-1_172.17.0.2
[root@sandbox-hdp ~]#
```

# Operating Kafka

- Zookeeper Tips

```
2018-01-20 05:47:00,559 - INFO [main:Environment@100] - Client environment:java.compiler=<NA>
2018-01-20 05:47:00,559 - INFO [main:Environment@100] - Client environment:os.name=Linux
2018-01-20 05:47:00,559 - INFO [main:Environment@100] - Client environment:os.arch=amd64
2018-01-20 05:47:00,559 - INFO [main:Environment@100] - Client environment:os.version=4.14.0-1.el7.elrepo.x86_64
2018-01-20 05:47:00,559 - INFO [main:Environment@100] - Client environment:user.name=root
2018-01-20 05:47:00,559 - INFO [main:Environment@100] - Client environment:user.home=/root
2018-01-20 05:47:00,559 - INFO [main:Environment@100] - Client environment:user.dir=/root
2018-01-20 05:47:00,561 - INFO [main:ZooKeeper@438] - Initiating client connection, connectString=localhost:2181 sessionTimeout=30000 watcher=org.apache.zookeeper.ZooKeeperMain$MyWatcher@4534b60d
Welcome to ZooKeeper!
2018-01-20 05:47:00,766 - INFO [main-SendThread(localhost:2181):ClientCnxn$SendThread@1019] - Opening socket connection to server localhost/127.0.0.1:2181. Will not attempt to authenticate using SASL (unknown error)
JLine support is enabled
2018-01-20 05:47:01,252 - INFO [main-SendThread(localhost:2181):ClientCnxn$SendThread@864] - Socket connection established, initiating session, client: /127.0.0.1:38944, server: localhost/127.0.0.1:2181
2018-01-20 05:47:01,343 - INFO [main-SendThread(localhost:2181):ClientCnxn$SendThread@1279] - Session establishment complete on server localhost/127.0.0.1:2181, sessionId = 0x161115ef3980019, negotiated timeout = 30000

WATCHER::

WatchedEvent state:SyncConnected type:None path:null
[zk: localhost:2181(CONNECTED) 0] ls /
[schema_registry, registry, cluster, controller, storm, brokers, zookeeper, infra-solr, hbase-unsecure, admin, isr_change_notification, log_dir_event_notification, hiveserver2, controller_epoch, consumers, latest_producer_id_block, config, kylin]
[zk: localhost:2181(CONNECTED) 1] get /consumers
null
cZxid = 0x25f
ctime = Fri Nov 10 14:58:37 UTC 2017
mZxid = 0x25f
mtime = Fri Nov 10 14:58:37 UTC 2017
pZxid = 0x111a
cversion = 14
dataVersion = 0
aclVersion = 0
ephemeralOwner = 0x0
dataLength = 0
numChildren = 4
[zk: localhost:2181(CONNECTED) 2] █
```

# Operating Kafka

- Config

```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-configs.sh --zookeeper localhost:2181 --describe --entity-type topics --entity-name class9Topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
Configs for topic 'class9Topic' are
[root@sandbox-hdp ~]#
```



# Operating Kafka

- Performance

```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-producer-perf-test.sh --topic class9Topic --broker-list sandbox-hdp.hortonworks.com:6667 --messages 20000
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
start.time, end.time, compression, message.size, batch.size, total.data.sent.in.MB, MB.sec, total.data.sent.in.nMsg, nMsg.sec
2018-01-20 06:15:51:400, 2018-01-20 06:16:03:088, 0, 100, 200, 1.91, 0.1632, 20000, 1711.1567
[root@sandbox-hdp ~]#
```



# Operating Kafka

- ACLs

```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-acls.sh --authorizer-properties zookeeper.connect=localhost:2181 --add --allow
-principal User:elastic --producer --topic class9Topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of p
arallel GC threads appropriately using -XX:ParallelGCThreads=N
Adding ACLs for resource `Topic:class9Topic`:
    User:elastic has Allow permission for operations: Write from hosts: *
    User:elastic has Allow permission for operations: Describe from hosts: *

Adding ACLs for resource `Cluster:kafka-cluster`:
    User:elastic has Allow permission for operations: Create from hosts: *

Current ACLs for resource `Topic:class9Topic`:
    User:elastic has Allow permission for operations: Write from hosts: *
    User:elastic has Allow permission for operations: Describe from hosts: *

[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-acls.sh --authorizer-properties zookeeper.connect=localhost:2181 --list --topi
c class9Topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of p
arallel GC threads appropriately using -XX:ParallelGCThreads=N
Current ACLs for resource `Topic:class9Topic`:
    User:elastic has Allow permission for operations: Write from hosts: *
    User:elastic has Allow permission for operations: Describe from hosts: *

[root@sandbox-hdp ~]#
```

# Monitoring

- Monitoring server statistics

localhost:8080/#/main/services/KAFKA/configs

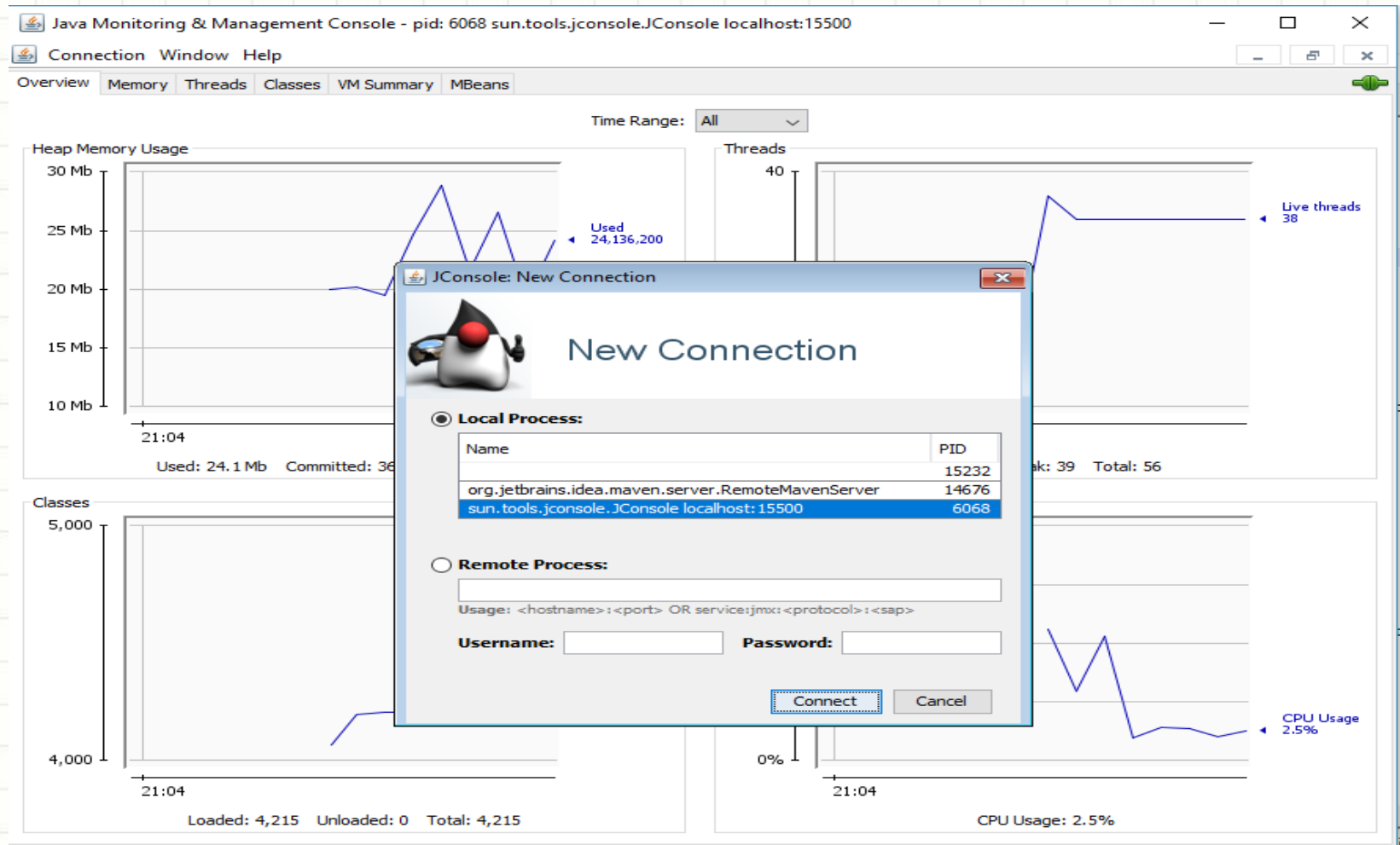
is\_supported\_kafka true

V4 ✓ admin authored on Fri, Nov 10, 2017 09:58 Discard Save

kafka_log_dir	/var/log/kafka	+	C
Kafka PID dir	/var/run/kafka		
kafka_user_nofile_limit	128000	+	C
kafka_user_nproc_limit	65536	+	C
kafka-env template	<pre>#!/bin/bash  # Set KAFKA specific environment variables here.  # The java implementation to use. export JAVA_HOME={{java64_home}} export PATH=\$PATH:\$JAVA_HOME/bin export PID_DIR={{kafka_pid_dir}} export LOG_DIR={{kafka_log_dir}} export KAFKA_KERBEROS_PARAMS={{kafka_kerberos_params}} JMX_PORT=15500 # Add kafka sink to classpath and related dependencies if [ -e "/usr/lib/ambari-metrics-kafka-sink/ambari-metrics-kafka-sink.jar" ]; then   export CLASSPATH=\$CLASSPATH:/usr/lib/ambari-metrics-kafka-sink/ambari-metrics-kafka-sink.jar   export CLASSPATH=\$CLASSPATH:/usr/lib/ambari-metrics-kafka-sink/lib/* fi if [ -f /etc/kafka/conf/kafka-ranger-env.sh ]; then   . /etc/kafka/conf/kafka-ranger-env.sh</pre>		

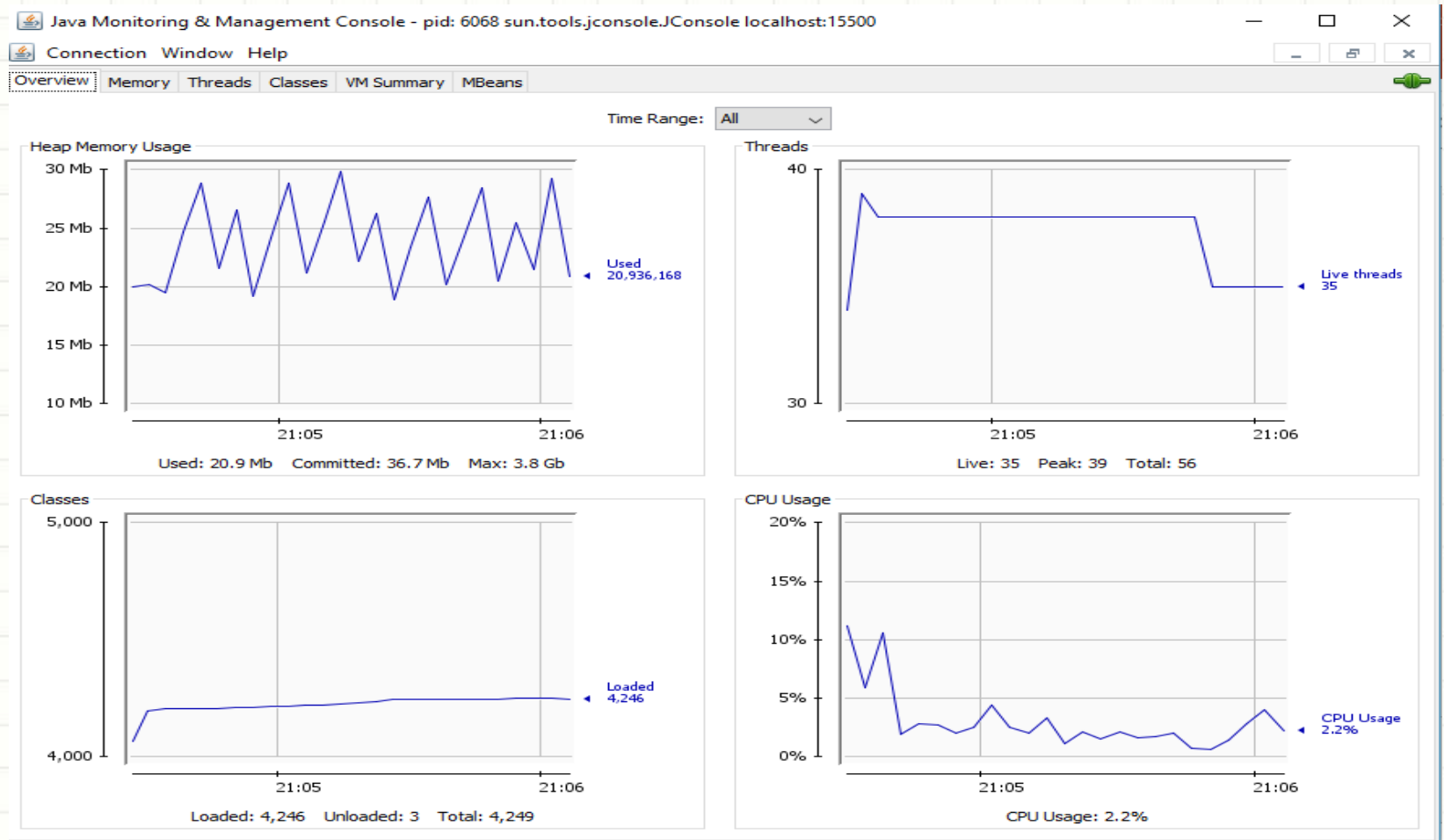
# Monitoring

- Monitoring server statistics



# Monitoring

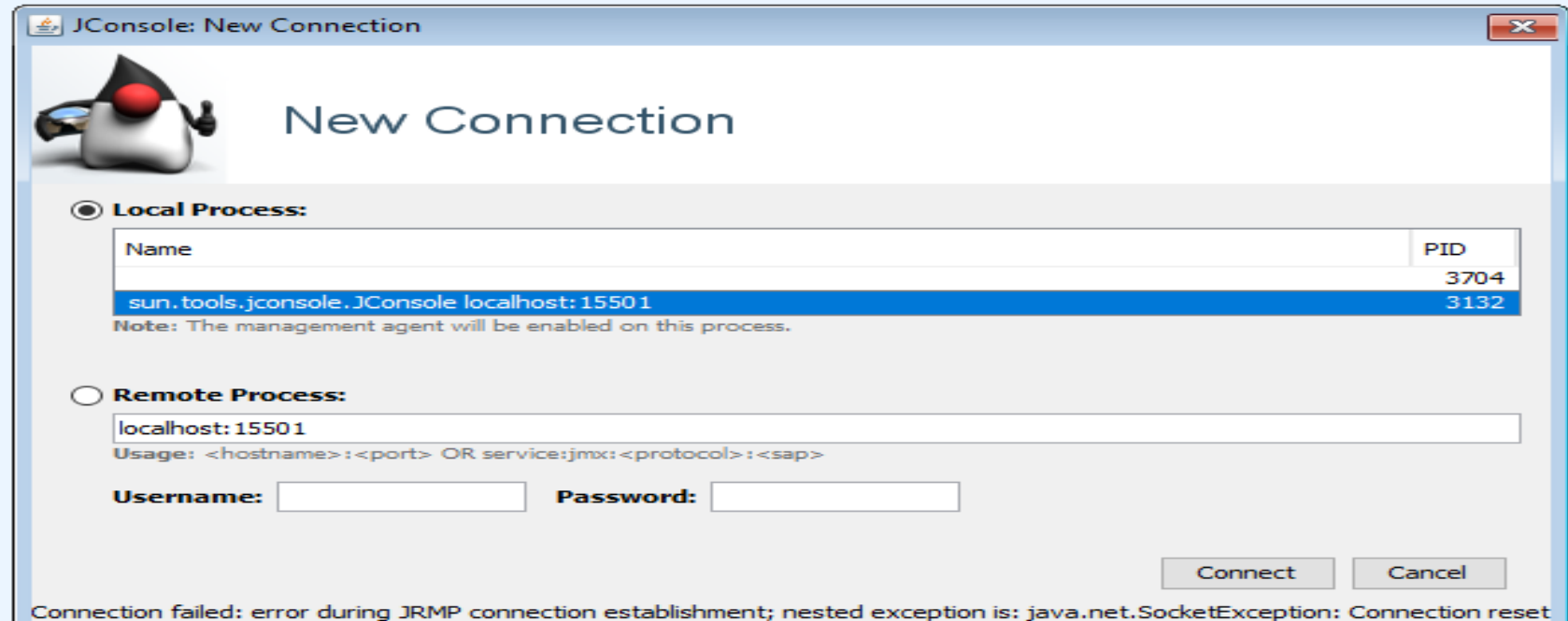
- Monitoring server statistics



# Monitoring

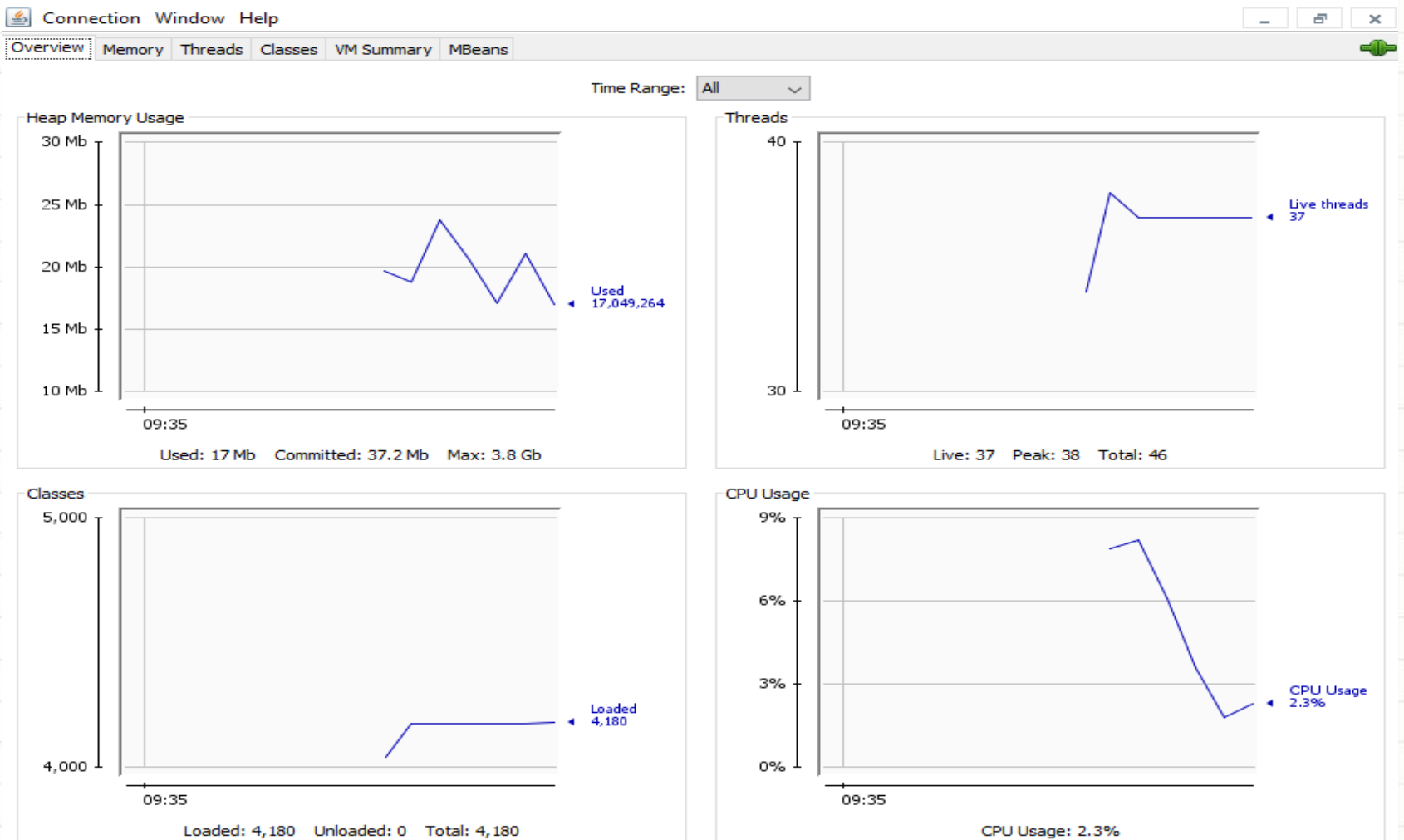
- Monitoring producer statistics

```
[root@sandbox-hdp ~]# JMX_PORT=15501 /usr/hdp/2.6.3.0-235/kafka/bin/kafka-console-producer.sh --broker-list sandbox-hdp.hortonworks.com:6667 --topic test_topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
```



# Monitoring

- Monitoring producer statistics

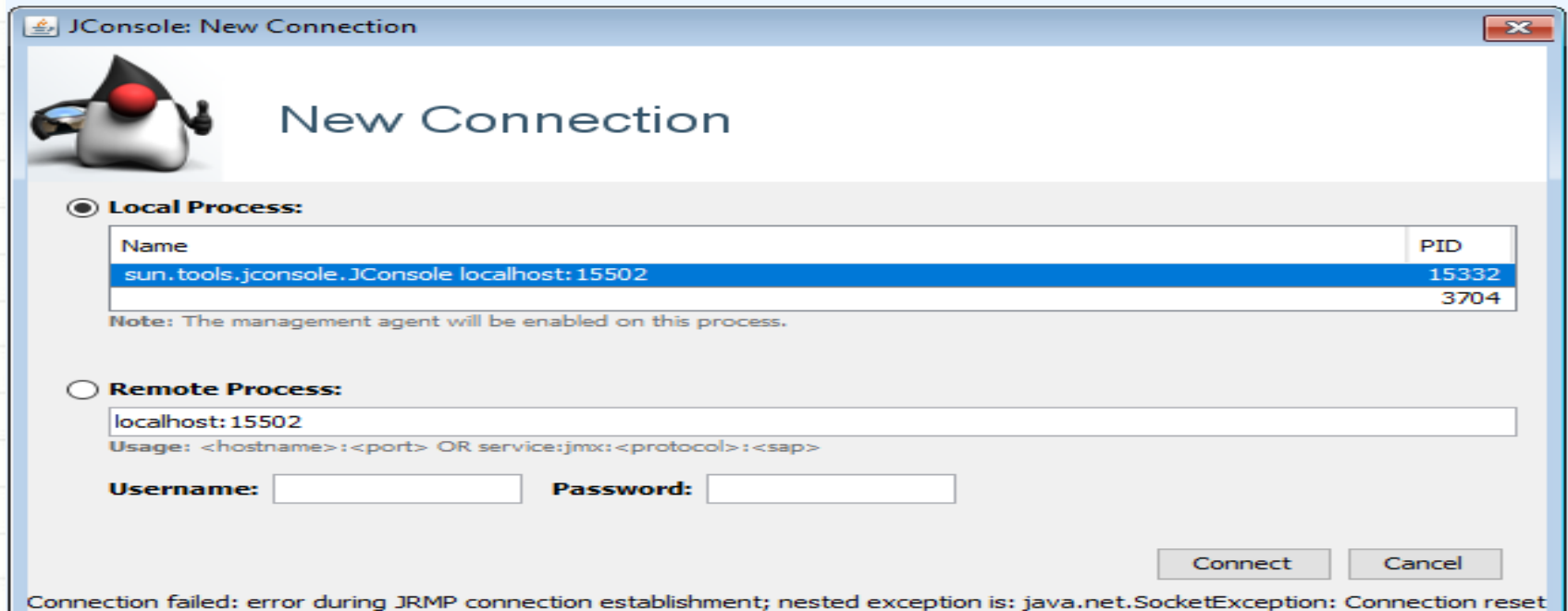




# Monitoring

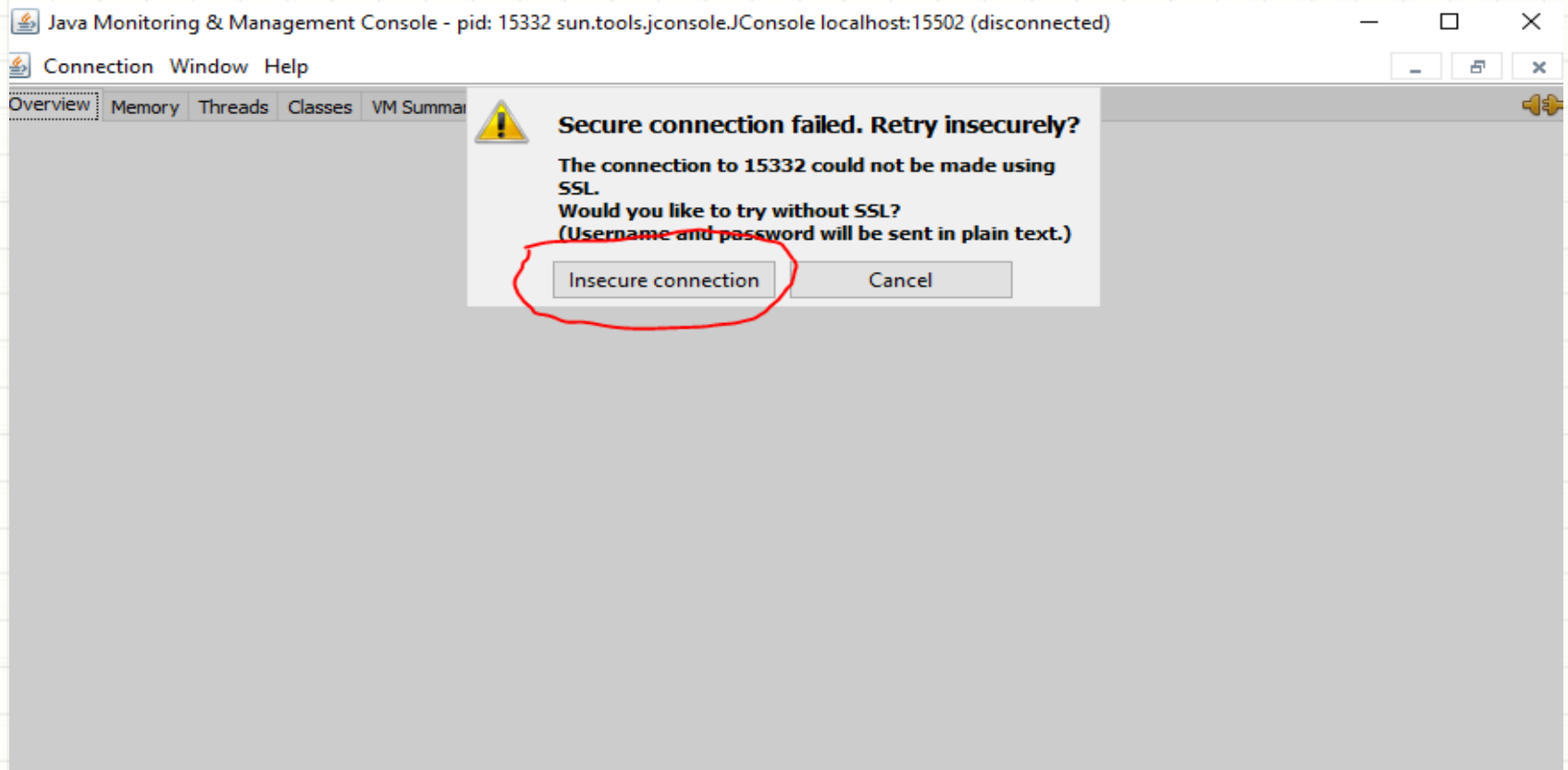
- Monitoring consumer statistics

```
[root@sandbox-hdp ~]# JMX_PORT=15502 /usr/hdp/2.6.3.0-235/kafka/bin/kafka-console-consumer.sh --bootstrap-server sandbox-hdp.hortonworks.com:6667 --from-beginning --topic test_topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
```



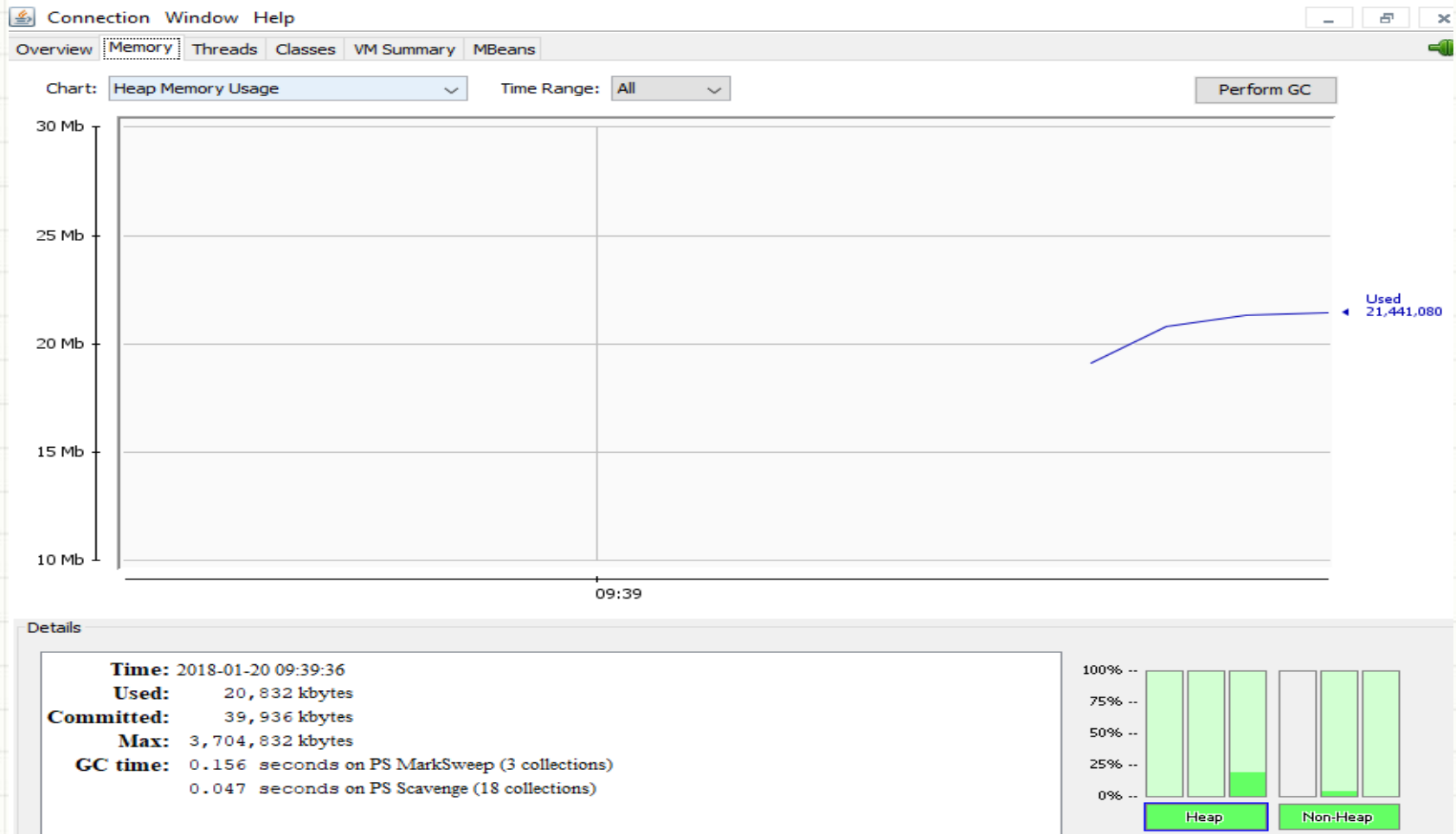
# Monitoring

- **Monitoring consumer statistics**



# Monitoring

- Monitoring consumer statistics



# Creating Custom Partition for Producer

```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-topics.sh --create --zookeeper localhost:2181 --replication-factor 1 --partitions 3 --topic partition-topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
Created topic "partition-topic".
[root@sandbox-hdp ~]# java -cp /root/TrainingOnHDP/StreamingApplicationOnKafka/target/StreamingApplicationOnKafka-1.0-SNAPSHOT-jar-with-dependencies.jar ca.training.bigdata.kafka.partition.ProcessingApp
log4j:WARN No appenders could be found for logger (org.apache.kafka.clients.producer.ProducerConfig).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.
Sent:Hello Microsoft, Key: MSFT, Partition: 1
Receive message: Hello Microsoft, Partition: 1, Offset: 0
Receive message: Hello Google, Partition: 2, Offset: 0
Sent:Hello Google, Key: GOOGL, Partition: 2
Sent:Hello Apple, Key: APPL, Partition: 0
Receive message: Hello Apple, Partition: 0, Offset: 0
```

# Creating Kafka Producer Example

← → ↻ localhost:4200



```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-topics.sh --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic demoTopic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
Created topic "demoTopic".
[root@sandbox-hdp ~]# java -cp /root/TrainingOnHDP/StreamingApplicationOnKafka/target/StreamingApplicationOnKafka-1.0-SNAPSHOT-jar-with-dependencies.jar ca.training.bigdata.kafka.demo.DemoProducer
log4j:WARN No appenders could be found for logger (org.apache.kafka.clients.producer.ProducerConfig).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.
[root@sandbox-hdp ~]#
```

# Creating Kafka Consumer Example

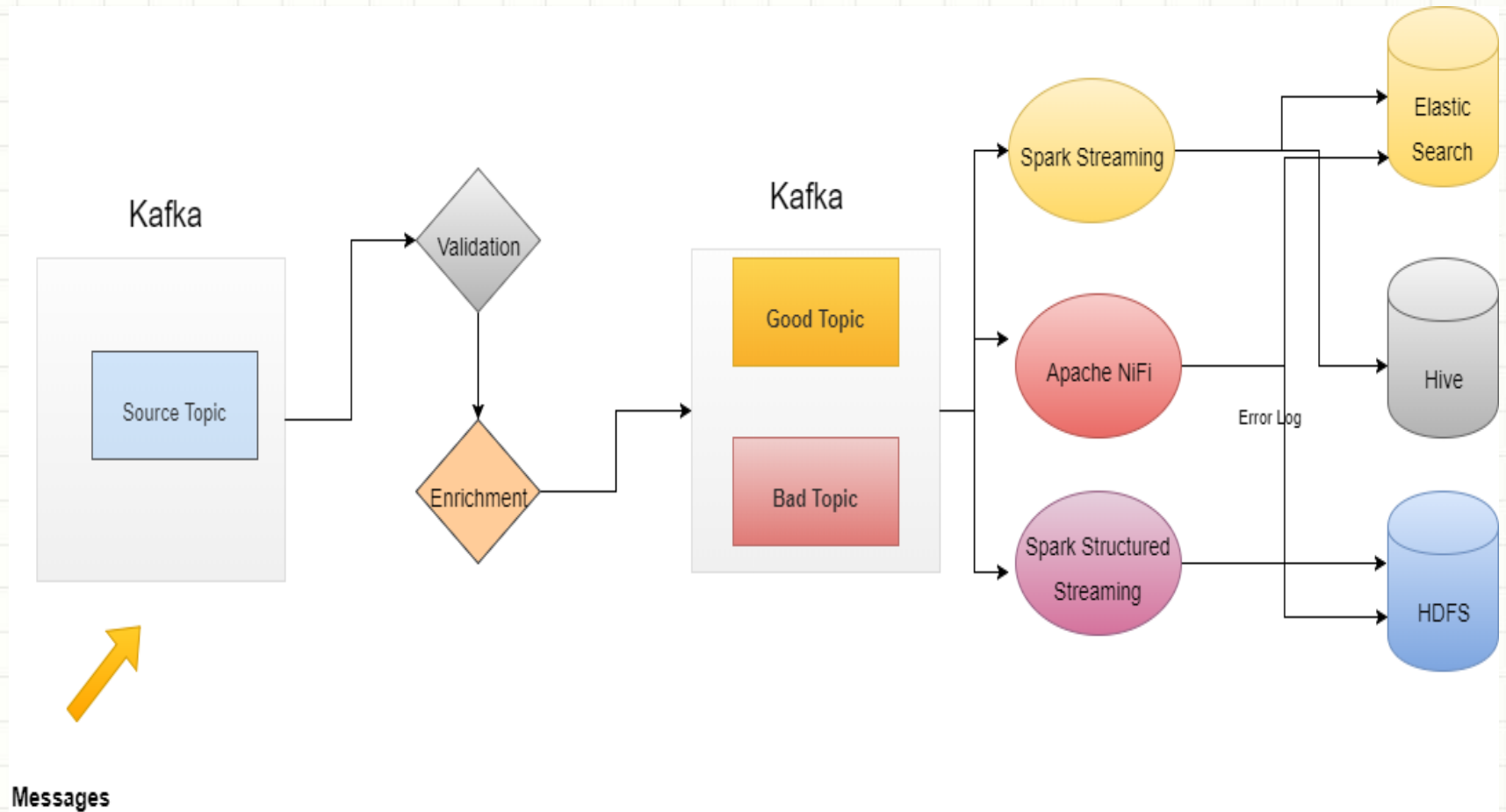
← → ↻ ⓘ localhost:4200



```
[root@sandbox-hdp ~]# java -cp /root/TrainingOnHDP/StreamingApplicationOnKafka/target/StreamingApplicationOnKafka-1.0-SNAPSHOT-jar-with-dependencies.jar ca.training.bigdata.kafka.demo.DemoConsumer
log4j:WARN No appenders could be found for logger (org.apache.kafka.clients.consumer.ConsumerConfig).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.
Subscribed to topic demoTopic
offset = 0key =nullvalue =Hello this is record1
Hello this is record1offset = 1key =nullvalue =Hello this is record2
Hello this is record2offset = 2key =nullvalue =Hello this is record3
Hello this is record3offset = 3key =nullvalue =Hello this is record4
Hello this is record4offset = 4key =nullvalue =Hello this is record5
Hello this is record5offset = 5key =nullvalue =Hello this is record6
Hello this is record6offset = 6key =nullvalue =Hello this is record7
Hello this is record7offset = 7key =nullvalue =Hello this is record8
Hello this is record8offset = 8key =nullvalue =Hello this is record9
Hello this is record9offset = 9key =nullvalue =Hello this is record10
Hello this is record10offset = 10key =nullvalue =Hello this is record11
```



# Building Spark Streaming Applications with Kafka



# Building Spark Streaming Applications with Kafka

- **Start Elasticsearch**

```
[elastic@sandbox-hdp root]$ /root/TrainingOnHDP/elasticsearch-6.1.1/bin/elasticsearch
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
[2018-01-24T02:28:16,860][INFO ][o.e.n.Node               ] [] initializing ...
[2018-01-24T02:28:17,584][INFO ][o.e.e.NodeEnvironment ] [Z9s05Kg] using [1] data paths, mounts [[/ (overlay)], net usable_space [96.2gb], net total_space [142.2gb], types [overlay]
[2018-01-24T02:28:17,585][INFO ][o.e.e.NodeEnvironment ] [Z9s05Kg] heap size [1015.6mb], compressed ordinary object pointers [true]
[2018-01-24T02:28:17,682][INFO ][o.e.n.Node               ] [Z9s05Kg] node name [Z9s05Kg] derived from node ID [Z9s05KgkSoi6nSzC7pZ_9w]; set [node.name] to override
[2018-01-24T02:28:17,682][INFO ][o.e.n.Node               ] version[6.1.1], pid[16887], build[bd92e7f/2017-12-17T20:23:25.338Z], OS[Linux/4.14.0-1.el7.elrepo.x86_64/amd64], JVM[Oracle Corporation/OpenJDK 64-Bit Server VM/1.8.0_151/25.151-b12]
[2018-01-24T02:28:17,682][INFO ][o.e.n.Node               ] JVM arguments [-Xms1g, -Xmx1g, -XX:+UseConcMarkSweepGC, -XX:CMSInitiatingOccupancyFraction=75, -XX:+UseCMSInitiatingOccupancyOnly, -XX:+AlwaysPreTouch, -Xss1m, -Djava.awt.headless=true, -Dfile.encoding=UTF-8, -Djna.nosys=true, -XX:-OmitStackTraceInFastThrow, -Dio.netty.noUnsafe=true, -Dio.netty.noKeySetOptimization=true, -Dio.netty.recycler.maxCapacityPerThread=0, -Dlog4j.shutdownHookEnabled=false, -Dlog4j2.disable.jmx=true, -XX:+HeapDumpOnOutOfMemoryError, -Des.path.home=/root/TrainingOnHDP/elasticsearch-6.1.1, -Des.path.conf=/root/TrainingOnHDP/elasticsearch-6.1.1/config]
[2018-01-24T02:28:24,174][INFO ][o.e.p.PluginsService     ] [Z9s05Kg] loaded module [aggs-matrix-stats]
[2018-01-24T02:28:24,174][INFO ][o.e.p.PluginsService     ] [Z9s05Kg] loaded module [analysis-common]
[2018-01-24T02:28:24,175][INFO ][o.e.p.PluginsService     ] [Z9s05Kg] loaded module [ingest-common]
[2018-01-24T02:28:24,175][INFO ][o.e.p.PluginsService     ] [Z9s05Kg] loaded module [lang-expression]
[2018-01-24T02:28:24,175][INFO ][o.e.p.PluginsService     ] [Z9s05Kg] loaded module [lang-mustache]
[2018-01-24T02:28:24,175][INFO ][o.e.p.PluginsService     ] [Z9s05Kg] loaded module [lang-painless]
[2018-01-24T02:28:24,175][INFO ][o.e.p.PluginsService     ] [Z9s05Kg] loaded module [mapper-extras]
```

# Building Spark Streaming Applications with Kafka

- Start Kibana

```
[root@sandbox-hdp ~]# /root/TrainingOnHDP/kibana-6.1.1-linux-x86_64/bin/kibana
log [02:33:28.481] [info][status][plugin:kibana@6.1.1] Status changed from uninitialized to green - Ready
log [02:33:28.617] [info][status][plugin:elasticsearch@6.1.1] Status changed from uninitialized to yellow - Waiting for Elasticsearch
log [02:33:28.694] [info][status][plugin:console@6.1.1] Status changed from uninitialized to green - Ready
log [02:33:28.819] [info][status][plugin:metrics@6.1.1] Status changed from uninitialized to green - Ready
log [02:33:31.903] [error][status][plugin:elasticsearch@6.1.1] Status changed from yellow to red - Request Timeout after 3000ms
log [02:33:31.904] [info][status][plugin:timelion@6.1.1] Status changed from uninitialized to green - Ready
log [02:33:31.913] [info][listening] Server running at http://0.0.0.0:8744
log [02:33:34.785] [info][status][plugin:elasticsearch@6.1.1] Status changed from red to green - Ready
```

# Building Spark Streaming Applications with Kafka

- **Kibana Admin Console**

The screenshot displays the Kibana Admin Console interface. The top navigation bar includes the Kibana logo and a sidebar with links to Discover, Visualize, Dashboard, Timelion, Dev Tools, and Management. The main content area shows search results for the index pattern 'stocks-index\*'. The search bar at the top contains the query 'search... (e.g. status:200 AND extension:PHP)' and indicates 'Uses lucene query syntax'. The results section shows 1,047 hits. The first three results are displayed as JSON objects, each containing fields like volume, symbol, price, ts, \_id, \_type, \_index, \_score, and ds.

Volume	Symbol	Price	TS	ID	Type	Index	Score	DS
20,999,961	AAPL	170.53	December 26th 2017, 07:38:18.000	oOy8kmABnwN1NivTSXD7	sto	cks	1	2017-12-26
4,208,681	MSFT	85.32	December 26th 2017, 07:38:47.000	oey8kmABnwN1NivTSXD7	stock	s	1	2017-12-26
21,003,767	AAPL	170.53	December 26th 2017, 07:38:35.000	ouy8kmABnwN1NivTSXD7	sto	cks	1	2017-12-26

# Building Spark Streaming Applications with Kafka

- Start Hive and Kafka Service

localhost:8080/#/main/services/KAFKA/summary

Ambari Sandbox 1 op 0 alerts Dashboard Services Hosts Alerts Admin admin

## 1 Background Operation Running

Operations	Start Time	Duration	Show: All (10)
⚙ Start Kafka ⓧ	Today 21:55	892 ms	0%
✓ Start Hive	Today 21:37	542.55 secs	100%
✓ Stop NiFi	Today 21:22	42.89 secs	100%
✓ Start MapReduce2	Today 21:08	119.05 secs	100%
✓ Start YARN	Today 20:47	496.15 secs	100%
— Start All Services	Today 19:46	21.01 mins	100%
! Start HDFS	Today 19:12	34.08 mins	100%
— Start All Services	Mon Jan 22 2018 21:41	21.00 mins	100%
! Start HDFS	Mon Jan 22 2018 21:07	34.01 mins	100%

☐ Do not show this dialog again when starting a background operation OK

Left sidebar: HDFS, YARN, MapReduce2, Tez, Hive, HBase, Pig, Sqoop, Oozie, ZooKeeper, Falcon, Storm, Flume, Ambari Infra, Atlas, Kafka, Knox, Ranger, Spark2

Right sidebar: Service Action, No a, Last 1 hou, Replica Manager, No Data Available

# Building Spark Streaming Applications with Kafka

- **Create Hive Database**

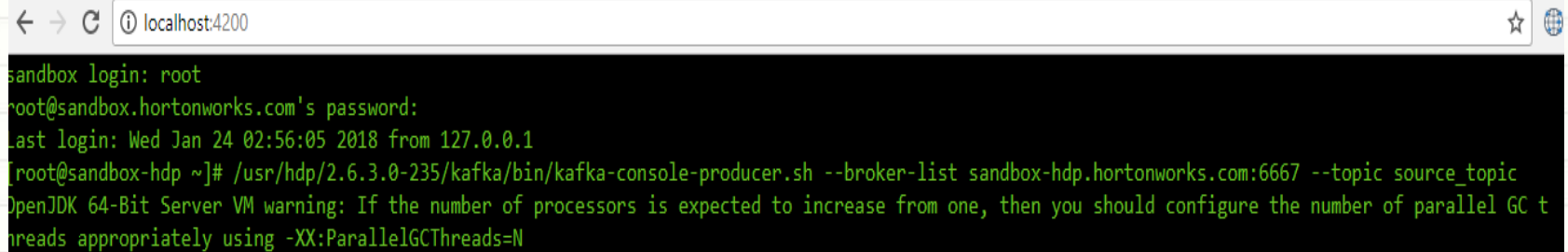
```
[root@sandbox-hdp ~]# hive
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
log4j:WARN No such property [maxFileSize] in org.apache.log4j.DailyRollingFileAppender.

Logging initialized using configuration in file:/etc/hive/2.6.3.0-235/0/hive-log4j.properties
hive> create database kafka;
OK
Time taken: 22.299 seconds
hive> show databases;
OK
default
foodmart
foodmart_spark
hbase
kafka
oozie
xademo
Time taken: 0.881 seconds, Fetched: 7 row(s)
hive> 
```



# Building Spark Streaming Applications with Kafka

- **Start Producer Console**

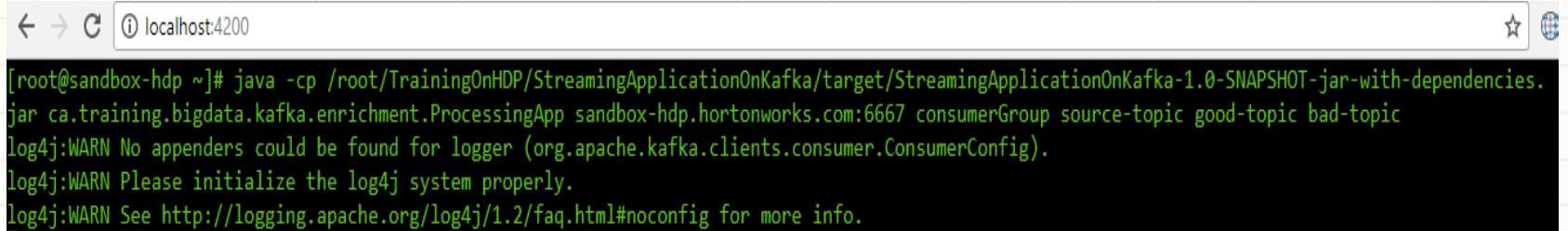


The screenshot shows a terminal window with the following text:

```
sandbox login: root
root@sandbox.hortonworks.com's password:
Last login: Wed Jan 24 02:56:05 2018 from 127.0.0.1
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-console-producer.sh --broker-list sandbox-hdp.hortonworks.com:6667 --topic source_topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
```

# Building Spark Streaming Applications with Kafka

- **Start up the processing application**



A terminal window with a black background and green text. The window title bar shows navigation icons and the address 'localhost:4200'. The terminal content shows a Java command being executed, followed by log4j warning messages about missing appenders and log4j initialization.

```
[root@sandbox-hdp ~]# java -cp /root/TrainingOnHDP/StreamingApplicationOnKafka/target/StreamingApplicationOnKafka-1.0-SNAPSHOT-jar-with-dependencies.jar ca.training.bigdata.kafka.enrichment.ProcessingApp sandbox-hdp.hortonworks.com:6667 consumerGroup source-topic good-topic bad-topic
log4j:WARN No appenders could be found for logger (org.apache.kafka.clients.consumer.ConsumerConfig).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.
```

# Building Spark Streaming Applications with Kafka

- Start up the streaming process application

```
ssl.keystore.type = JKS
ssl.protocol = TLS
ssl.provider = null
ssl.secure.random.implementation = null
ssl.trustmanager.algorithm = PKIX
ssl.truststore.location = null
ssl.truststore.password = null
ssl.truststore.type = JKS
value.deserializer = class org.apache.kafka.common.serialization.StringDeserializer
```

```
18/01/24 03:25:31 INFO AppInfoParser: Kafka version : 0.10.1.1
18/01/24 03:25:31 INFO AppInfoParser: Kafka commitId : f10ef2720b03b247
18/01/24 03:25:32 INFO AbstractCoordinator: Discovered coordinator sandbox-hdp.hortonworks.com:6667 (id: 2147482646 rack: null) for group streamingGroup.
18/01/24 03:25:32 INFO ConsumerCoordinator: Revoking previously assigned partitions [] for group streamingGroup
18/01/24 03:25:32 INFO AbstractCoordinator: (Re-)joining group streamingGroup
18/01/24 03:25:32 INFO AbstractCoordinator: Successfully joined group streamingGroup with generation 1
18/01/24 03:25:32 INFO ConsumerCoordinator: Setting newly assigned partitions [good-topic-0] for group streamingGroup
18/01/24 03:25:32 INFO RecurringTimer: Started timer for JobGenerator at time 1516764360000
18/01/24 03:25:32 INFO JobGenerator: Started JobGenerator at 1516764360000 ms
18/01/24 03:25:32 INFO JobScheduler: Started JobScheduler
18/01/24 03:25:32 INFO ContextHandler: Started o.s.j.s.ServletContextHandler@3e7b65d7{/streaming,null,AVAILABLE,@Spark}
18/01/24 03:25:32 INFO ContextHandler: Started o.s.j.s.ServletContextHandler@3ddeaa5f{/streaming/json,null,AVAILABLE,@Spark}
18/01/24 03:25:32 INFO ContextHandler: Started o.s.j.s.ServletContextHandler@3bec5821{/streaming/batch,null,AVAILABLE,@Spark}
18/01/24 03:25:32 INFO ContextHandler: Started o.s.j.s.ServletContextHandler@66236a0a{/streaming/batch/json,null,AVAILABLE,@Spark}
18/01/24 03:25:32 INFO ContextHandler: Started o.s.j.s.ServletContextHandler@7c781c42{/static/streaming,null,AVAILABLE,@Spark}
18/01/24 03:25:32 INFO StreamingContext: StreamingContext started
```

# Building Spark Streaming Applications with Kafka

- **Publish the messages**

```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-console-producer.sh --broker-list sandbox-hdp.hortonworks.com:6667 --topic source_topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
{"event": "CUSTOMER_SEES_BTCPRI", "customer": {"id": "86689427", "name": "Edward S.", "ipAddress": "95.31.18.119"}, "currency": {"name": "bitcoin", "price": "USD"}, "timestamp": "2017-07-03T12:00:35Z"}
{"event": "CUSTOMER_SEES_BTCPRI", "customer": {"id": "18313440", "name": "Julian A.", "ipAddress": "185.86.151.11"}, "currency": {"name": "bitcoin", "price": "USD"}, "timestamp": "2017-07-04T15:00:35Z"}
{"event": "CUSTOMER_SEES_BTCPRI", "customer": {"id": "56886468", "name": "Lindsay M.", "ipAddress": "186.46.129.15"}, "currency": {"name": "bitcoin", "price": "USD"}, "timestamp": "2017-07-11T19:00:35Z"}
```

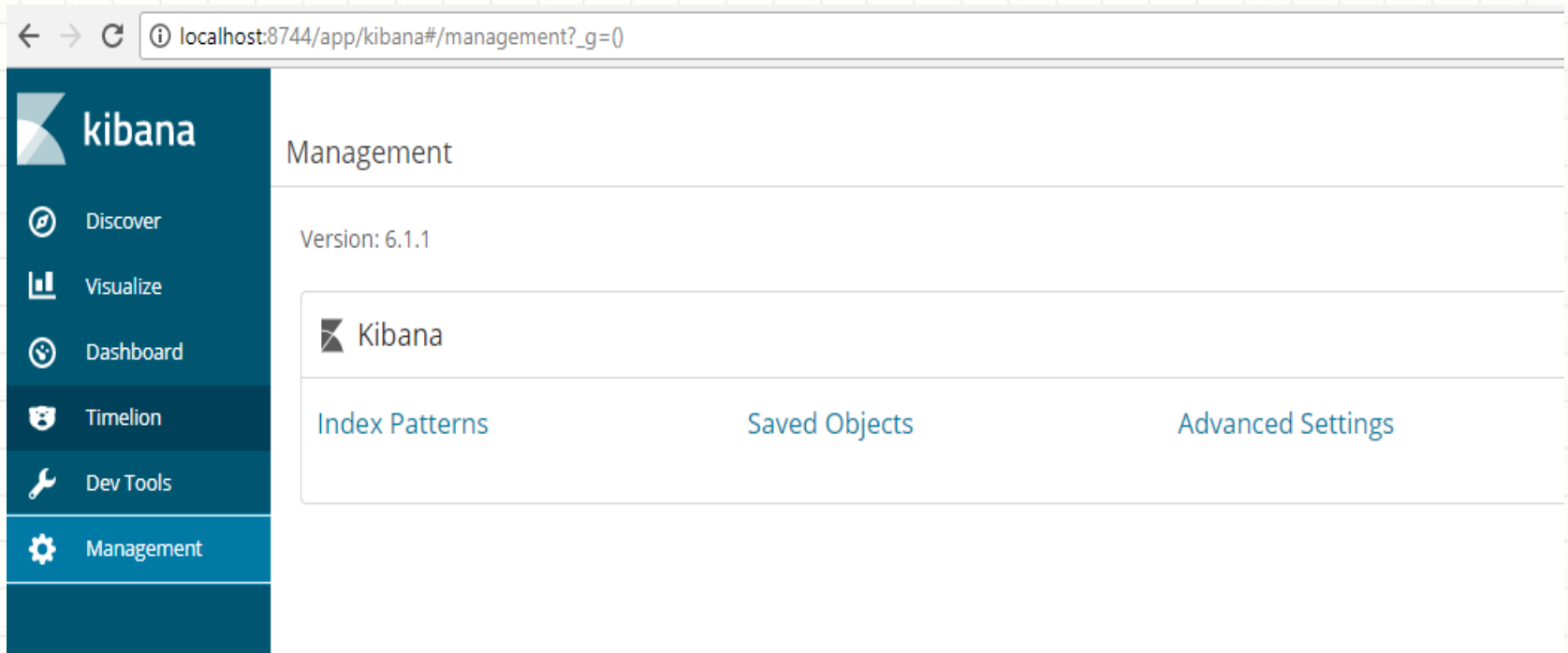
# Building Spark Streaming Applications with Kafka

- Run hive query to check

```
hive> select * from kafka.customer_bitcoin;
OK
{"name":"bitcoin","price":"USD","rate":9.3300451E-5} {"city":"London","country":"United Kingdom","id":"18313440","ipAddress":"185.86.1
51.11","name":"Julian A."} CUSTOMER_SEES_BTCPPRICE 2017-07-04T15:00:35Z
{"name":"bitcoin","price":"USD","rate":9.3300451E-5} {"city":"Quito","country":"Ecuador","id":"56886468","ipAddress":"186.46.129.15",
name":"Lindsay M."} CUSTOMER_SEES_BTCPPRICE 2017-07-11T19:00:35Z
{"name":"bitcoin","price":"USD","rate":9.3300451E-5} {"city":"Moscow","country":"Russian Federation","id":"86689427","ipAddress":"95.3
1.18.119","name":"Edward S."} CUSTOMER_SEES_BTCPPRICE 2017-07-03T12:00:35Z
{"name":"bitcoin","price":"USD","rate":9.3300451E-5} {"city":"Moscow","country":"Russian Federation","id":"86689427","ipAddress":"95.3
1.18.119","name":"Edward S."} CUSTOMER_SEES_BTCPPRICE 2017-07-03T12:00:35Z
{"name":"bitcoin","price":"USD","rate":9.3300451E-5} {"city":"Quito","country":"Ecuador","id":"56886468","ipAddress":"186.46.129.15",
name":"Lindsay M."} CUSTOMER_SEES_BTCPPRICE 2017-07-11T19:00:35Z
Time taken: 0.442 seconds, Fetched: 5 row(s)
```

# Building Spark Streaming Applications with Kafka

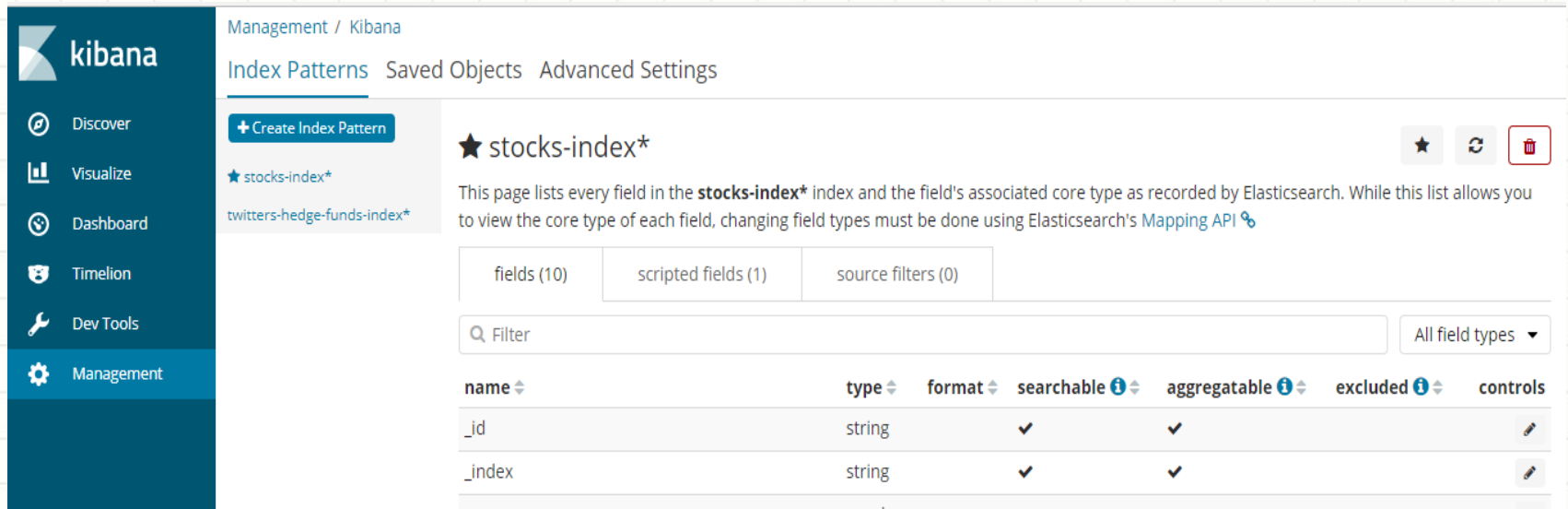
- **Kibana console**





# Building Spark Streaming Applications with Kafka

- **Kibana console**



The screenshot displays the Kibana Management console interface. On the left is a dark blue sidebar with the Kibana logo and navigation links: Discover, Visualize, Dashboard, Timelion, Dev Tools, and Management (which is highlighted). The main content area has a light blue header with 'Management / Kibana' and tabs for 'Index Patterns', 'Saved Objects', and 'Advanced Settings'. Below the tabs is a '+ Create Index Pattern' button and a list of index patterns, with 'stocks-index\*' selected. The main view shows details for 'stocks-index\*', including a star icon, a refresh icon, and a delete icon. A descriptive text explains that the page lists fields in the index and their core types as recorded by Elasticsearch, noting that field types must be changed using the Mapping API. Below this text are three tabs: 'fields (10)', 'scripted fields (1)', and 'source filters (0)'. A search filter input is present. A table lists the fields with columns for name, type, format, searchable, aggregatable, excluded, and controls. The visible rows are for '\_id' (string, searchable, aggregatable) and '\_index' (string, searchable, aggregatable).

Management / Kibana

Index Patterns Saved Objects Advanced Settings

+ Create Index Pattern

★ stocks-index\*

twitters-hedge-funds-index\*

★ stocks-index\*

This page lists every field in the **stocks-index\*** index and the field's associated core type as recorded by Elasticsearch. While this list allows you to view the core type of each field, changing field types must be done using Elasticsearch's [Mapping API](#).

fields (10) scripted fields (1) source filters (0)

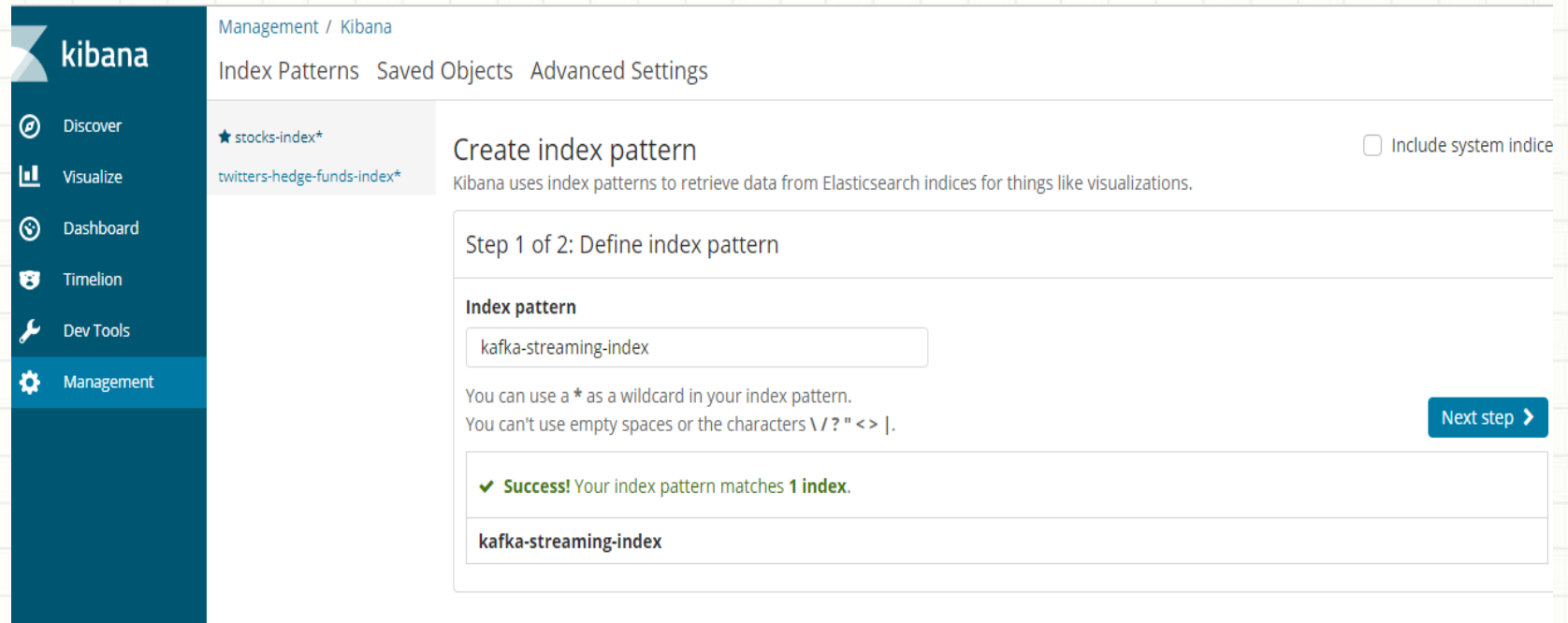
Filter

All field types

name	type	format	searchable	aggregatable	excluded	controls
_id	string		✓	✓		
_index	string		✓	✓		
_____	number					

# Building Spark Streaming Applications with Kafka

- **Kibana console**



The screenshot displays the Kibana console interface. On the left is a dark blue sidebar with the 'kibana' logo and a menu containing 'Discover', 'Visualize', 'Dashboard', 'Timelion', 'Dev Tools', and 'Management' (which is highlighted). The main content area has a header 'Management / Kibana' and sub-navigation links 'Index Patterns', 'Saved Objects', and 'Advanced Settings'. Below this, there's a list of index patterns: '★ stocks-index\*' and 'twitters-hedge-funds-index\*'. The main section is titled 'Create index pattern' with a checkbox 'Include system indices' on the right. Below the title is a description: 'Kibana uses index patterns to retrieve data from Elasticsearch indices for things like visualizations.' The current step is 'Step 1 of 2: Define index pattern'. There is a text input field labeled 'Index pattern' containing the text 'kafka-streaming-index'. Below the input field, a message states: 'You can use a \* as a wildcard in your index pattern. You can't use empty spaces or the characters \ / ? " < > |.' To the right of this message is a 'Next step >' button. At the bottom, a green success message reads: '✓ Success! Your index pattern matches 1 index.' Below this message, the matched index pattern 'kafka-streaming-index' is listed.

Management / Kibana

Index Patterns Saved Objects Advanced Settings

★ stocks-index\*  
twitters-hedge-funds-index\*

## Create index pattern

☐ Include system indices

Kibana uses index patterns to retrieve data from Elasticsearch indices for things like visualizations.

Step 1 of 2: Define index pattern

**Index pattern**

kafka-streaming-index

You can use a \* as a wildcard in your index pattern.  
You can't use empty spaces or the characters \ / ? " < > |.

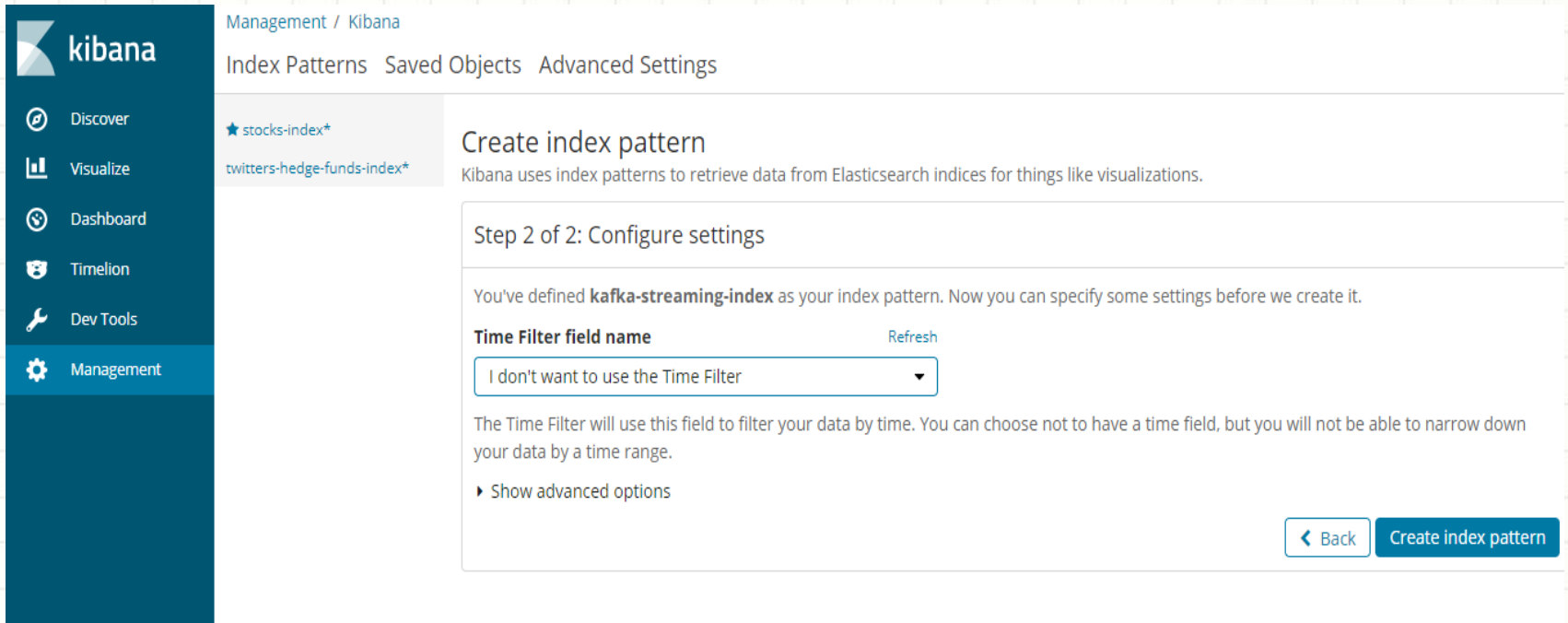
**Next step >**

✓ **Success!** Your index pattern matches **1 index**.

kafka-streaming-index

# Building Spark Streaming Applications with Kafka

- **Kibana console**



The screenshot shows the Kibana Management console interface. On the left is a dark blue sidebar with the Kibana logo and navigation links: Discover, Visualize, Dashboard, Timelion, Dev Tools, and Management (which is highlighted). The main content area has a light blue header with 'Management / Kibana' and links for 'Index Patterns', 'Saved Objects', and 'Advanced Settings'. Below the header, there's a list of index patterns: 'stocks-index\*' (with a star icon) and 'twitters-hedge-funds-index\*'. The 'Create index pattern' page is displayed, showing 'Step 2 of 2: Configure settings'. A text block explains that the user has defined 'kafka-streaming-index' and can now specify settings. A 'Time Filter field name' section contains a dropdown menu with the text 'I don't want to use the Time Filter' and a 'Refresh' link. Below this, a paragraph explains the purpose of the Time Filter. At the bottom right, there are two buttons: 'Back' and 'Create index pattern'.

Management / Kibana

Index Patterns Saved Objects Advanced Settings

★ stocks-index\*

twitters-hedge-funds-index\*

## Create index pattern

Kibana uses index patterns to retrieve data from Elasticsearch indices for things like visualizations.

### Step 2 of 2: Configure settings

You've defined **kafka-streaming-index** as your index pattern. Now you can specify some settings before we create it.

**Time Filter field name** Refresh

I don't want to use the Time Filter ▼

The Time Filter will use this field to filter your data by time. You can choose not to have a time field, but you will not be able to narrow down your data by a time range.

► Show advanced options

◀ Back Create index pattern

# Building Spark Streaming Applications with Kafka

- **Kibana console**

The screenshot displays the Kibana console interface. The top navigation bar includes the Kibana logo and a sidebar with menu items: Discover, Visualize, Dashboard, Timelion, Dev Tools, and Management. The main content area shows search results for the 'kafka-streaming-index'. The search bar contains the query 'search... (e.g. status:200 AND extension:PHP)' and indicates 'Uses lucene query syntax'. The results section shows 5 hits, with the first four displayed. Each hit is a JSON object containing fields such as currency.name, currency.price, currency.rate, customer.city, customer.country, customer.id, customer.ipAddress, customer.name, event, timestamp, \_id, \_type, \_index, and \_score.

5 hits New Save Open Share

search... (e.g. status:200 AND extension:PHP) Uses lucene query syntax Q

Discover Add a filter +

Visualize kafka-streaming-index

Dashboard

Timelion

Dev Tools

Management

Selected Fields

? \_source

Available Fields ⚙

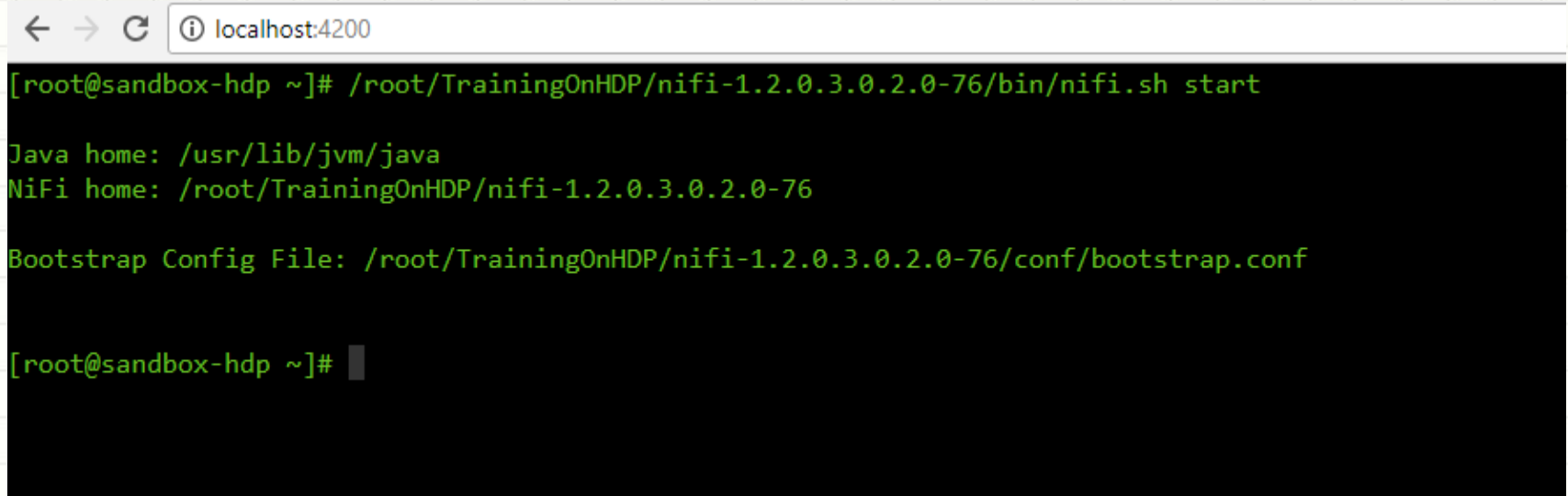
- t \_id
- t \_index
- # \_score
- t \_type
- t currency.name
- t currency.price
- # currency.rate
- t customer.city

**\_source**

- currency.name:** bitcoin **currency.price:** USD **currency.rate:** 0 **customer.city:** Moscow **customer.country:** Russian Federation **customer.id:** 86689427 **customer.ipAddress:** 95.31.18.119 **customer.name:** Edward S. **event:** CUSTOMER\_SEES\_BTCPRIE **timestamp:** July 3rd 2017, 08:00:35.000 **\_id:** EUJFJmEB11zoNlbVtFvM **\_type:** bitcoin **\_index:** kafka-streaming-index **\_score:** 1
- currency.name:** bitcoin **currency.price:** USD **currency.rate:** 0 **customer.city:** Quito **customer.country:** Ecuador **customer.id:** 56886468 **customer.ipAddress:** 186.46.129.15 **customer.name:** Lindsay M. **event:** CUSTOMER\_SEES\_BTCPRIE **timestamp:** July 11th 2017, 15:00:35.000 **\_id:** FEJHJmEB11zoNlbVL1t\_ **\_type:** bitcoin **\_index:** kafka-streaming-index **\_score:** 1
- currency.name:** bitcoin **currency.price:** USD **currency.rate:** 0 **customer.city:** Moscow **customer.country:** Russian Federation **customer.id:** 86689427 **customer.ipAddress:** 95.31.18.119 **customer.name:** Edward S. **event:** CUSTOMER\_SEES\_BTCPRIE **timestamp:** July 3rd 2017, 08:00:35.000 **\_id:** EEJFJmEB11zoNlbViFvj **\_type:** bitcoin **\_index:** kafka-streaming-index **\_score:** 1
- currency.name:** bitcoin **currency.price:** USD **currency.rate:** 0 **customer.city:** London **customer.country:** United Kingdom **customer.id:** 18313440 **customer.ipAddress:** 185.86.151.11 **customer.name:** Julian A. **event:** CUSTOMER\_SEES\_BTCPRIE **timestamp:** July 4th 2017, 11:00:35.000 **\_id:** EkJFJmEB11zoNlbV71u7 **\_type:** bitcoin **\_index:** kafka-streaming-index **\_score:** 1

# Building Spark Streaming Applications with Kafka

- **Start the NiFi**

A terminal window with a dark background and light green text. The window title bar shows navigation icons and the address 'localhost:4200'. The terminal content shows a user at a root prompt running the command to start NiFi. The output displays the Java and NiFi home paths and the bootstrap configuration file path. The prompt returns to the root user.

```
< > ↻ ⓘ localhost:4200
[root@sandbox-hdp ~]# /root/TrainingOnHDP/nifi-1.2.0.3.0.2.0-76/bin/nifi.sh start

Java home: /usr/lib/jvm/java
NiFi home: /root/TrainingOnHDP/nifi-1.2.0.3.0.2.0-76

Bootstrap Config File: /root/TrainingOnHDP/nifi-1.2.0.3.0.2.0-76/conf/bootstrap.conf

[root@sandbox-hdp ~]#
```

# Building Spark Streaming Applications with Kafka

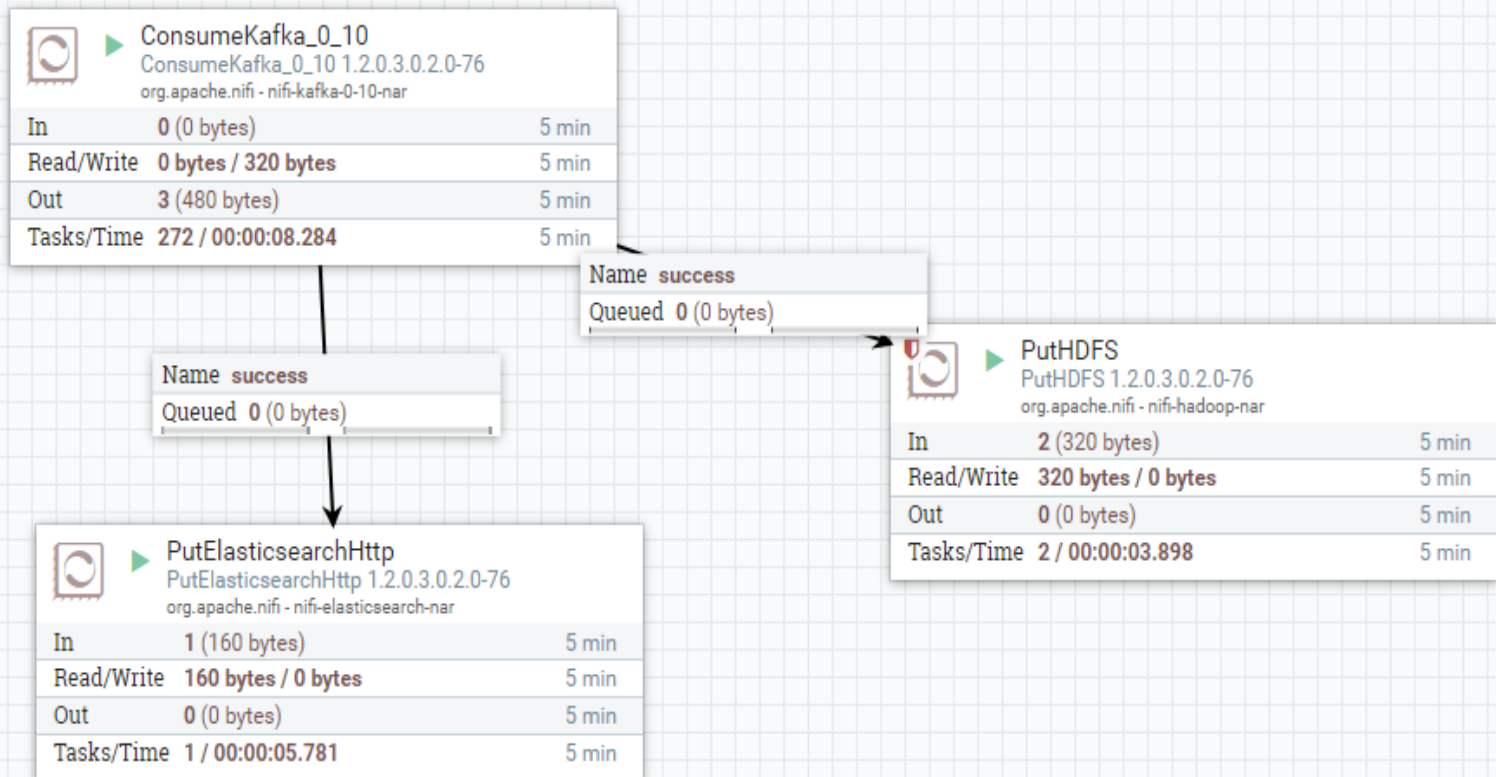
- **Publish the invalid message**

```
[root@sandbox-hdp ~]# /usr/hdp/2.6.3.0-235/kafka/bin/kafka-console-producer.sh --broker-list sandbox-hdp.hortonworks.com:6667 --topic source-topic
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one, then you should configure the number of parallel GC threads appropriately using -XX:ParallelGCThreads=N
{"event": "CUSTOMER_SEES_BTCPRI", "customer": {"id": "86689427", "name": "Edward S.", "ipAddress": "95.31.18.119"}, "currency": {"name": "bitcoin", "price": "USD"}, "timestamp": "2017-07-03T12:00:35Z"}
{"event": "CUSTOMER_SEES_BTCPRI", "customer": {"id": "86689427", "name": "Edward S.", "ipAddress": "95.31.18.119"}, "currency": {"name": "bitcoin", "price": "USD"}, "timestamp": "2017-07-03T12:00:35Z"}
{"event": "CUSTOMER_SEES_BTCPRI", "customer": {"id": "18313440", "name": "Julian A.", "ipAddress": "185.86.151.11"}, "currency": {"name": "bitcoin", "price": "USD"}, "timestamp": "2017-07-04T15:00:35Z"}
{"event": "CUSTOMER_SEES_BTCPRI", "customer": {"id": "56886468", "name": "Lindsay M.", "ipAddress": "186.46.129.15"}, "currency": {"name": "bitcoin", "price": "USD"}, "timestamp": "2017-07-11T19:00:35Z"}
{"event": "CUSTOMER_SEES_BTCPRI", "customer": {"id": "56886468", "name": "Lindsay M.", "ipAddress": "186.46.129.15"}, "currency": {"name": "bitcoin", "price": "USD"}, "timestamp": "2017-07-11T19:00:35Z"}
This is the test message
```



# Building Spark Streaming Applications with Kafka

- NiFi Status



# Building Spark Streaming Applications with Kafka

- Check HDFS folder

```
[root@sandbox-hdp ~]# hadoop fs -ls /user/root/kafka/error
Found 1 items
-rw-r--r--  1 root hdfs          160 2018-01-24 04:04 /user/root/kafka/error/14295956977544
[root@sandbox-hdp ~]#
```

# Building Spark Streaming Applications with Kafka

- **Kibana console**

The screenshot displays the Kibana console interface. The browser address bar shows the URL: `localhost:8744/app/kibana#/discover?_g=()&_a=(columns:!(source),index:a5493490-00bc-11e8-9f9f-77bb74b9e617,interval:auto,query:(language:luce,query:),sort:!(score,desc))`. The Kibana logo and navigation menu are on the left. The main area shows a search for `source` in the `kafka-streaming-error-index`, resulting in 1 hit. The search bar contains `search... (e.g. status:200 AND extension:PHP)` and a note `Uses lucene query syntax`. The left sidebar lists navigation options: Discover, Visualize, Dashboard, Timelion, Dev Tools, and Management. The main content area shows the search results for the `source` field. The results table has columns `_id`, `_type`, `_index`, and `_score`. The single result shows an error message: `error: JsonParseException: Unrecognized token 'This': was expecting ('true', 'false' or 'null')` at [Source: This is the test message; line: 1, column: 5]. The `_id` is `FUJbJmEB11zoNlbVYFsD`, `_type` is `error`, `_index` is `kafka-streaming-error-index`, and `_score` is `1`.

1 hit

New Save Open Share

search... (e.g. status:200 AND extension:PHP) Uses lucene query syntax

Discover

Visualize

Dashboard

Timelion

Dev Tools

Management

kafka-streaming-error-index

Selected Fields

? \_source

Available Fields

t \_id

t \_index

\_source

error: JsonParseException: Unrecognized token 'This': was expecting ('true', 'false' or 'null') at [Source: This is the test message; line: 1, column: 5] \_id: FUJbJmEB11zoNlbVYFsD \_type: error \_index: kafka-streaming-error-index \_score: 1