Kafka Integration with Database(MemSQL(Free Tier))

Kafka Installation

1. Download kafka on Master Aggregator of MemSQL

   wget http://mirrors.estointernet.in/apache/kafka/2.4.0/kafka_2.12-2.4.0.tgz
   tar -xzf kafka_2.12-2.4.0.tgz
   cd kafka_2.12-2.4.0

2. Kafka uses ZooKeeper so you need to first start a ZooKeeper server.

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

3. Now start the Kafka server:

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

4. Create a Topic
   Let's create a topic named "test" with a single partition and only one replica:

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

   We can now see that topic if we run the list topic command:
   - bin/kafka-topics.sh --list --bootstrap-server localhost:9092
   test

5. Send some messages
   Run the producer and then type a few messages into the console to send to the server.

   - bin/kafka-console-producer.sh --broker-list localhost:9092 --topic test

   >India is large country.

6. Kafka also has a command line consumer that will dump out messages to standard output.

   - bin/kafka-console-consumer.sh --bootstrap-server localhost:9092 --topic test --from-beginning

     India is large country.

Database integration with Kafka Topic

1. Create a table to ingest data from kafka topic

```
CREATE TABLE `kafka_topic_store` (
  `description` text CHARACTER SET utf8 COLLATE utf8_general_ci
  /*!90618 , SHARD KEY () */
);
```

2. Create a Pipeline to receive data from kafka topic

```
memsql> CREATE PIPELINE my_pipeline AS LOAD DATA KAFKA
'192.168.11.21:9092/test' INTO TABLE kafka_topic_store;
```

Note: by default kafka works on port 9092 so ensure that this port is reachable.

3. Start Pipeline to finally get data from kafka topic

```
memsql> start pipeline my_pipeline;
```

4. See ingest data from table

```
memsql> select * from kafka_topic_store;
+-------------------------------+
| description                   |
+-------------------------------+
| India is large country.       |
```

Setting up a multi-broker cluster

```
cp config/server.properties config/server-1.properties
```

Now edit these new files and set the following properties:

config/server-1.properties:

```
broker.id=1
listeners=PLAINTEXT://:9093
log.dirs=/tmp/kafka-logs-1
```

We already have Zookeeper and our single node started, so we just need to start the two new nodes:

> bin/kafka-server-start.sh config/server-1.properties &

Now create a new topic with a replication factor of Two (one for earlier already running :

> bin/kafka-topics.sh --create --bootstrap-server localhost:9092 --replication-factor 2 --partitions 1 --topic my-replicated-topic

Okay but now that we have a cluster how can we know which broker is doing what? To see that run the "describe topics" command:

> bin/kafka-topics.sh --describe --bootstrap-server localhost:9092 --topic my-replicated-topic

Let's publish a few messages to our new topic:

> bin/kafka-console-producer.sh --broker-list localhost:9092 --topic my-replicated-topic

:END: