

Week 6

Felipe Rodriguez

Bellevue University

DSC 650 Big Data

Professor Nasheb Ismaily

January 28, 2024

Exercise 1 and 2 Kafka topic named 'my-topic'.

```
[feliperodriguez@big-data:~/dsc650-infra/bellevue-bigdata/kafka$ docker exec -it kafka_kafka_1 bash ]
[root@65e7a18d47e4:/# /opt/kafka_2.13-2.8.1/bin/kafka-topics.sh --create --topic my-topic --bootstrap-server localhost:9092
Created topic my-topic.
[root@65e7a18d47e4:/# /opt/kafka_2.13-2.8.1/bin/kafka-topics.sh --list --bootstrap-server localhost:9092
my-topic
root@65e7a18d47e4:/# █
```

Exercise 3 and 4 Producing and Consuming Messages in Kafka

```
[feliperodriguez@big-data:~/dsc650-infra/bellevue-bigdata/kafka$ docker exec -it kafka_kafka_1 bash ]
[root@65e7a18d47e4:/# /opt/kafka_2.13-2.8.1/bin/kafka-console-producer.sh --topic my-topic --bootstrap-server localhost:9092
>hello world
>█
```

```
my-topic
[root@65e7a18d47e4:/# /opt/kafka_2.13-2.8.1/bin/kafka-console-consumer.sh --topic my-topic --from-beginning --bootstrap-server localhost:9092
hello world
█
```

Exercise 5 and 6 Kafka Performance Tests

```
root@65e7a18d47e4:/#
/opt/kafka_2.13-2.8.1/bin/kafka-producer-perf-test.sh --topic my-topic --num-records 50000 --record-size 100 --throughput 1000 --producer-props bootstrap.servers=localhost:9092
n.serialization.StringSerializer value.serializer=org.apache.kafka.common.serialization.StringSerializer
4999 records sent, 999.6 records/sec (0.10 MB/sec), 2.8 ms avg latency, 389.0 ms max latency.
5006 records sent, 1001.2 records/sec (0.10 MB/sec), 0.7 ms avg latency, 7.0 ms max latency.
5002 records sent, 1000.2 records/sec (0.10 MB/sec), 0.6 ms avg latency, 12.0 ms max latency.
5002 records sent, 1000.2 records/sec (0.10 MB/sec), 0.5 ms avg latency, 5.0 ms max latency.
5002 records sent, 1000.4 records/sec (0.10 MB/sec), 0.4 ms avg latency, 3.0 ms max latency.
5001 records sent, 1000.0 records/sec (0.10 MB/sec), 0.4 ms avg latency, 6.0 ms max latency.
5000 records sent, 1000.0 records/sec (0.10 MB/sec), 0.4 ms avg latency, 4.0 ms max latency.
5002 records sent, 1000.4 records/sec (0.10 MB/sec), 0.4 ms avg latency, 8.0 ms max latency.
5003 records sent, 1000.4 records/sec (0.10 MB/sec), 0.4 ms avg latency, 2.0 ms max latency.
50000 records sent, 999.800040 records/sec (0.10 MB/sec), 0.71 ms avg latency, 389.00 ms max latency, 1 ms 50th, 1 ms 95th, 4 ms 99th, 30 ms 99.9th.
root@65e7a18d47e4:/# /opt/kafka_2.13-2.8.1/bin/kafka-consumer-perf-test.sh --broker-list localhost:9092 --topic my-topic --messages 50000
start.time, end.time, data.consumed.in.MB, MB.sec, data.consumed.in.nMsg, nMsg.sec, rebalance.time.ms, fetch.time.ms, fetch.MB.sec, fetch.nMsg.sec
2024-01-29 01:05:15:352, 2024-01-29 01:05:16:070, 4.7684, 6.6412, 50001, 69639.2758, 438, 280, 17.0299, 178575.0000
root@65e7a18d47e4:/# █
```

Interpretation

This performance test allows us to understand how the producer and consumer will act with a certain number of messages. In this example, we are using 50,000 messages. A couple of items to look at are the amount of time it took to generate the messages and the amount of data consumed. This test was small, so it performed well. This test requires us to pay attention to two important metrics; Latency measures - how long it takes to process one event and throughput measures - how many events arrive within a specific amount of time.

Exercise 5 through 8 Expanding Kafka and Running Additional Performance Tests

```
root@65e7a18d47e4:/# /opt/kafka_2.13-2.8.1/bin/kafka-producer-perf-test.sh --topic my-partitioned-topic --num-records 50000 --record-size 100 --throughput 1000
serializer=org.apache.kafka.common.serialization.StringSerializer value.serializer=org.apache.kafka.common.serialization.StringSerializer
4993 records sent, 998.4 records/sec (0.10 MB/sec), 16.5 ms avg latency, 483.0 ms max latency.
5011 records sent, 1002.2 records/sec (0.10 MB/sec), 3.9 ms avg latency, 27.0 ms max latency.
5005 records sent, 1001.0 records/sec (0.10 MB/sec), 2.4 ms avg latency, 50.0 ms max latency.
5002 records sent, 1000.4 records/sec (0.10 MB/sec), 1.3 ms avg latency, 26.0 ms max latency.
5001 records sent, 1000.2 records/sec (0.10 MB/sec), 1.4 ms avg latency, 32.0 ms max latency.
5004 records sent, 1000.8 records/sec (0.10 MB/sec), 0.9 ms avg latency, 10.0 ms max latency.
5002 records sent, 1000.2 records/sec (0.10 MB/sec), 0.8 ms avg latency, 14.0 ms max latency.
5001 records sent, 1000.2 records/sec (0.10 MB/sec), 0.7 ms avg latency, 5.0 ms max latency.
5002 records sent, 1000.2 records/sec (0.10 MB/sec), 0.6 ms avg latency, 7.0 ms max latency.
50000 records sent, 999.780048 records/sec (0.10 MB/sec), 2.91 ms avg latency, 483.00 ms max latency, 1 ms 50th, 10 ms 95th, 26 ms 99th, 131 ms 99.9th.
root@65e7a18d47e4:/# /opt/kafka_2.13-2.8.1/bin/kafka-consumer-perf-test.sh --broker-list localhost:9092 --topic my-partitioned-topic --messages 50000
start.time, end.time, data.consumed.in.MB, MB.sec, data.consumed.in.nMsg, nMsg.sec, rebalance.time.ms, fetch.time.ms, fetch.MB.sec, fetch.nMsg.sec
2024-01-29 01:23:18:651, 2024-01-29 01:23:19:409, 4.7684, 6.2907, 50000, 65963.0607, 434, 324, 14.7172, 154320.9877
root@65e7a18d47e4:/#
```

Interpretation

Between the two performance tests there are not many differences however when performing these tests most systems are optimized for either latency or throughput. The first test consumed 178,575 in a second while the second test was 154,320. The other factor is MB.sec or how much data is transferred in megabytes per second. The first test consumed more data than the second. This makes sense with the number of messages consumed.