

DSC 650 – Big Data

Week 7 Assignment

Eyram Kueviakoe

04/23/2024

Screenshot showing the 'my-topic' listed amongst the topics.

```
rsa-key-20240315@dsc650-kueviakoe: ~/dsc650-infra/bellevue-bigdata/kafka
rsa-key-20240315@dsc650-kueviakoe:~/dsc650-infra/bellevue-bigdata/kafka$ sudo docker exec -it kafka_kafka_1 bash
root@f236ce45878c:/# /opt/kafka_2.13-2.8.1/bin/kafka-topics.sh --create --topic my-topic --bootstrap-server localhost:9092
Created topic my-topic.
root@f236ce45878c:/# /opt/kafka_2.13-2.8.1/bin/kafka-topics.sh --list --bootstrap-server localhost:9092
my-topic
root@f236ce45878c:/#
```

Screenshot of the producer

```
rsa-key-20240315@dsc650-kueviakoe: ~/dsc650-infra/bellevue-bigdata/kafka
root@f236ce45878c:/# /opt/kafka_2.13-2.8.1/bin/kafka-console-producer.sh --topic my-topic --bootstrap-server localhost:9092
>This is DSC 650 Week 7 Assignment
>Today's date is Tuesday April 23rd, 2024
>
```

Screenshot of the consumer terminal

```
rsa-key-20240315@dsc650-kueviakoe: ~/dsc650-infra/bellevue-bigdata/kafka
root@f236ce45878c:/# /opt/kafka_2.13-2.8.1/bin/kafka-console-consumer.sh --topic my-topic --from-beginning --bootstrap-server localhost:9092
This is DSC 650 Week 7 Assignment
Today's date is Tuesday April 23rd, 2024
```

Screenshot of the producer performance test results

```
rsa-key-20240315@dsc650-kueviakoe: ~/dsc650-infra/bellevue-bigdata/kafka
root@f236ce45878c:/# /opt/kafka_2.13-2.8.1/bin/kafka-producer-perf-test.sh --topic my-topic --num-records 50000 --record-size 100 --throughput 1000 --
ingSerializer value.serializer=org.apache.kafka.common.serialization.StringSerializer
5001 records sent, 1000.2 records/sec (0.10 MB/sec), 3.5 ms avg latency, 401.0 ms max latency.
5003 records sent, 1000.4 records/sec (0.10 MB/sec), 0.9 ms avg latency, 12.0 ms max latency.
5000 records sent, 1000.0 records/sec (0.10 MB/sec), 0.7 ms avg latency, 6.0 ms max latency.
5001 records sent, 1000.2 records/sec (0.10 MB/sec), 0.6 ms avg latency, 8.0 ms max latency.
5001 records sent, 1000.2 records/sec (0.10 MB/sec), 0.6 ms avg latency, 13.0 ms max latency.
5002 records sent, 1000.2 records/sec (0.10 MB/sec), 0.6 ms avg latency, 5.0 ms max latency.
5001 records sent, 1000.0 records/sec (0.10 MB/sec), 0.5 ms avg latency, 6.0 ms max latency.
5001 records sent, 1000.0 records/sec (0.10 MB/sec), 0.5 ms avg latency, 2.0 ms max latency.
5001 records sent, 1000.2 records/sec (0.10 MB/sec), 0.5 ms avg latency, 10.0 ms max latency.
50000 records sent, 999.820032 records/sec (0.10 MB/sec), 0.89 ms avg latency, 401.00 ms max latency, 1 ms 50th, 1 ms 95th, 5 ms 99th, 33 ms 99.9th.
```

Screenshot of the consumer performance test results

```
rsa-key-20240315@dsc650-kueviakoe: ~/dsc650-infra/bellevue-bigdata/kafka
root@f236ce45878c:/# /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
2024-04-24 00:00:17:215, 2024-04-24 00:00:18:015, 4.7684, 5.9606, 50002, 62502.5000, 443, 357, 13.3570, 140061.6246
root@f236ce45878c:/#
```

Discussions of the results

Producer Test: Sent 50,000 records with an average throughput of around 999.82 records/sec and an average latency of 0.89 ms, with occasional spikes in max latency up to 401 ms.

Consumer Test: Consumed 50,000 messages with a consumption rate of 62,502.5 messages/sec and a fetch throughput of 13.3570 MB/sec, showing efficient message retrieval and processing.

The test results indicate that the Kafka producer was able to achieve the target throughput with relatively low latency. The consumer was also able to keep up with the producer, consuming all messages efficiently.

Screenshots of the performance tests on the partitioned topic

```
rsa-key-20240315@dsc650-kueviakoe: ~/dsc650-infra/bellevue-bigdata/kafka
root@f236ce45878c:/# /opt/kafka-2.13-2.8.1/bin/kafka-topics.sh --create --topic my-partitioned-topic --replication-factor 3 --partitions 3 --bootstrap-server localhost:9092
Created topic my-partitioned-topic.
root@f236ce45878c:/# /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 --producer
rvers=localhost:9092 key.serializer=org.apache.kafka.common.serialization.StringSerializer value.serializer=org.apache.kafka.common.serialization.StringSerializer
4999 records sent, 999.6 records/sec (0.10 MB/sec), 20.7 ms avg latency, 423.0 ms max latency.
5007 records sent, 1001.4 records/sec (0.10 MB/sec), 4.8 ms avg latency, 40.0 ms max latency.
4995 records sent, 998.8 records/sec (0.10 MB/sec), 4.9 ms avg latency, 80.0 ms max latency.
5010 records sent, 1002.0 records/sec (0.10 MB/sec), 3.0 ms avg latency, 50.0 ms max latency.
5000 records sent, 1000.0 records/sec (0.10 MB/sec), 1.9 ms avg latency, 15.0 ms max latency.
5007 records sent, 1001.2 records/sec (0.10 MB/sec), 1.6 ms avg latency, 29.0 ms max latency.
5002 records sent, 1000.4 records/sec (0.10 MB/sec), 1.1 ms avg latency, 16.0 ms max latency.
5002 records sent, 1000.4 records/sec (0.10 MB/sec), 1.4 ms avg latency, 18.0 ms max latency.
5001 records sent, 1000.2 records/sec (0.10 MB/sec), 1.0 ms avg latency, 18.0 ms max latency.
50000 records sent, 999.720078 records/sec (0.10 MB/sec), 4.12 ms avg latency, 423.00 ms max latency, 1 ms 50th, 12 ms 95th, 41 ms 99th, 153 ms 99.9th.
root@f236ce45878c:/#
```

```
root@f236ce45878c:/# /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
2024-04-24 00:10:50:172, 2024-04-24 00:10:51:006, 4.7684, 5.7175, 50000, 59952.0384, 442, 392, 12.1642, 127551.0204
root@f236ce45878c:/#
```

Observations on performance variations between a single Kafka instance and the scaled setup

Producer Test: Sent 50,000 records with an average throughput of approximately 999.72 records/sec and an average latency of 4.12 ms, with occasional spikes in max latency up to 423 ms.

Consumer Test: Consumed 50,000 messages with a consumption rate of 59,952.0384 messages/sec and a fetch throughput of 12.1642 MB/sec, indicating efficient message retrieval and processing from the partitioned topic.

There is a significant difference in average latency between the single partition test (0.89ms) and the portioned topic test (4.12ms). This variation in performance is due to how the workload is distributed across the 3 partitions in the partitioned topic, which reduces the pressure on any single partition.