Jamie Tran

Professor Nasheb Ismaily

DSC650 Big Data

17 July 2024

Week 7 Screenshots

**Screenshot showing the ‘my-topic’ listed amongst the topics:**

**A screenshot of a computer screen

Description automatically generated**

**Screenshots of the producer terminal with your entered text and the consumer terminal showing the received message:**

**A screenshot of a computer screen

Description automatically generated**

**Screenshots of both the producer and consumer performance test results:**

**A screenshot of a computer screen

Description automatically generated**

**Discuss the meaning of the results:**

The first value in each row of the producer output lets us know the total number of records sent, which according to our results would be anywhere between 5000-5004 records. The second value in each row tells us the average rate sending rate of the records. For instance, the first row is telling us 1000.2 records are sent each second. From there the next number tells us the average time it takes for each record to be sent. Based on our results it takes anywhere from 0.5-3.3 milliseconds on average to send one record. The last value in the rows informs us of the maximum amount of time it has taken for a record to be sent. This would be anywhere between 2-400 milliseconds. The final row in these results gives us a summary of all the previous rows.

As for the consumer results they start and end time of the report, this shows us that the results took 736 milliseconds to generate. The second and third value lets us know that the data consumed 4.7684 megabytes of storage and the speed at which it was able to be processed. Turns out the system processed 6.4788 megabytes of data per second. Next the output tells us the 50,004 messages were consumed during this test and 67,940.2174 messages are processed per second. The next two values tell us that the program took 315 milliseconds to retrieve the data from the program, the values paired with it lets us know that the program retrieved 15.14 megabytes of data per second and can retrieve 158,742.8571 messages per second.

**Screenshots of the performance tests on the portioned topic:**

**A screen shot of a computer

Description automatically generated**

**Include your observations on performance variations between a single Kafka instance and the scaled setup:**

For the most part the results show us a similar result as the single Kafka instance versus the scaled setup. The main differences would be that the producer output gives us slightly better results with the single instance and the consumer output is split on performance results. For example, in the producer results the single instance was able to send 0.059974 more records per second than the scaled. The single instance also required a lot less time to send a record with their average latency at 0.82 and the scala at 3.67. Final call out from the producer output is the maximum time for a record to be sent was 55 milliseconds higher for the scala than the single instance.

Moving on to the consumer output the scala gives us better results in the first portion consuming data 0.6298 megabytes per second faster and processing 6,515.156 more messages per second. However, the single instance takes less time rebalancing and fetching data while also fetching more megabytes and messages per second.

All in all, the single instance is the better and more efficient option given our current data. However, they are both so similar you would be just fine using either.