**Performance Experiment - MQ - Actor Count**

The goal of this experiment is to benchmark how this technology performs when hardware resources are plentiful and the number of actors (publishers and subscribers) in the system varies.

Contact [@jmbeach](https://github.com/jmbeach) or [@laynemoseley](https://github.com/laynemoseley) for clarifications.

**Preliminary Steps**

**Producer Program**

Design a **python** program utilizing your assigned MQ technology that produces as many messages per second as possible (For example: while true: produce\_message())

For MQ’s that run under multiple operating modes, run under the mode that is most reliable. An example of this would be RabbitMQ which can use “at most once” and “at least once” delivery by choosing whether or not to use acknowledgements. [Reference](https://www.rabbitmq.com/reliability.html). You should choose “at least once”.

For data collection, the program should output (to standard output) a CSV of message ID’s, the time they were sent, and the time the acknowledge (ACK) response was received (if applicable).

Ex output:

1,1617516000,1617516005

2,1617516010,1617516015

...

Each message from the producer should be the exact same. A byte UTF-8 string of the following quote:

There are not more than five primary colors (blue, yellow, red, white, and black), yet in combination they produce more hues than can ever been seen.

This is a quote from [Sun Tzu's Art of War](https://www.gutenberg.org/files/132/132-h/132-h.htm#:~:text=There%20are%20not%20more%20than%20five%20musical%20notes) and it's a fun, small, public-domain message that will ensure that everyone is using a message of the exact same size (149 bytes).

The message ID’s need to be unique so it may be a good idea to use a unique ID prefix per publisher.

**Note:** The reason we are using standard out is to prevent using file operations that could slow down the overall performance.

The program should send 1,000 messages and then terminate once all messages are sent and all acknowledgements are received.

Link to your program’s usage here:

[Replace me](Replace me)

**Consumer Program**

Design a **python** program utilizing your assigned MQ technology that consumes as many messages per second as possible (For example: while true: consume\_message())

For MQ’s that run under multiple operating modes, run under the mode that is most reliable. An example of this would be RabbitMQ which can use “at most once” and “at least once” delivery by choosing whether or not to use acknowledgements. [Reference](https://www.rabbitmq.com/reliability.html). You should choose “at least once”.

The consumer doesn’t have to do anything with the message; only retrieve it.

The program should terminate once all messages are consumed.

For data collection, the program should output (to standard output) a CSV of message ID’s, the time they were received, and the time they were acknowledged (if applicable).

Ex output:

1,1617516000,1617516005

2,1617516010,1617516015

...

Link to your program’s usage here:

[Replace me](Replace me)

**Data Processor**

The experiment designers will provide a program that accepts a list of producer csv files and a list of consumer csv files. It should output 2 CSV’s that marry up the matching ID’s of messages and contain calculated results.

The first csv should contain the following durations:

* Processing latency - total duration in milliseconds from the time the message is sent to the time the message is received
* Send time - total duration in milliseconds from the time the message is sent to the time the message arrives at the broker/queue/system (this is likely the ACK received time)

Ex:

1,15,8

2,8,4

...

The above output indicates that message 1 took 15ms to be fully processed and 8ms to be sent to the broker.

The second csv should output the number of messages sent in each second of the experiment and the number of messages received in each second of the experiment. This is the throughput.

1,15,30

2,8,16

...

The above output indicates that the end to end throughput in second 1 was 15 and the send to broker throughput in second 1 was 16.

Link to program pending.

**Hardware Configuration**

Each test should be ran using Ubuntu 18.04 chameleon cloud VM’s with 8 CPU and 32 GB of RAM. If these hardware specs aren’t attainable, use the nearest available and document the configuration.

**Experiment Configurations**

These are the configurations your program should be ran in.

| **Producer Machines** | **Producer Instances** | **Consumer Machines** | **Consumer Instances** |
| --- | --- | --- | --- |
| 1 | 1 | 1 | 1 |
| 1 | 1 | 1 | 5 |
| 1 | 5 | 1 | 1 |
| 3 | 8 | 3 | 8 |
| 4 | 25 | 4 | 25 |

If possible, implement these configurations with redundancy enabled and document the configuration.

These tests are described in detail below and have placeholders for results.

**Redundancy implementation Details**

Place the redundancy implementation details here.

**1 Producer - 1 Consumer**

Run one instance of your producer program on one VM.

Ensure you run your programs in such a way that the output is not lost. For example: python main.py > results.csv.

Run your consumer program with a matching configuration.

Fill this table with the appropriate links:

| **File Name** | **Link** |
| --- | --- |
| producer1.csv |  |
| consumer1.csv |  |
| latencies.csv |  |
| throughput.csv |  |

Calculate the following data using Excel or a custom program by analyzing the data in latencies.csv and throughput.csv

| **Metric** | **Value** |
| --- | --- |
| Processing Latency Min |  |
| Processing Latency Max |  |
| Processing Latency Average |  |
| Processing Latency Standard Deviation |  |
| Send Time Min |  |
| Send Time Max |  |
| Send Time Average |  |
| Send Time Standard Deviation |  |
| Processing Throughput Min |  |
| Processing Throughput Max |  |
| Processing Throughput Average |  |
| Processing Throughput Standard Deviation |  |
| Send Throughput Min |  |
| Send Throughput Max |  |
| Send Throughput Average |  |
| Send Throughput Standard Deviation |  |

**1 Producer VM - 1 Consumer VM: 5 Instances**

Run one instance of your producer program on one VM.

Ensure you run your programs in such a way that the output is not lost. For example: python main.py > results.csv.

Run five consumer program instances on one VM.

Fill this table with the appropriate links:

| **File Name** | **Link** |
| --- | --- |
| producer1.csv |  |
| consumer1.csv |  |
| latencies.csv |  |
| throughput.csv |  |

Calculate the following data using Excel or a custom program by analyzing the data in latencies.csv and throughput.csv

| **Metric** | **Value** |
| --- | --- |
| Processing Latency Min |  |
| Processing Latency Max |  |
| Processing Latency Average |  |
| Processing Latency Standard Deviation |  |
| Send Time Min |  |
| Send Time Max |  |
| Send Time Average |  |
| Send Time Standard Deviation |  |
| Processing Throughput Min |  |
| Processing Throughput Max |  |
| Processing Throughput Average |  |
| Processing Throughput Standard Deviation |  |
| Send Throughput Min |  |
| Send Throughput Max |  |
| Send Throughput Average |  |
| Send Throughput Standard Deviation |  |

**1 Producer VM: 5 Instances - 1 Consumer VM: 1 Instance**

Run five producer program instances on one VM.

Ensure you run your programs in such a way that the output is not lost. For example: python main.py > results.csv.

Run one instance of the consumer program on one VM.

Fill this table with the appropriate links:

| **File Name** | **Link** |
| --- | --- |
| producer1.csv |  |
| consumer1.csv |  |
| latencies.csv |  |
| throughput.csv |  |

Calculate the following data using Excel or a custom program by analyzing the data in latencies.csv and throughput.csv

| **Metric** | **Value** |
| --- | --- |
| Processing Latency Min |  |
| Processing Latency Max |  |
| Processing Latency Average |  |
| Processing Latency Standard Deviation |  |
| Send Time Min |  |
| Send Time Max |  |
| Send Time Average |  |
| Send Time Standard Deviation |  |
| Processing Throughput Min |  |
| Processing Throughput Max |  |
| Processing Throughput Average |  |
| Processing Throughput Standard Deviation |  |
| Send Throughput Min |  |
| Send Throughput Max |  |
| Send Throughput Average |  |
| Send Throughput Standard Deviation |  |

**3 Producer VM's: 8 Instances - 3 Consumers VM's: 8 Instances**

Run eight instances of your producer on three different machines (24 instances total).

Ensure you run your programs in such a way that the output is not lost. For example: python main.py > results.csv.

Run eight instances of your consumer program on three different machines as well.

Fill this table with the appropriate links:

| **File Name** | **Link** |
| --- | --- |
| producer1.csv |  |
| producer2.csv |  |
| producer3.csv |  |
| consumer1.csv |  |
| consumer2.csv |  |
| consumer3.csv |  |
| latencies.csv |  |
| throughput.csv |  |

Calculate the following data using Excel or a custom program by analyzing the data in latencies.csv and throughput.csv

| **Metric** | **Value** |
| --- | --- |
| Processing Latency Min |  |
| Processing Latency Max |  |
| Processing Latency Average |  |
| Processing Latency Standard Deviation |  |
| Send Time Min |  |
| Send Time Max |  |
| Send Time Average |  |
| Send Time Standard Deviation |  |
| Processing Throughput Min |  |
| Processing Throughput Max |  |
| Processing Throughput Average |  |
| Processing Throughput Standard Deviation |  |
| Send Throughput Min |  |
| Send Throughput Max |  |
| Send Throughput Average |  |
| Send Throughput Standard Deviation |  |

**4 Producer VM's: 25 Threads - 4 Consumer VM's: 25 Threads**

Run 25 instances of the producer program on four different VM's (100 instances total).

Ensure you run your programs in such a way that the output is not lost. For example: python main.py > results.csv.

Run 25 instances of the consumer program on four different VM's (100 instances total).

Fill this table with the appropriate links:

| **File Name** | **Link** |
| --- | --- |
| producer1.csv |  |
| producer2.csv |  |
| producer3.csv |  |
| producer4.csv |  |
| consumer1.csv |  |
| consumer2.csv |  |
| consumer3.csv |  |
| consumer4.csv |  |
| latencies.csv |  |
| throughput.csv |  |

Calculate the following data using Excel or a custom program by analyzing the data in latencies.csv and throughput.csv

| **Metric** | **Value** |
| --- | --- |
| Processing Latency Min |  |
| Processing Latency Max |  |
| Processing Latency Average |  |
| Processing Latency Standard Deviation |  |
| Send Time Min |  |
| Send Time Max |  |
| Send Time Average |  |
| Send Time Standard Deviation |  |
| Processing Throughput Min |  |
| Processing Throughput Max |  |
| Processing Throughput Average |  |
| Processing Throughput Standard Deviation |  |
| Send Throughput Min |  |
| Send Throughput Max |  |
| Send Throughput Average |  |
| Send Throughput Standard Deviation |  |