Kafka Capacity Calculation Report :

This report contains the capacity based on network and disk utilisation.

Test Machine Hardware Spec

Disk Capacity : 1 TB
NIC : 1000 Mbps
RAM : 32 GB
Processors : 32

Network Observation :

Kafka Client JMX metrics monitored :

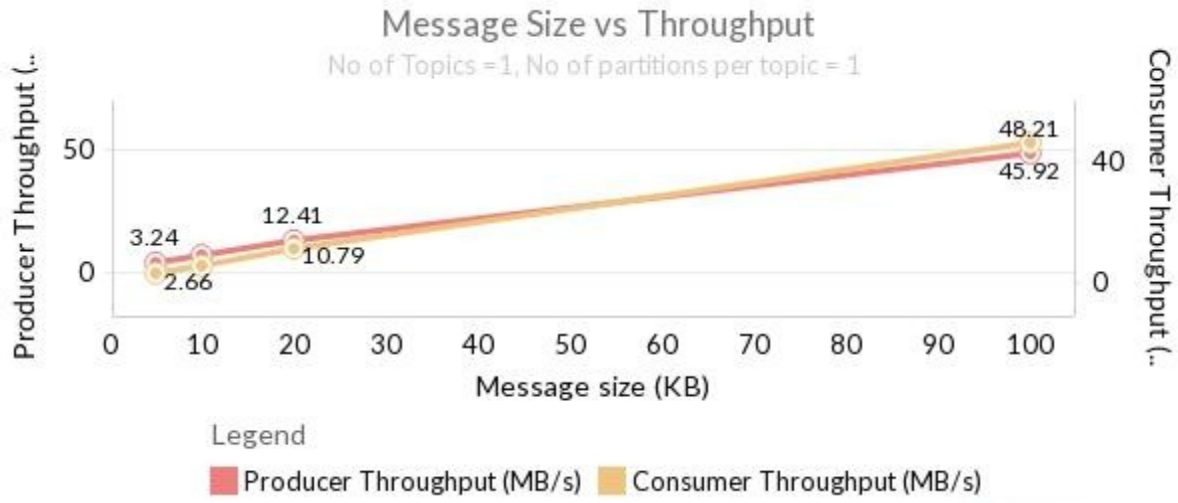
1. producer-outgoing-byte-rate
2. consumer-incoming-byte-rate
3. records-send-rate
4. records-consumed-rate
5. produce-request-latency
6. fetch-request-latency

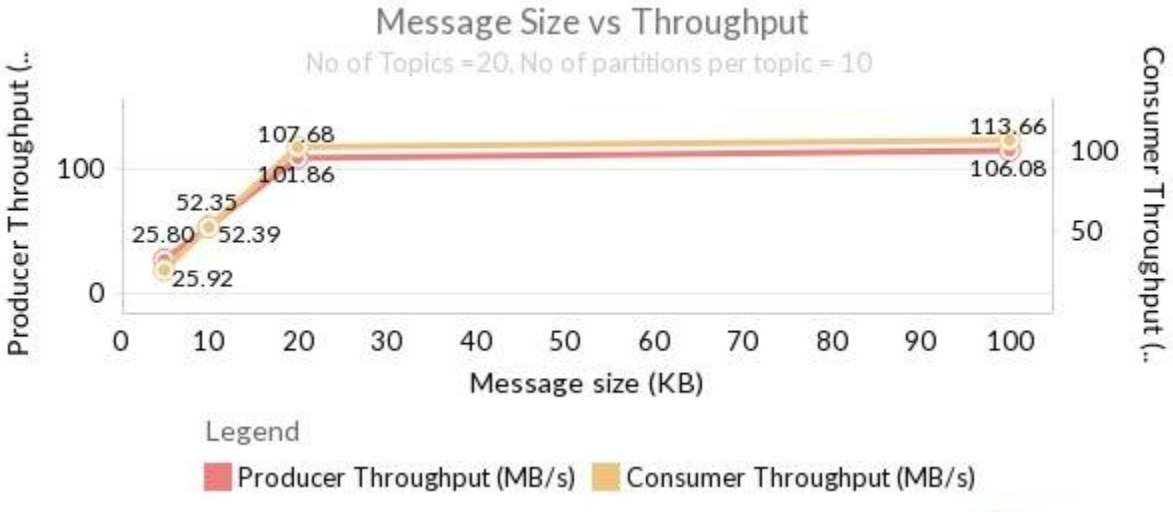
Consolidated Test Results :
https://sheet.zoho.com/sheet/open/4rd5s258426f425b744e09e9aacf677af060e/sheets/ConsolidatedReport-clientPerf/ranges/E38

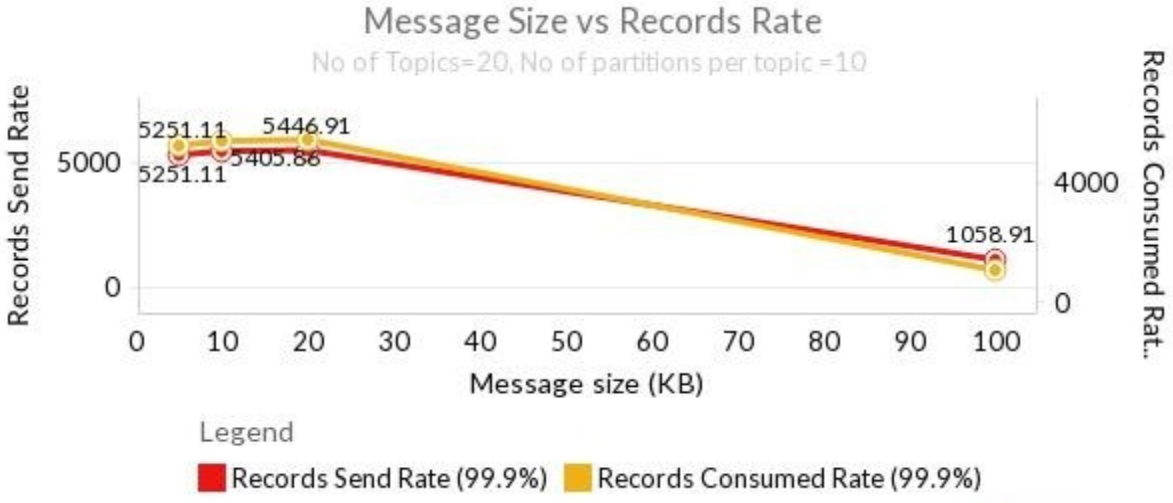
The results which are marked here are 99.9th percentile.

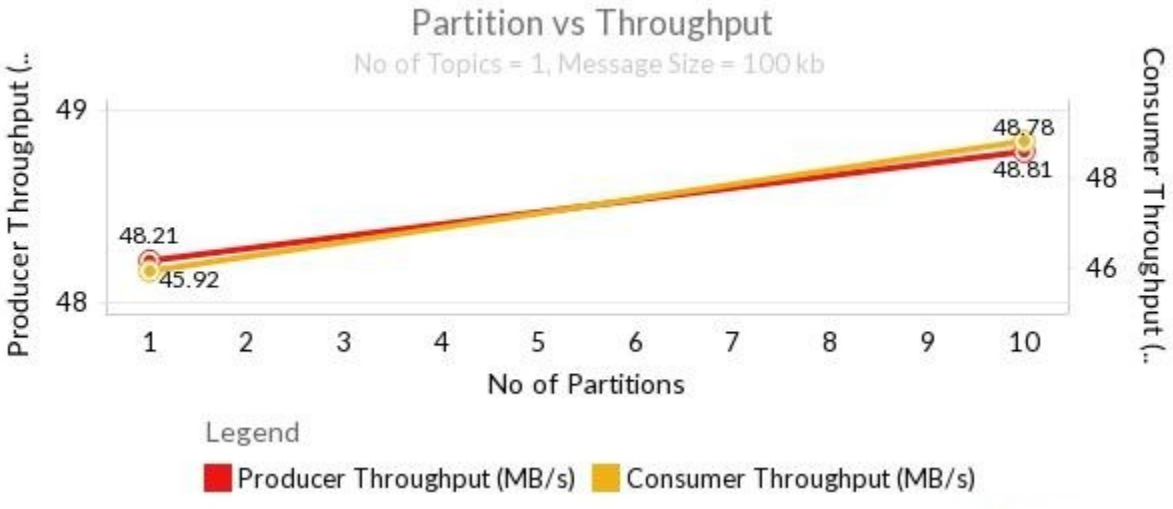
Throughput :

Test has been conducted with three brokers



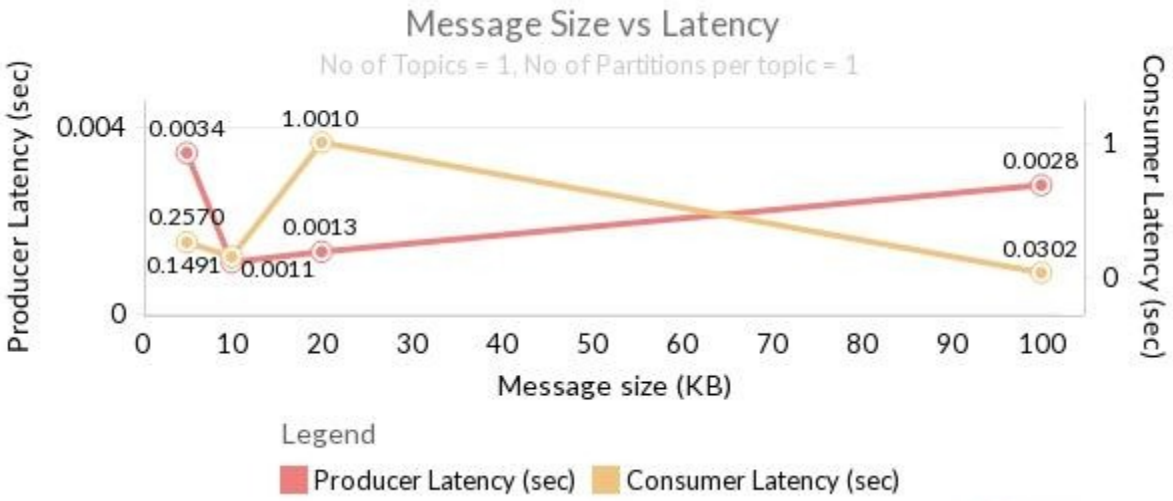


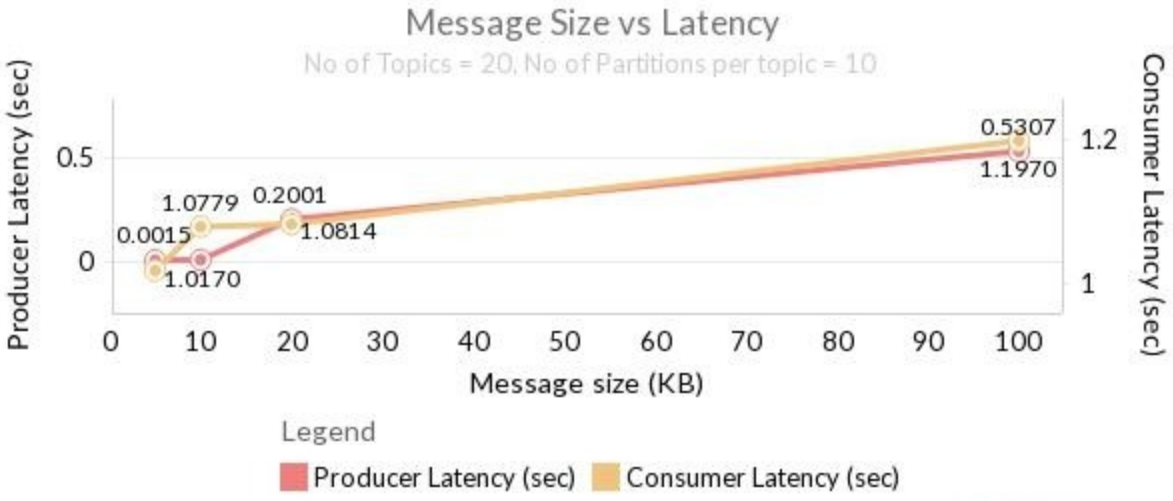

Message Size vs Records Rate
No of Topics=20, No of partitions per topic =10
Records Send Rate
Records Consumed Rat..
5000
0
4000
0
5251.11
5446.91
5251.11
5405.86
1058.91
0 10 20 30 40 50 60 70 80 90 100
Message size (KB)
Legend
Records Send Rate (99.9%)
Records Consumed Rate (99.9%)

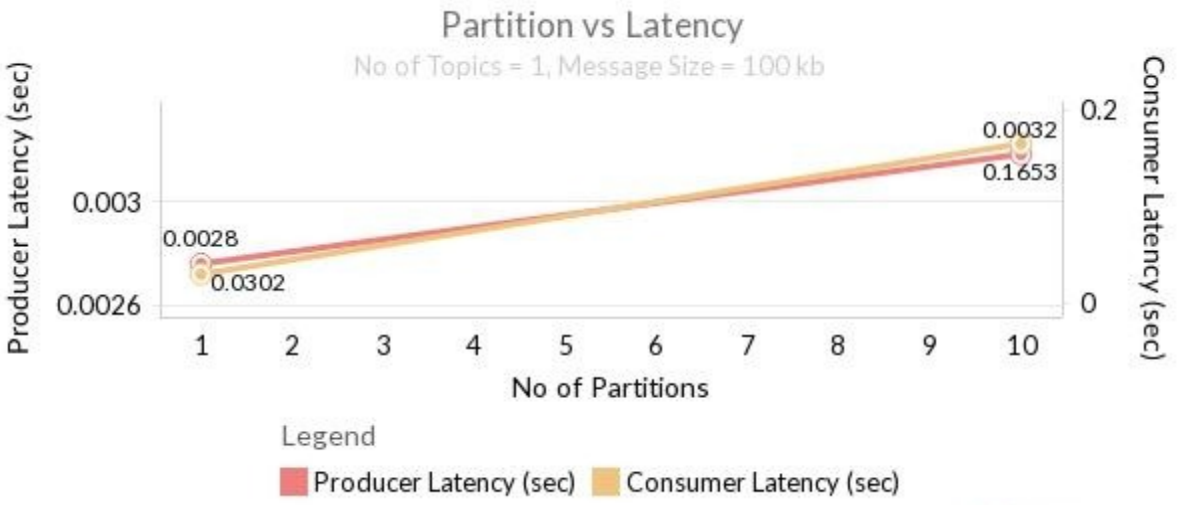

Partition vs Throughput
No of Topics = 1, Message Size = 100 kb
Producer Throughput (..
Consumer Throughput (..
49
48
48
46
48.78
48.81
48.21
45.92
1 2 3 4 5 6 7 8 9 10
No of Partitions
Legend
Producer Throughput (MB/s)
Consumer Throughput (MB/s)

Latency:

Test has been conducted with three brokers

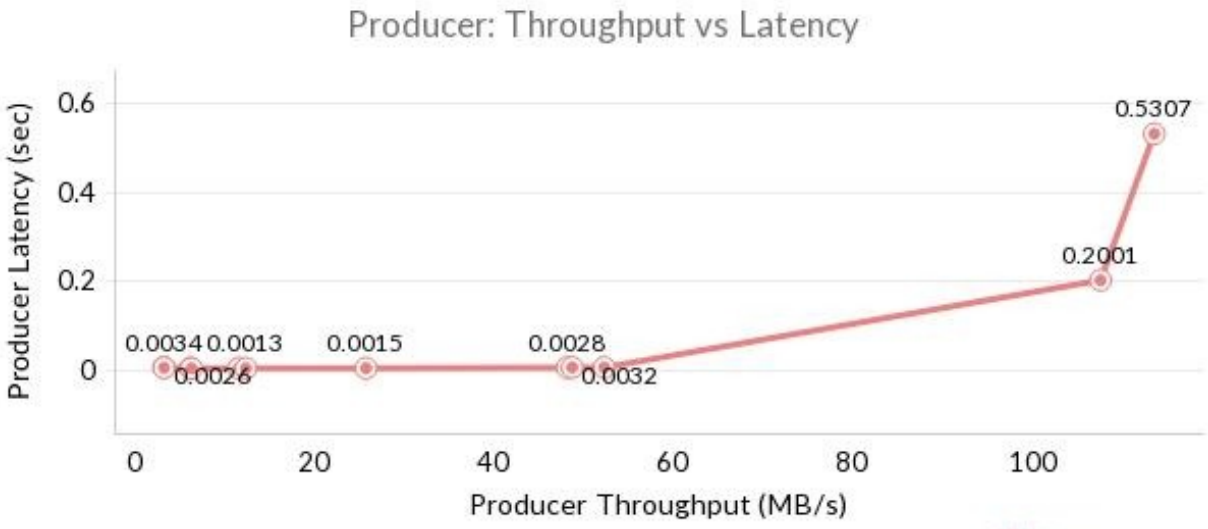


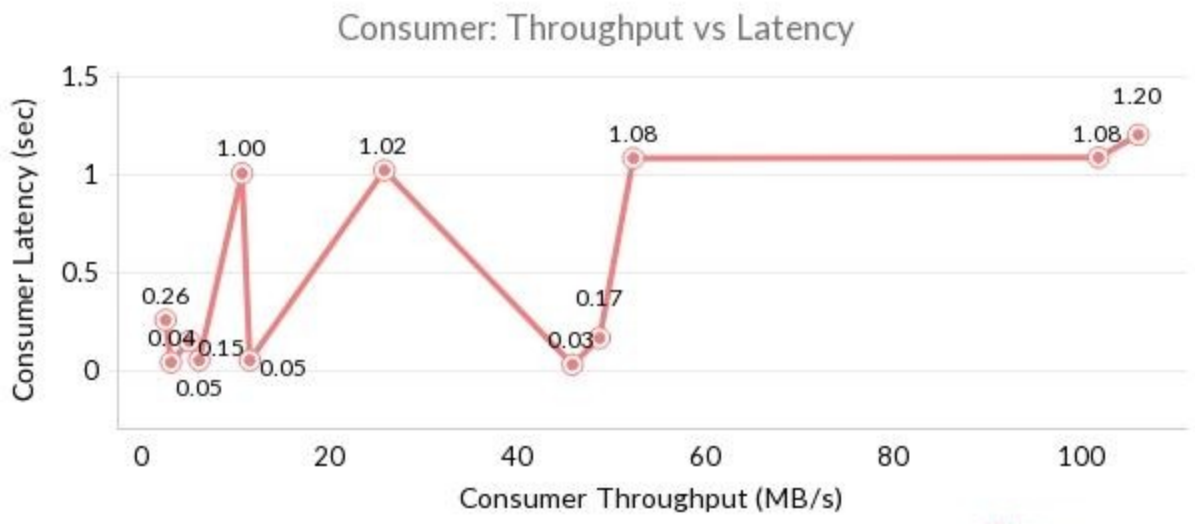




Throughput vs Latency

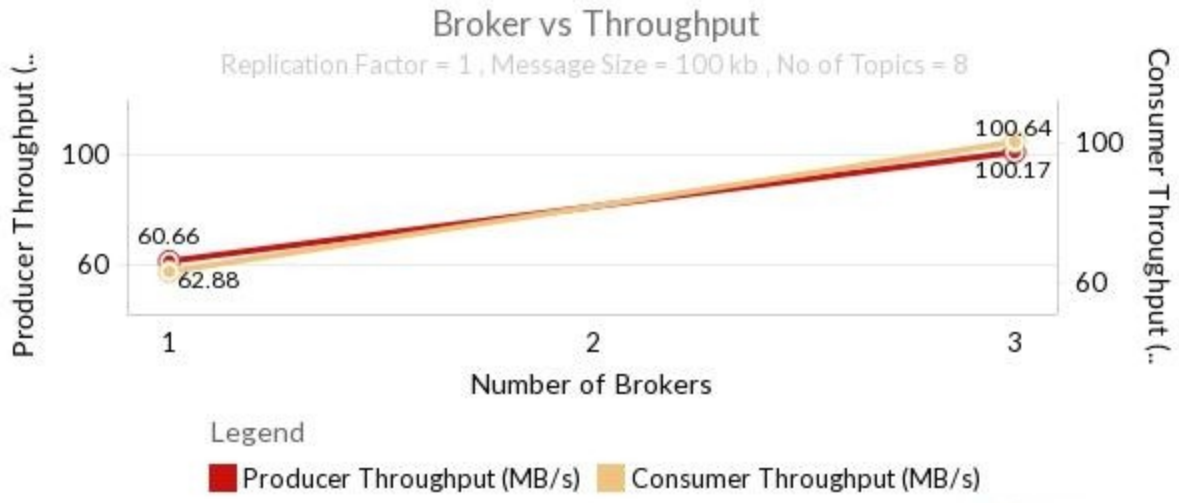
Test has been conducted with three brokers





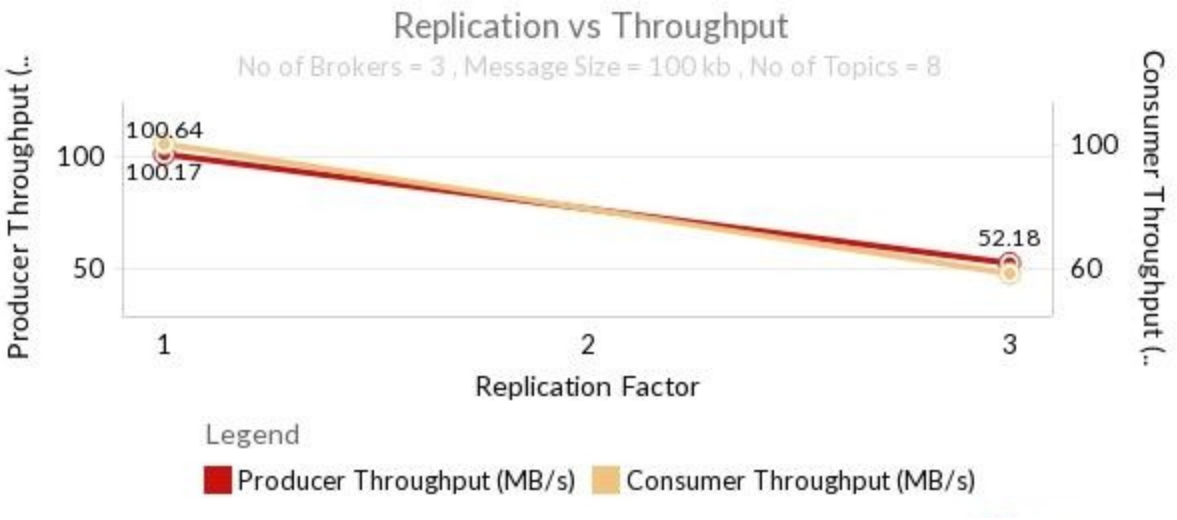
Broker Vs Throughput :

Consolidated Test Report :
https://sheet.zoho.com/sheet/open/4rd5s258426f425b744e09e9aacf677af060e/sheets/ConolidatedReport-Broker/ranges/A1



Replication Vs Throughput :

Consolidated Test Report :
https://sheet.zoho.com/sheet/open/4rd5s258426f425b744e09e9aacf677af060e/sheets/ConolidatedReport-Broker/ranges/A1



Client LoadTest outcome :

a. Message size is directly proportional to Throughput.
b. Message size is inversely proportional to Record Rate.
c. Message size is directly proportional to Latency.
d. Throughput is directly proportional to Latency.
e. No of Brokers is directly proportional to Throughput, with Replication Factor = 1.
f. Replication Factor is inversely proportional to Throughput, for the constant number of brokers.

Number of machines based on Network Utilization (NN)

Lets say,

MS - Message size
MPS - Message Produced per sec
IBR - Incoming Byte Rate
OBR - Outgoing Byte Rate
RF - Replication Factor
NT - Network Throughput = 1 Gbps
NUP - Network Utilization in percent = 80%

IBR = MPS * MS

((IBR + OBR) * NP * RF) / (NUP * NT)

To have a smooth data transfer for both producers and consumers,

IBR = OBR

NN = (2 * IBR * NP * RF) / (NUP * NT)

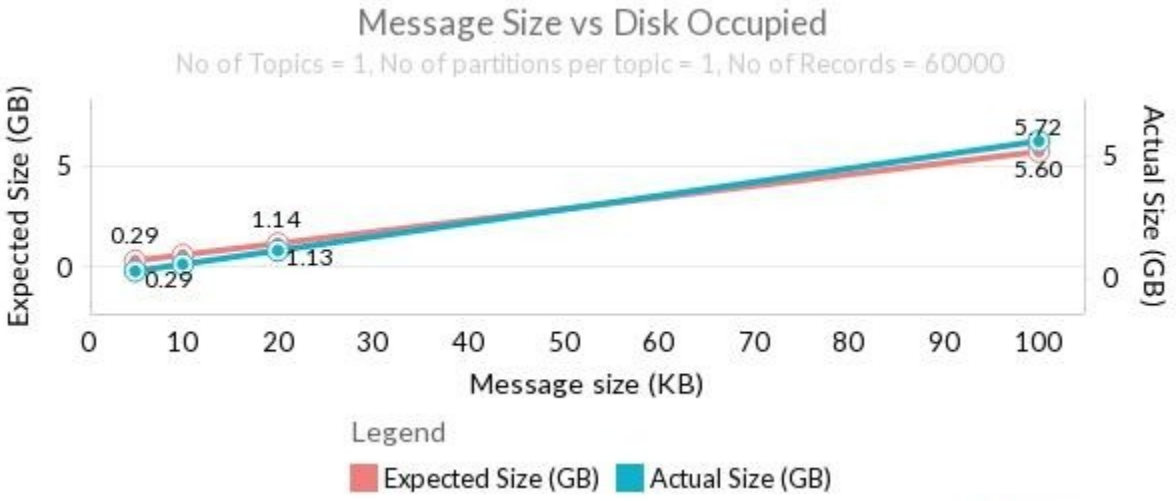
Disk Capacity observation :

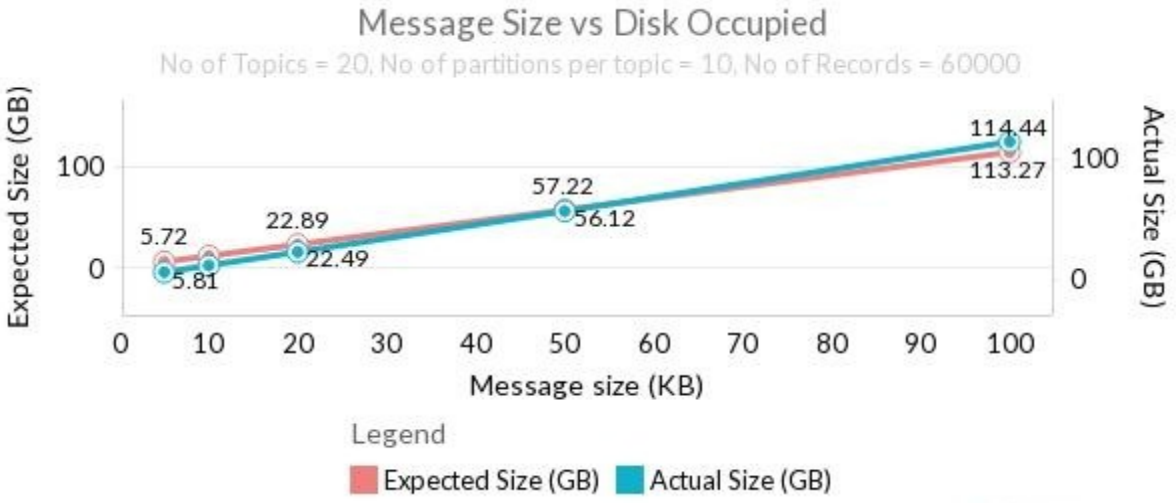
Test has been conducted with a single broker machine

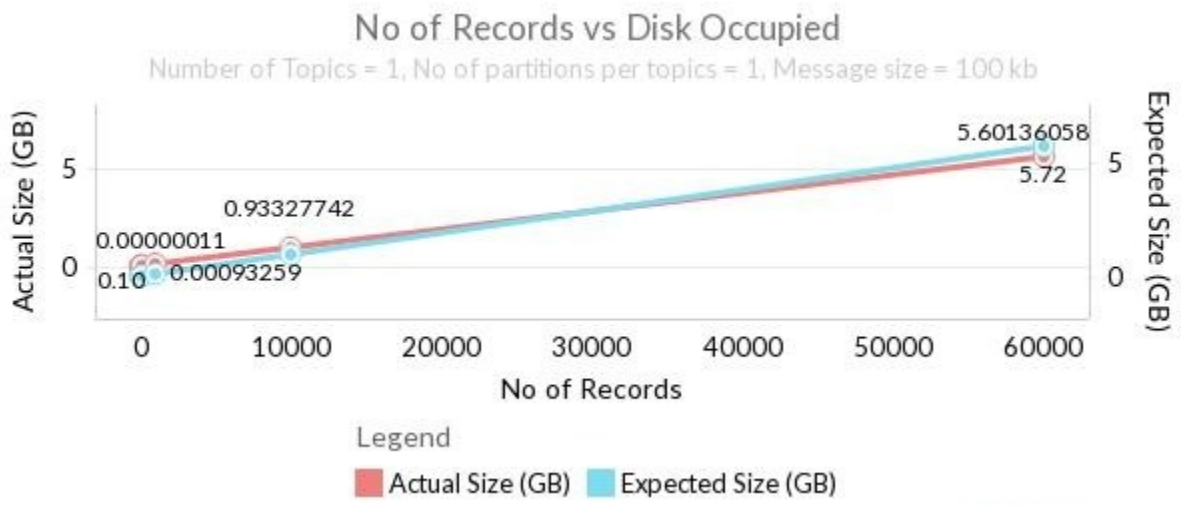
Consolidated Test Results :
https://sheet.zoho.com/sheet/open/4rd5s258426f425b744e09e9aacf677af060e/sheets/Report-Disk/ranges/G5

Command used to calculate kafka's data folder : **du -b data/**

Report : -


Message Size vs Disk Occupied
No of Topics = 1, No of partitions per topic = 1, No of Records = 60000
Expected Size (GB)
Actual Size (GB)
5
0
0.29
0.29
1.14
1.13
5.72
5.60
0 10 20 30 40 50 60 70 80 90 100
Message size (KB)
Legend
Expected Size (GB)
Actual Size (GB)


Message Size vs Disk Occupied
No of Topics = 20, No of partitions per topic = 10, No of Records = 60000
Expected Size (GB)
Actual Size (GB)
100
0
5.72
5.81
22.89
22.49
57.22
56.12
114.44
113.27
0 10 20 30 40 50 60 70 80 90 100
Message size (KB)
Legend
Expected Size (GB)
Actual Size (GB)



Number of machines based on Disk Utilization (ND)

Lets say,

MS - Message size
MPD - Message per Day
RP - Retention Period
RF - Replication Factor
DC - Disk Capacity per machine = 1 TB
DUP - Disk Utilization in percent = 80%

ND = (MS * MPD * RF * RP) / (DUP * DC)

Conclusion :

Number of machines = MAX (NN , ND)