

## **Kafka Capacity Calculation Report :**

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

### **Test Machine Spec**

Disk Capacity : 1 TB

NIC : 1000 Mbps (125 MBps)

RAM : 32 GB

Processors : 32

### **Factors taken for test :**

1. Message Size
2. Topic count
3. Partition Count
4. Replication Factor
5. Number of Brokers

### **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

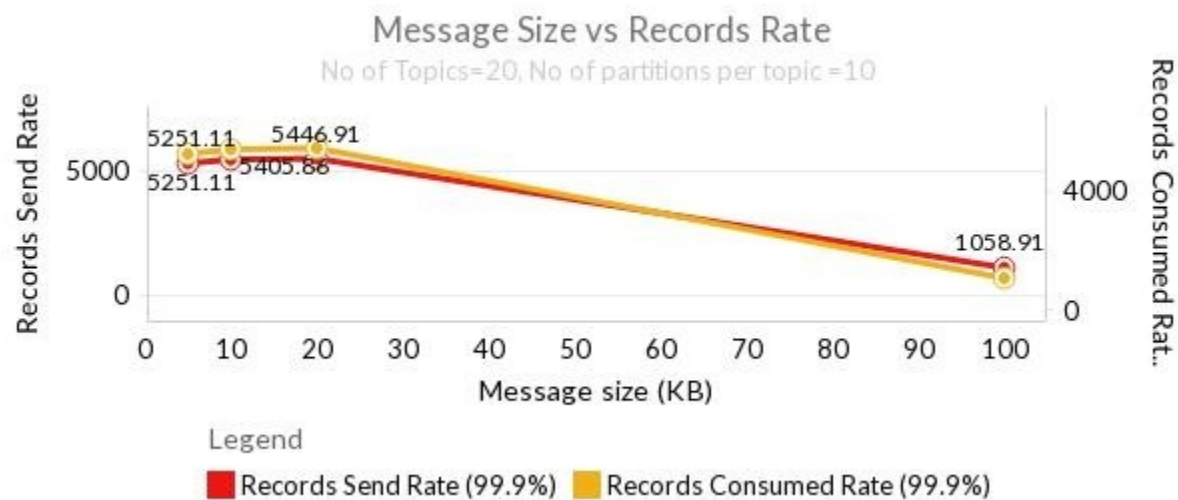
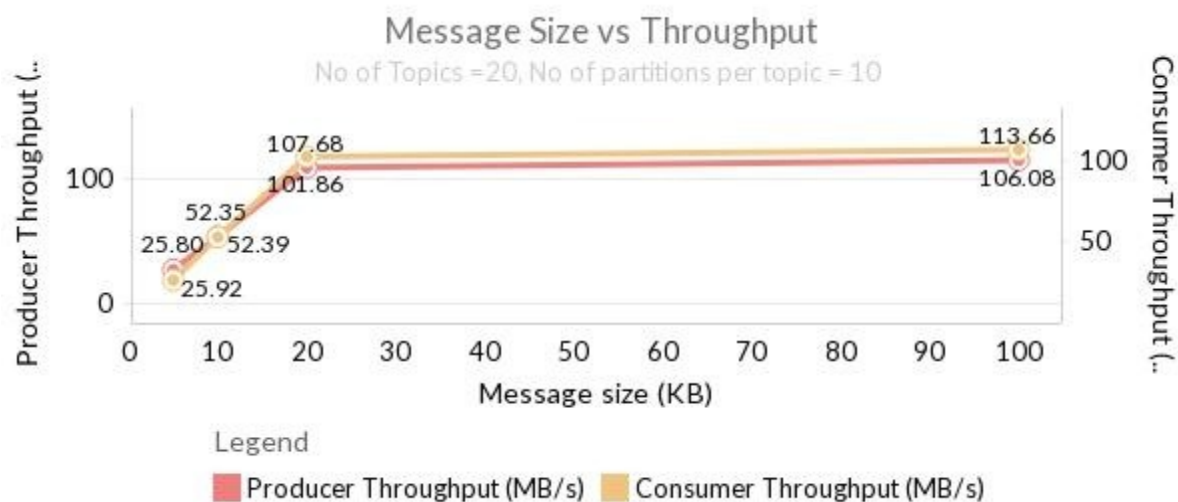
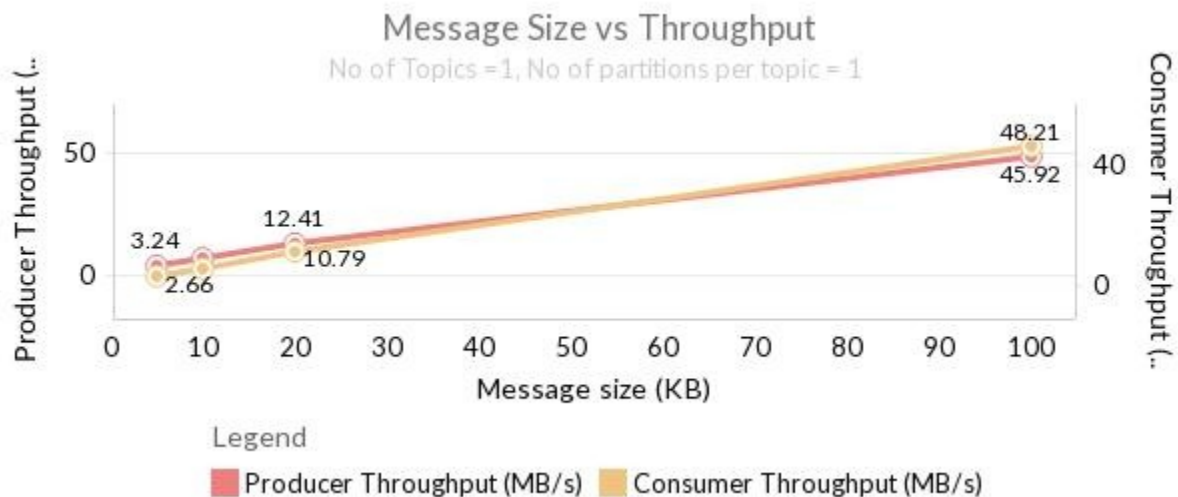
The results which are marked here are 99.9th percentile.

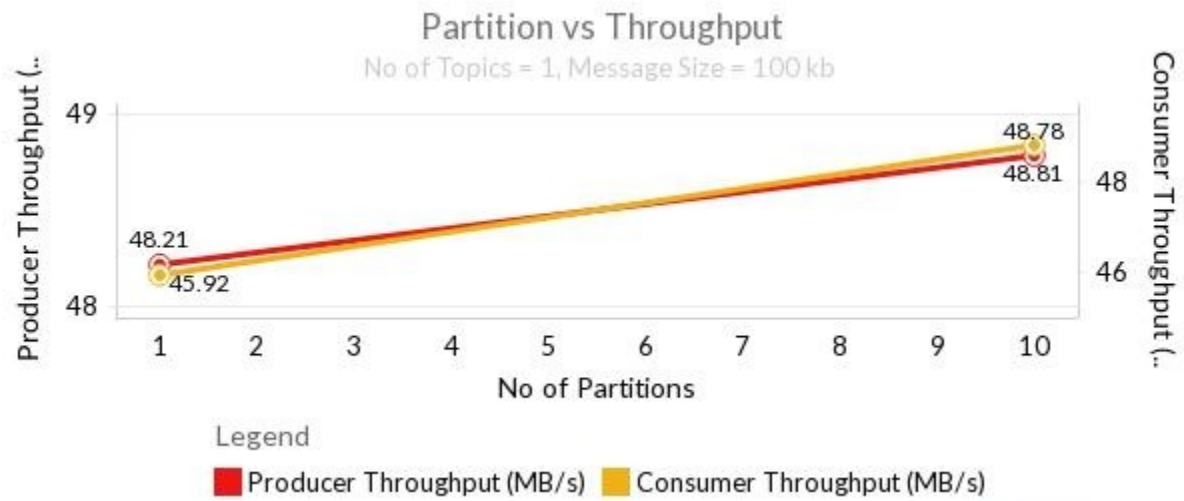
### **Network Utilisation :**

1. Throughput
2. Latency

### **Throughput :**

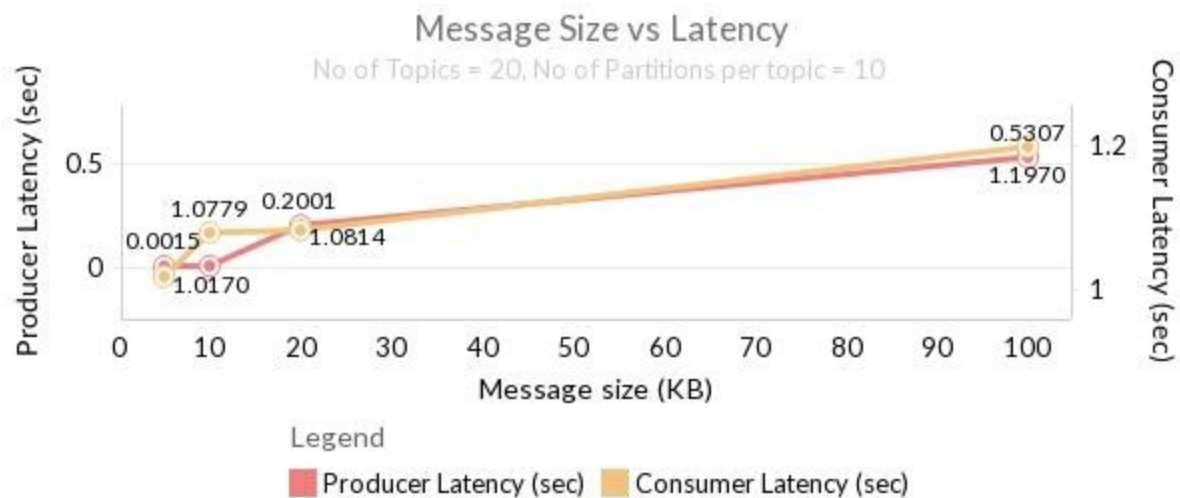
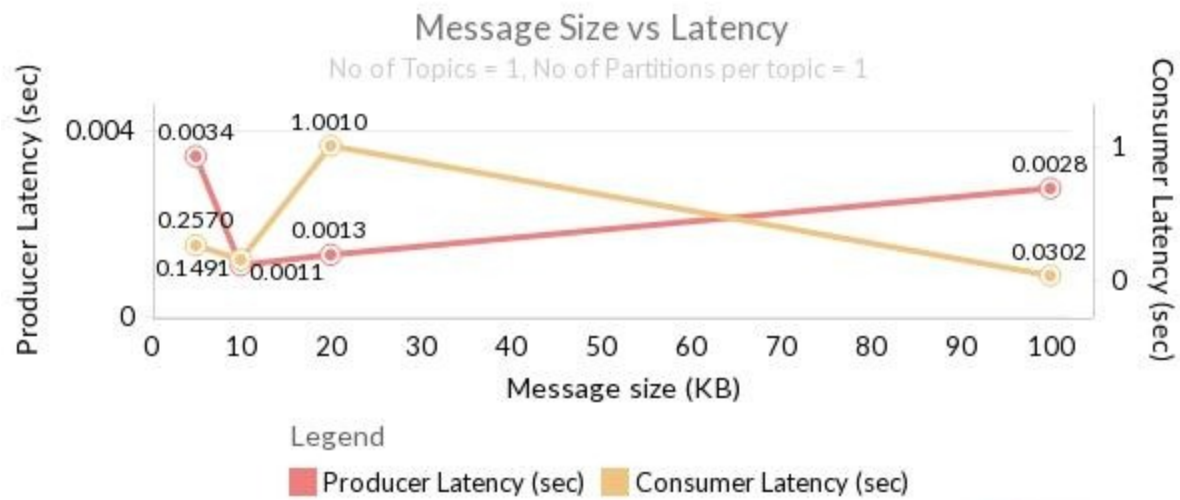
Test has been conducted with three brokers

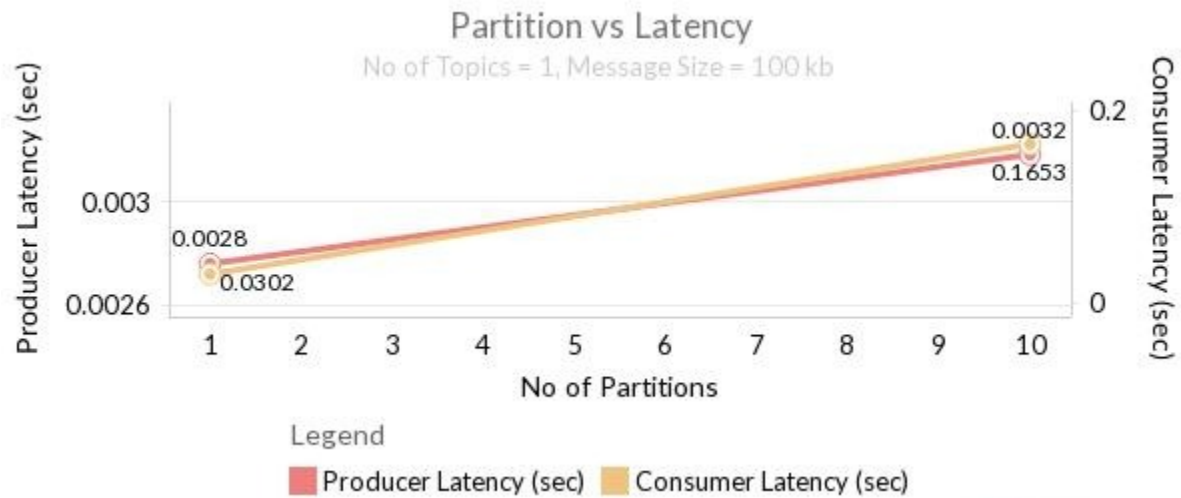




### Latency:

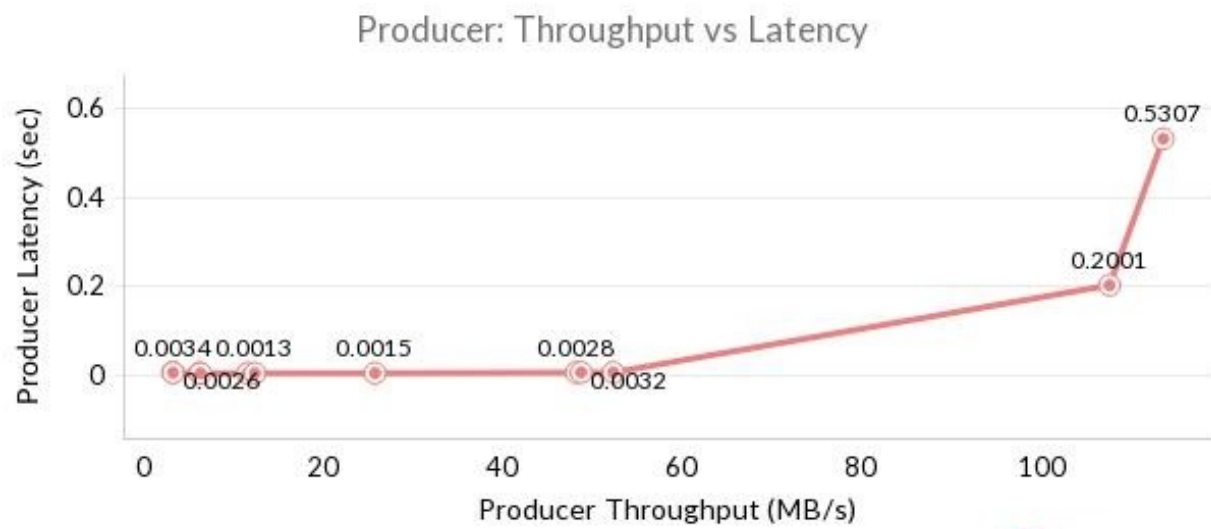
Test has been conducted with three brokers

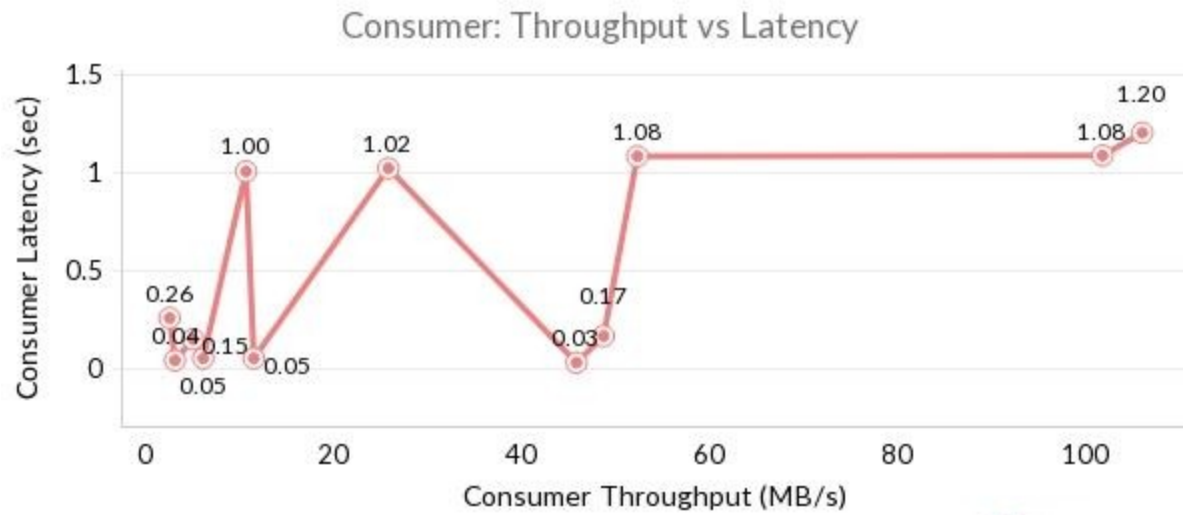




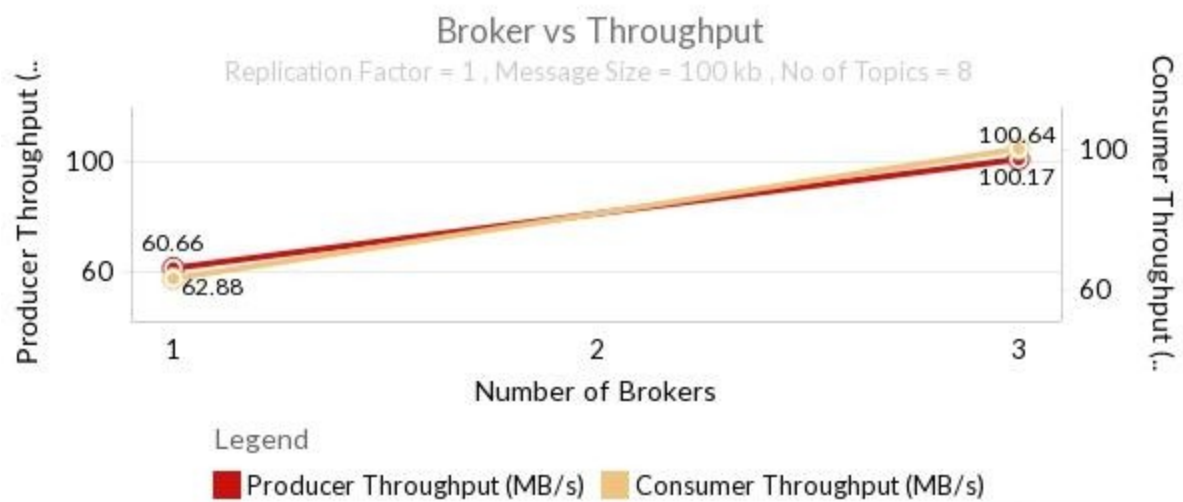
## Throughput vs Latency

Test has been conducted with three brokers

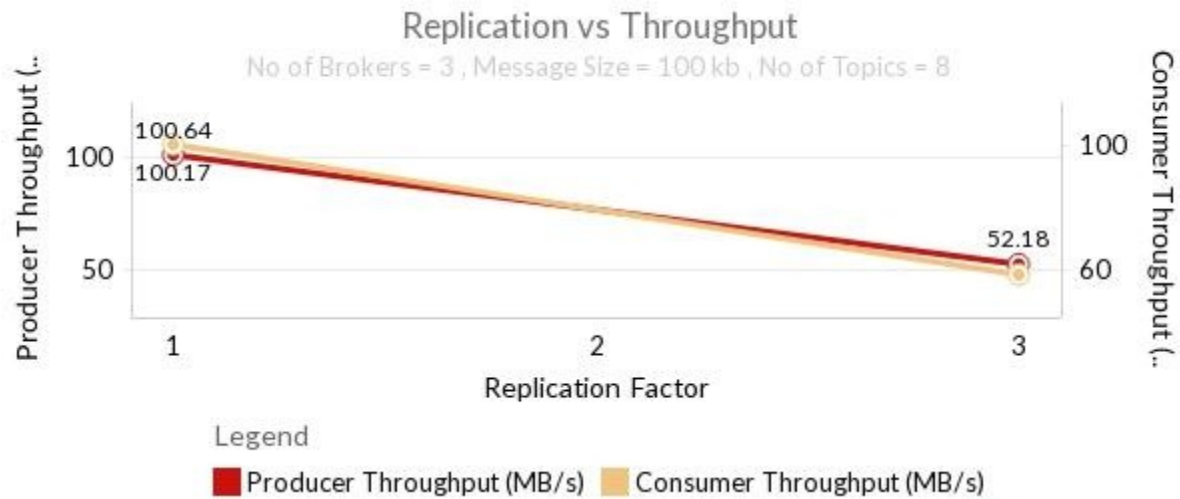




### Broker Vs Throughput :



## Replication Vs Throughput :



## Client LoadTest outcome :

- Message size  $\propto$  Throughput.
- Message size  $\propto$  (1 / Record Rate) .
- Message size  $\propto$  Latency.
- Throughput  $\propto$  Latency.
- No of Brokers  $\propto$  Throughput, with Replication Factor = 1.
- Replication Factor  $\propto$  (1 / Throughput) , where Replication factor = number of brokers.

## **Number of machines based on Network Utilization (NN)**

Lets say,

MS - Message size in bytes

MD - Size of Metadata per message = 80 bytes

MPS - Message Produced per sec

IBR - Incoming Byte Rate

OBR - Outgoing Byte Rate

RF - Replication Factor

NT - Network Throughput = 1 Gbps

DTT - Disk Transfer Throughput = 3 Gbps

NUP - Network Utilization in percent = 80%

$$\text{IBR} = \text{MPS} * (\text{MS} + \text{MD})$$

$$((\text{IBR} + \text{OBR}) * \text{NP} * \text{RF}) / (\text{NUP} * \text{MIN}(\text{NT}, \text{DTT}))$$

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

$$\text{IBR} = \text{OBR},$$

$$\text{NN} = (2 * \text{IBR} * \text{NP} * \text{RF}) / (\text{NUP} * \text{MIN}(\text{NT}, \text{DTT}))$$

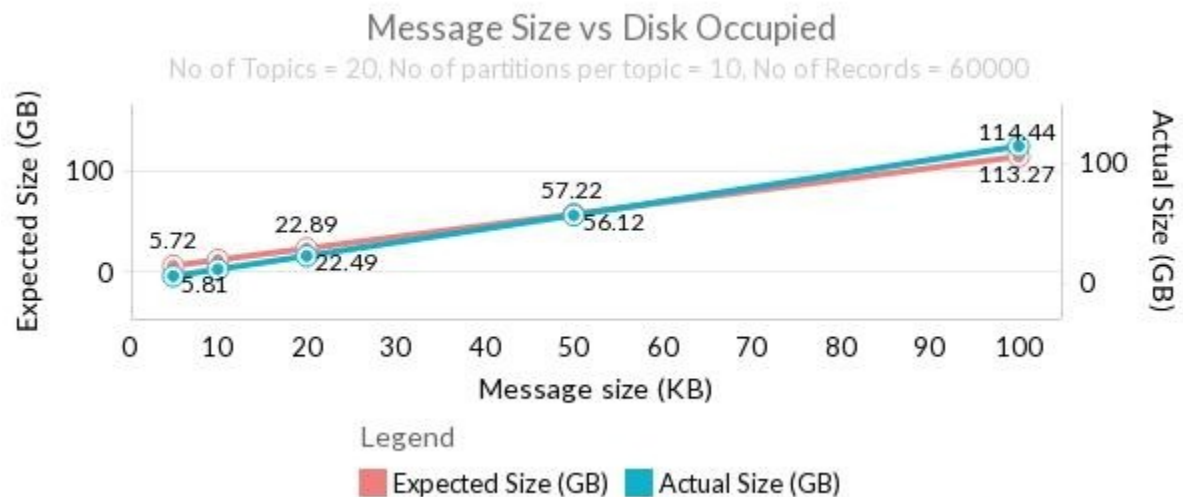
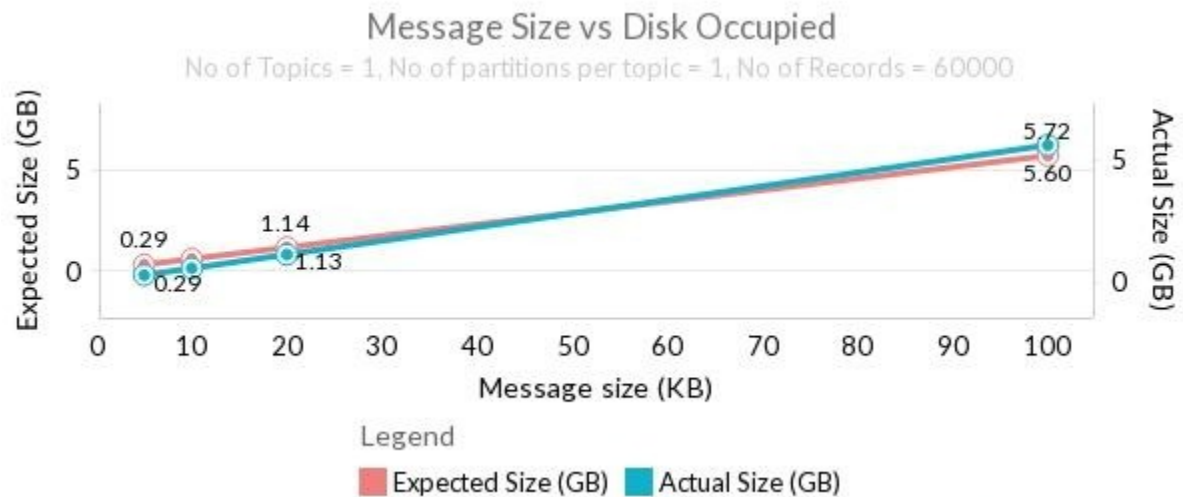


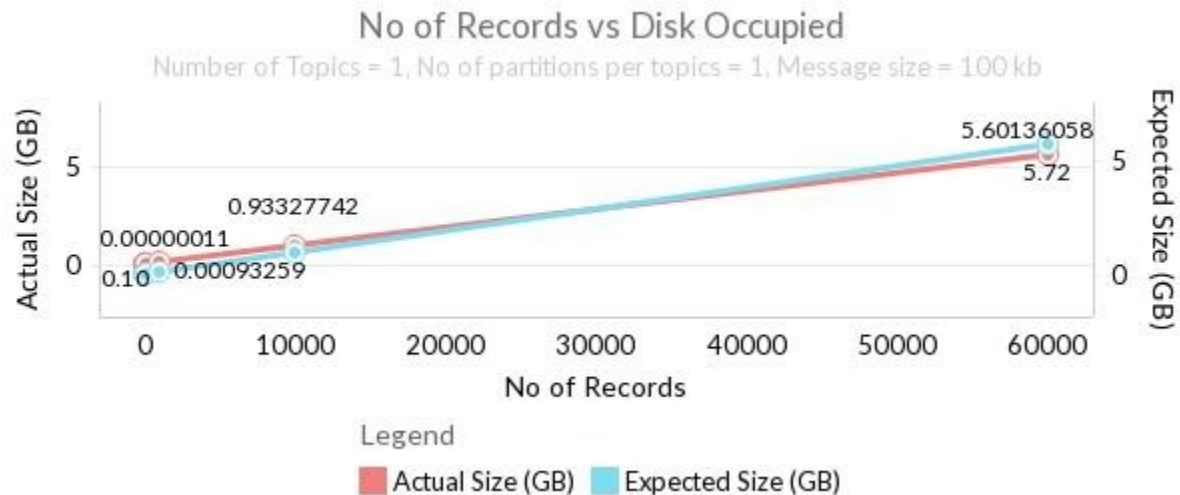
## Disk Capacity observation :

Test has been conducted with a single broker machine

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

Report :-





### Number of machines based on Disk Utilization (ND)

Lets say,

MS - Message size in bytes

MD - Size of Metadata per message = 80 bytes

MPD - Message per Day

RP - Retention Period

RF - Replication Factor

DC - Disk Capacity per machine = 1 TB

DA - Disk Actual Capacity per machine = 865 GB

BS - Kafka build size including logs,backup,working directory.

DUP - Disk Utilization in percent = 80% of DA

$$ND = ((MS + MD) * MPD * RF * RP) / ((DUP - BS) * DA)$$

### Conclusion :

$$\text{Number of machines} = \text{MAX} (NN, ND)$$