perfmatrix.com

Latency, Bandwidth, Throughput and Response Time - Practical Example I

PerfMatrix

4-5 minutes

Latency, Throughput, Bandwidth and Response Time; somehow these terms are very confusing. A new performance tester faces difficulty to understand these terms without example. This article will help to get the knowledge of Network Latency, Network Bandwidth, Data Throughput and Response Time. So, let's start now,

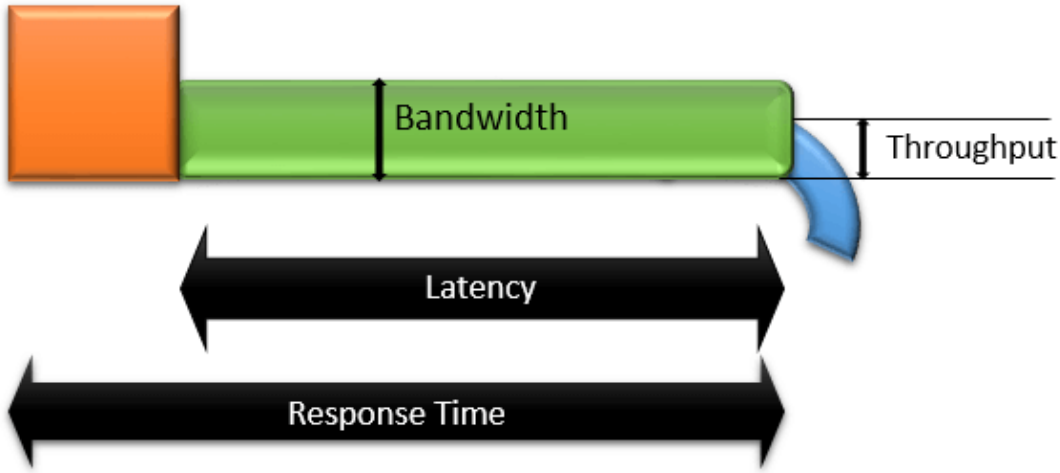

Figure 01

Look at the above figure (Figure 01). The illustration shows a water tank (Orange), water pipe (Green), water (Blue). The water tank represents a server, pipe represents communication channel with certain width and water represents data. You will get an idea about all four important terms i.e. Latency, Bandwidth, Throughput and

Response Time with this diagram. If still, it seems confusing then let's have a detailed discussion on it.

The time taken by water to travel from one end to another end is called Latency. It's measuring units are millisecond, second, minute or hour. In performance testing, the term latency (or better to called Network Latency) of a request is travel time from client to server and server to the client. Some testers called it 'Network Delay'. Let's say:

- A request starts at t=0
- **Reaches to a server in 1 second (at t=1)**
- The server takes 2 seconds to process (at t=3)
- **Reaches to the client end in 1.2 seconds (at t=4)**

So, the network latency will be 2.2 seconds (= 1 + 1.2).

Bandwidth:

Bandwidth shows the capacity of the pipe (communication channel). It indicates the maximum water passes through the pipe. In performance testing term the maximum amount of data that can be transferred per unit of time through a communication channel is called channel's bandwidth. Let's say an ISDN having 64Kbps of bandwidth and we can increase it by adding one more 64Kbps channel, so total bandwidth will be 128Kbps, so maximum 128Kbps data can be transferred through ISDN channel.

Throughput:

The water is flowing from the pipe can be represented as 'Throughput'. In performance testing term '*The amount of data moved successfully from one place to another in a given time period is called Data Throughput*'. It is typically measured in bits per second (bps), as in megabits per second (Mbps) or gigabits per

second (Gbps). Let's say, 20bits data transferred at t=4th second, so throughput at t=4 is 20bps.

Remember: *Data Throughput can never be more than Network Bandwidth.*

Response Time:

Response time is the amount of time from the moment that a user sends a request until the time that the application indicates that the request has completed and reaches back to the user. In the Latency example, Response time will be 4 seconds. For more information on response time refer to the link.

Some important points to be remembered:

- Solving bandwidth is easier than solving latency.
- If throughput is nearly equal to bandwidth, it means the full capacity of the network is being utilized which may lead network bandwidth issue.
- Increase in response time with flat throughput graph shows a network bandwidth issue. This bottleneck can be rectified by adding extra channels i.e. by increasing network bandwidth.
- Ideally, consistent throughput indicates an expected capacity of network bandwidth.
- Some tools do not express the throughput in units per unit of time but in clock periods. This is incorrect but commonly used because of convenience.
- Ideally, response time is directly proportional to throughput during the user ramp-up period. If throughput decreases with an increase in response time then it indicates instability of application/system.
- Ideally, response time and throughput should be constant during

steady state. A less deviation in both the terms indicates the stability of the application.

- The Number of threads is directly proportional to the throughput.
- If you have low latency and small bandwidth then it will take a longer time for data to travel from point A to point B compared to a connection that has low latency and high bandwidth.
- Latency is affected by connection type, distance and network congestion.

Network Bandwidth Network Latency Performance Testing Basics Response Time Throughput Graph

- by
- on January 17, 2019

2