**SCALABILITY AND LOAD TESTING EVALUATION**

**VERIFACTS SERVICES PRIVATE LIMITED**

# Scalability Testing Attributes:

Response Time: Response time is the time consumed between the user’s request and the application’s response. Response time may increase or decrease based on different user load on the application. Basically, the response time of an application decreases as the user load increases. Application having the lesser response time is considered as the higher performance application.

Throughput: Throughput is the measurement of number of requests processed in a unit time by the application. It differs from one application to another as in web application it is measured in number of user requests processed in a unit time whereas in database application it is measured in number of queries processed in a unit time.

Performance measurement with number of users: Depending on the application type, it is always tested for the number of users that it can support without its breakdown or busy standby situation.

Threshold load: Threshold load is the number of requests or transactions the application can process with desired throughput.

CPU Usage: CPU Usage is the measurement of the CPU utilization while executing application code instructions. It is basically measured in terms of the unit Megahertz.

Memory Usage: Memory usage is the measurement of the memory consumed for performing a task by an application. It is basically measured in terms of the unit bytes.

Network Usage: Network usage is the measurement of the bandwidth consumed by an application under test. It is measured in terms of bytes received per second, frames received per second, segments received and sent per second etc.

# Load Testing Attributes:

During the execution of load tests, it is essential to look at the following metrics, as they help you know how accurately load testing takes place under different test scenarios.

Average Response Time – It is the average taken by the application to respond to user requests. In other words, it is the application’s speed in responding to user queries or requests.   
Error Rate – It is expressed in percentage. An error rate is the number of errors occurred during the user requests to the total number of user requests. A request could be anything, such as requesting images, text, web pages, documents, or any other resources. Errors during requests occur because the application is not capable of handling more requests at a specific time.   
Requests Per Second – This metric determines the number of requests the application’s server receives in one second. 

Throughput – In load testing, throughput refers to the number of transactions per second, which gauges the number of requests your application receives in a second.   
Concurrent Users – This metric determines the count of users using the application at a specific point in time or at any time. It helps you understand the time at which the maximum number of users use your app.   
Peak Response Time – The peak response time measures the time required for a request-response cycle. Rather than taking the average, it primarily concentrates on the longest cycle.

# Scalability and load testing tools:

|  |  |
| --- | --- |
| **Scalability Testing** | Gatling |
| **Load Testing** | Jmeter |