

Load Balancers

Let us suppose you are working for an IT Company and experience that the company doesn't have any management system. Some employees do not have any tasks, and some are overburdened with responsibilities and functions. You discuss the issue with the company's CEO, and he appoints a manager to your team. The manager now considers all aspects like capabilities, availability, additional responsibilities, etc., and assigns work. The team is now better managed, and no employee faces either shortage of work or feels overburdened. The team is now a better place to work mentally and physically. We can say that the manager worked towards load balancing and can be described as a load balancer.

Let's understand more about load balancing.

What is load balancing?

Load Balancing is the process of efficient distribution of application or network traffic across all nodes in a distributed system.

The devices used for ensuring load balancing are load balancers.

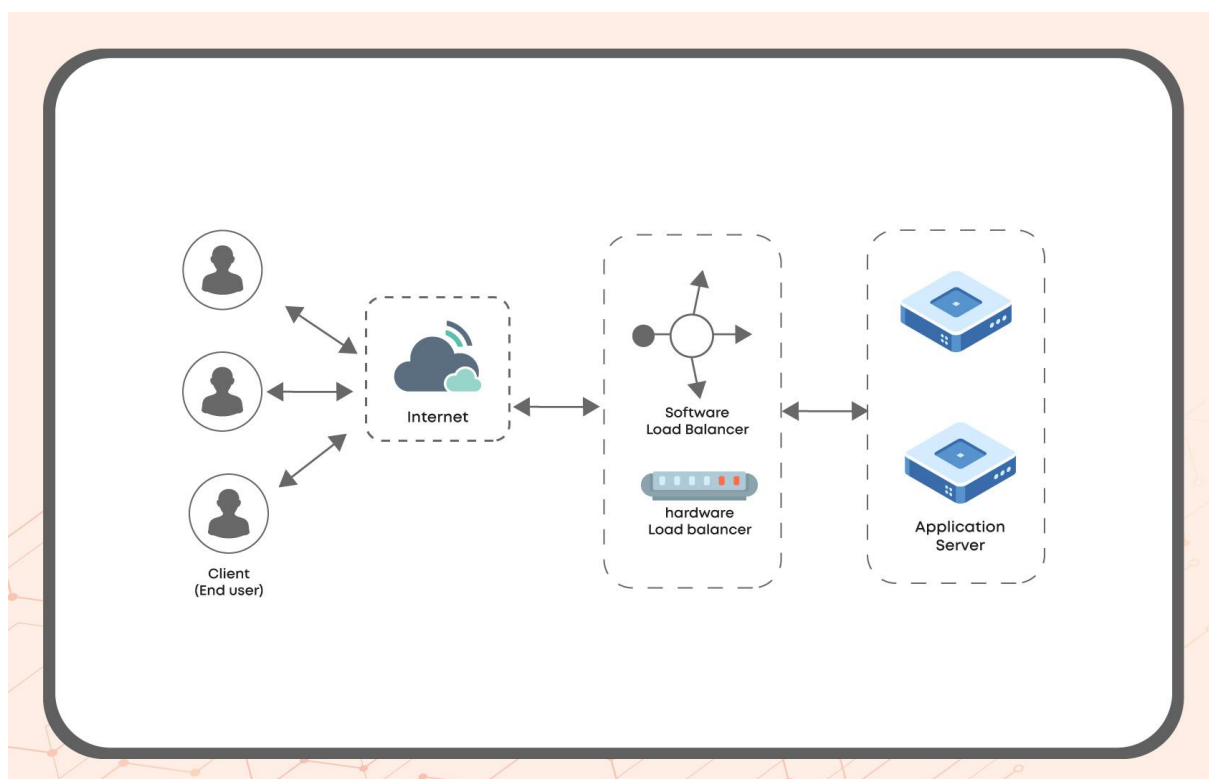


Fig: Load Balancing

Roles of Load Balancer

1. The work or load distribution is equal over every node.
2. Should keep a check on which nodes are not operational or available.
3. Efficiently distribute/manage work so that it is completed before time.
4. Distribution should be such that the total capacity is utilised and speed is maximised.
5. Load balancers should ensure high scalability, high throughput and high availability(if the node is not operational, the request is passed to another node that is up and running).

Where should the load balancer be placed?

Load balancers are placed where it has the capacity to terminate connections to the public IPs. Load balancers can be placed to manage multiple ICs(Industrial Control Systems). For example, the three layers of a system: FrontEnd, BackEnd and DataBase consist of various servers. The load balancer can be placed at multiple places like: in front of the Front-end , Back-end and data storage layers. Load Balancers helps to manage multiple instances and supports horizontal scaling.

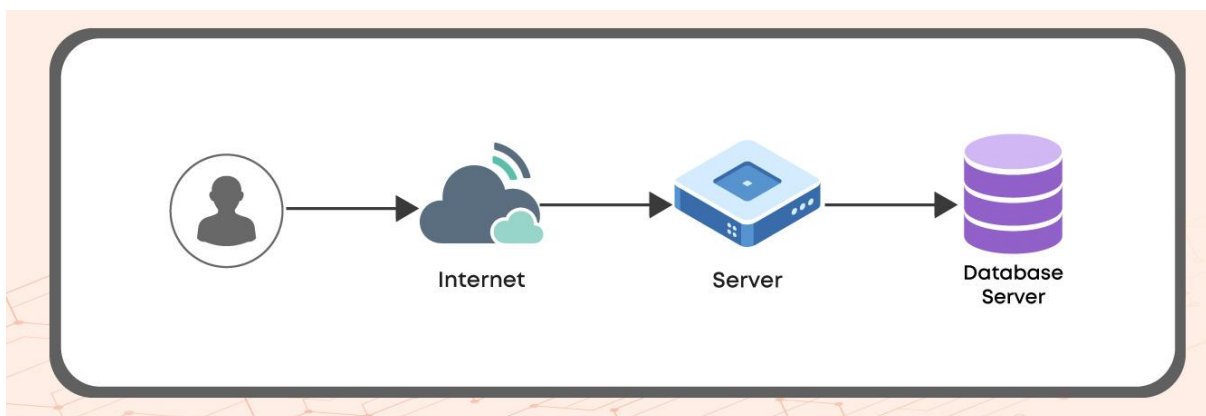


Fig: Load Balancers

Benefits of Load Balancing

1. Optimisation:

Load Balancers helps in resource utilisation and lower response time, thereby optimising the system in a high traffic environment.

2. Better User Experience:

Load balancers help in reducing the latency and increasing availability, making the user's request go smoothly and error-free.

3. Prevents Downtime:

Load Balancers maintain a record of servers that are non-operational and distribute the traffic accordingly, therefore ensuring security and preventing downtime, which also helps increase profits and productivity.

4. Flexibility:

Load Balancers have the flexibility to re-route the traffic in case of any breakdown and work on server maintenance to ensure efficiency.

5. Scalability:

When the traffic of a web application increases suddenly, load balancers can use physical or virtual servers to deliver the responses without any disruption.

6. Redundancy:

Load Balancing also provides in-build redundancy by re-routing the traffic in case of any failure.

When to use Load Balancers?

1. When the application has multiple instances or multiple servers, load balancers can be used for load management in such cases.
2. Application traffic is distributed over multiple nodes or servers.
3. In a high traffic environment, load balancers are essential for ensuring scalability, availability and latency.

Challenges of Load Balancing

1. Single Point Of Failure:

During a load balancer failure or breakdown, the complete system is interrupted and will be unavailable for the time periods leading to a bad user experience. During a load balancer malfunctioning, the communication between clients and servers would be broken. To solve this issue, we can use redundancy. The system can have an active load balancer and one passive load

balancer. In case the active load balancer goes down, the passive can become the active load balancer then.