LOAD BALANCING IN CLOUD

SABARNIKA S (21CSR168)

3RD DEPT. OF CSE

01

PROBLEEM

SERVER OVERLOADING

THE SILENT KILLER

SERVER OVERLOAD

- An issue that occurs due to conditions that exhaust server resources, following which the server fails to handle the requests

CAUSES

- Unexpected User Traffic spikes
- Issues with servers
- Malwares
- Denial-Of-Service(DOS) attacks
- Network slowdown

!?

Research shows that 37% of users abandon websites if loading pages take too long

02

SOLUTION

LOAD BALANCING

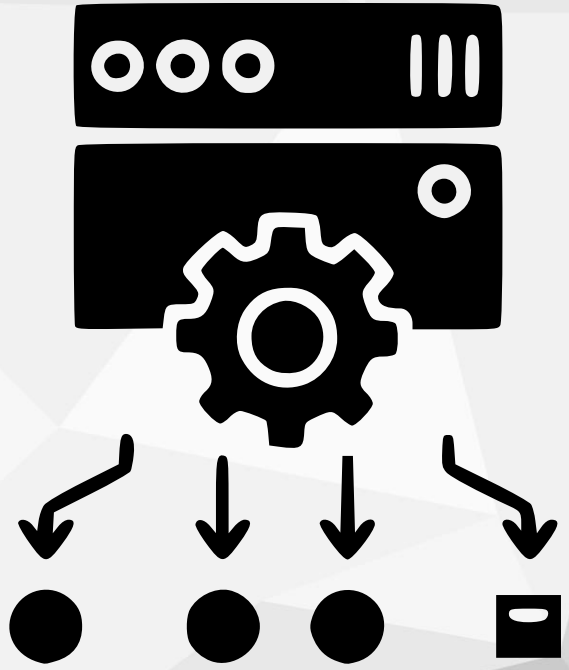
"One way to overcome server overloading is by managing the Unexpected User Traffic spikes which is achieved by the concept of LOAD BALANCING."

LOAD BALANCING

- A process of distributing the workloads across computing resources in a cloud computing environment and carefully balancing the network traffic accessing these resources

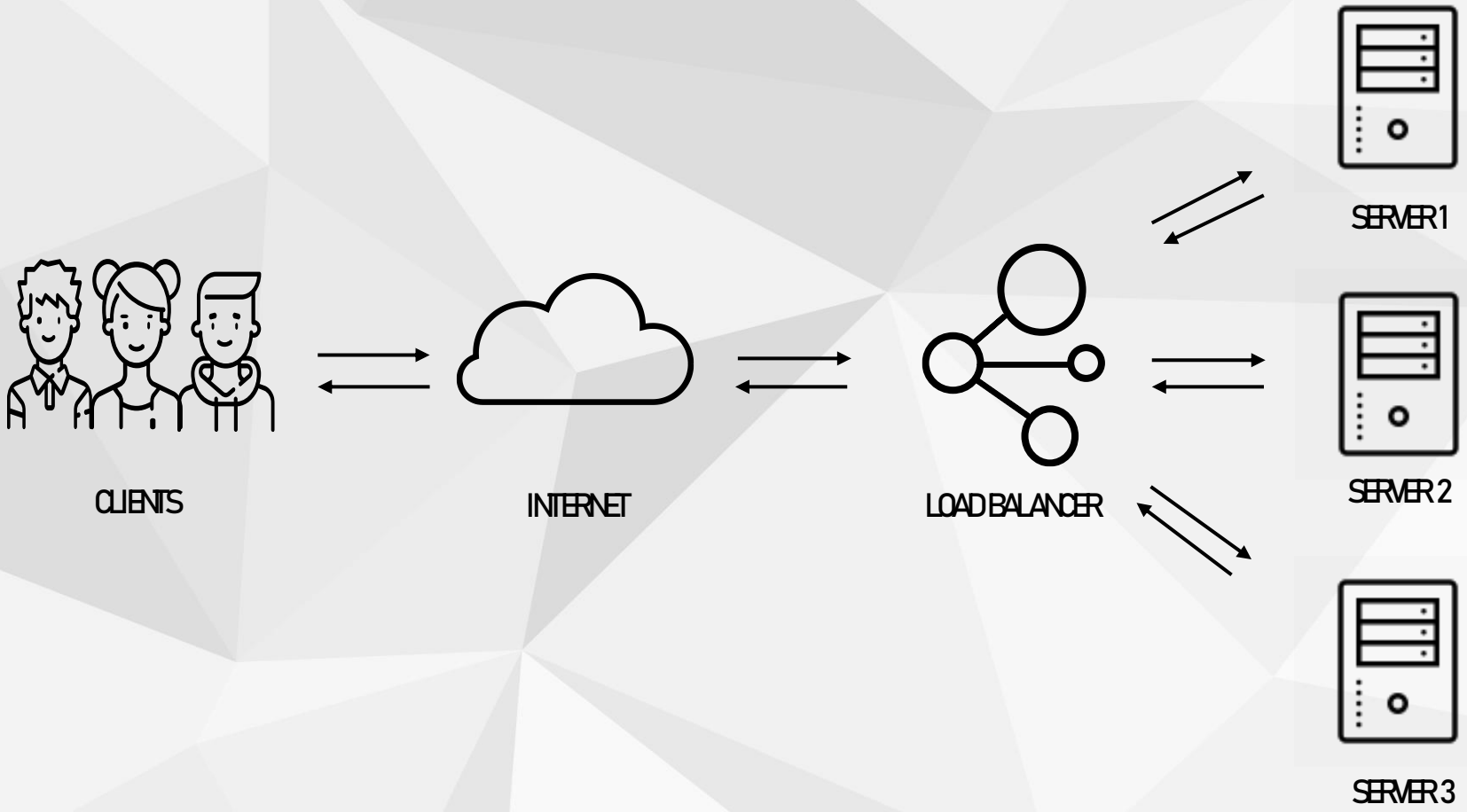
LOAD BALANCER

A device that sits between the user and the server group and acts as an invisible facilitator, ensuring that all resource servers are used equally



03 WORKING

HOW DOES THIS WORK !



WORKING OF LOAD BALANCING

HOW IT WORKS?

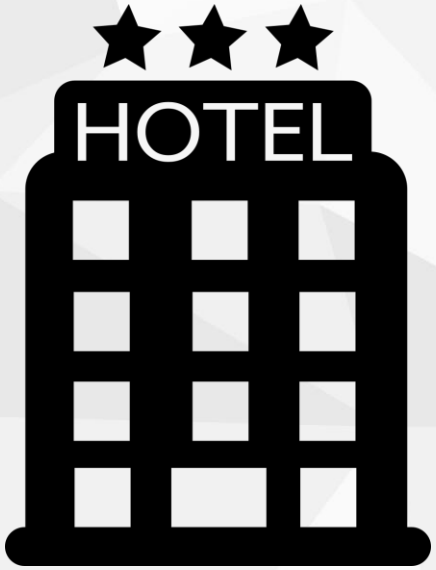
BASIC STEPS INVOLVED

1. A client gets a request and tries to connect with server
2. A load balancer receives the request and based on the preset pattern of the ALGORITHMS, it routes the request to one of the Server Group (or Farm).
3. The server receives the connection request and responds to the client via load balancer

Companies usually have their applications running on multiple servers. Such server arrangement is called a Server Farm or Group

ANALOGY

MANAGER IN A RESTAURANT...



Manger – Load Balancer
Waiters – Servers
Customers - Clients

- Consider a restaurant with 5 waiters
- If customers were allowed to choose their waiters, one or two waiters could be overloaded with work while others are idle
- To avoid this, Manager assigns customers to specific waiters who are better suited to serve them

04

ALGORITHMS

AND TYPES

"A load balancer follows a set of rules called **Load Balancing Algorithms** to determine the best server for each of the different client requests."

ALGORITHMS AND TYPES

STATIC LOAD BALANCING

- Follows a fixed rules and are independent of current server state
- Ex:

  Round Robin method

  IDHash method

  Weighted Round Robin method

DYNAMIC LOAD BALANCING

- Examine the current state of the servers before distributing traffic
- Ex:

  Least Connection method

  Weighted Least Connection method

  Least Response Time method

05

BENEFITS

- Application availability
- Application scalability
- Application security
- Application performance

AVAILABILITY

- Increase fault tolerance of our system by redirecting client traffic to available resources

SECURITY

- Load balancers come with built-in security features to add another layer of security to our internet applications

SCALABILITY

- Our applications can handle thousands of client requests

PERFORMANCE

- Improve application performance by increasing response time and reducing network latency

THANK YOU