

**Cloud Computing Architecture Lab**

**Name:** Pratham Kandari

**Sap Id:** 500097663

**Batch:** 7

**Experiment -4**

Application Load Balancer in AWS

Q:1. Enumerate and differentiate different Load balancers in AWS?

Ans:

1. Application Load Balancer: An application load balancer can route requests to one or more ports on each container instance in your cluster, enables path-based routing, and makes routing decisions at the application layer (HTTP/HTTPS). Dynamic host port mapping is supported by application load balancers.
2. Network Load Balancer: The transport layer (TCP/SSL) routing decisions are made by a network load balancer. Millions of queries can be handled per second. Using a flow hash routing technique, the load balancer chooses a target from the target group for the default rule after receiving a connection. On the port given in the listener configuration, it tries to establish a TCP connection to the chosen target. Without changing the headers, it transmits the request. Dynamic host port mapping is supported by network load balancers.

Q:2. Explain:

1. Load Balancer: Load balancing is the method of distributing network traffic equally across a pool of resources that support an application. Modern applications must process millions of users simultaneously and return the correct text, videos, images, and other data to each user in a fast and reliable manner.

1. Target Group: Target groups use the protocol and port number that you specify to route requests to one or more registered targets, such as EC2 instances. A target can be registered with various target groups. Health checks can be set up for each target group. All targets registered to a target group that is specified in a listener rule for your load balancer are subject to health checks.

1. Rules: Your devices can communicate with AWS services thanks to rules. Based on the MQTT topic stream, rules are examined, and subsequent actions are taken.

1. Listeners: A process that monitors connection requests is called a listener. It is set up with a protocol and a port for connections from the front end (client to load balancer) and a protocol and a port for connections from the back end (load balancer to back-end instance).

1. Load Scheduling Algorithms: There are two types of Load Scheduling Algorithms:

1. Round Robin Load Balancing Method: The simplest and most popular load balancing algorithm is round-robin load balancing. Application servers receive client requests in a straightforward rotation. The first client request is sent to the first application server in the list, the second client request is sent to the second application server, the third client request is delivered to the third application server, the fourth client request is sent to the first application server, and so on.

1. Weighted Round Robin Load Balancing Method: Similar to the round-robin load balancing method, weighted round robin has the added benefit of distributing incoming client requests across the server farm in accordance with the respective capacities of each server. It works well when distributing incoming client requests among a group of servers with various capacities or resources. Based on criteria of their choosing, the administrator gives a weight to each application server that represents the relative capacity of each server in the farm to handle traffic.

1. Least Connection Load Balancing Method: A dynamic load balancing mechanism called least connection load balancing distributes client requests to the application server that has the fewest active connections at the moment the client request is received. This technique takes the active connection load into account. When application servers have identical specs, one server may be overwhelmed due to longer-lasting connections. This method works well with incoming requests that have variable connection latency and a group of servers that have similar processing capabilities and resource availability.

1. Weighted Least Connection Load Balancing Method: The weighted least connection load balancing algorithm expands on the least connection load balancing technique to take into consideration various application server properties. Based on the relative processing power and resource availability of each server in the farm, the administrator gives a weight to each application server. Using the assigned server weights and active connections, the LoadMaster decides how much of each server should be loaded balanced (e.g., if there are two servers with the lowest number of connections, the server with the highest weight is chosen).

And so on.





















































