8. Autoscaling in AWS

Scalability:

- The ability to increase or decrease the IT resources usage according to the requirement.

- Scalability has 2 benefits: Fault tolerance and High Availability.

- The Scalability and its benefits depend upon the Autoscaling service within the EC2 instance.

- The auto scaling is the most important concept in the cloud computing.

Example:

Example using a private data center (not Cloud):

- When a website is created, it has traffic which is handled by the server.

- When the traffic increases, the load on the server increases.

- If the load is greater then the capacity, an alert is set which informs/suggests to buy a new server, join an Elastic load balancer and equally balance the traffic in the website.

- If the traffic is decreased due to unusual circumstances, one server can handle the traffic making the second server idle and of no use.

- We cannot sell the server since high traffic can occur gradually.

- The website creator cannot handle the loss since he invested a lot in the second server.

Example using Cloud Services:

- The auto scaling allows the user to select the required number of servers if the capacity of the load increases than the maximum capacity.

- The servers will be in the autoscaling zone, and will be used when required.

- In the scenario above, if the load on the traffic increases, the autoscaling enables to increase the number of servers as per the requirement where the load balancer service balances out the traffic and server traffic containing capacity.

- If the traffic decreases with the decrease in the load per server, the auto scaling service will descale the requirement of the server2 and the prices will be according to the usage and requirement of the server2.

- Due to Cloud Computing, the hardware, networking, space, time spent and availability is gradually less than owning a data center for a website.

- Not just for the load balancer and EC2, the autoscaling can be done on various AWS resources.

- Auto scaling ensures that you have the right number of AWS 2 instances for the needs at all the time.

- Auto scaling does not have a billing price; due to this we can always use the auto scaling service.

- Scale Out: Maximizes out the usage of the set servers in the autoscaling service.

- Scale In: Decreases the number of servers if its manageable in the existing number of servers.

Auto Scaling Components:

1. Launch Configuration:

- The Launch configuration is used when the types of instances are used whether general purpose,

1. Autoscaling Group:

-

1. Scaling Policy:

-