Explain about Auto Scaling?

- As the name suggests, auto scaling allows you to scale your Amazon EC2 instances up or down automatically as per the instructions set by the user.
- Parameters like minimum and maximum number of instances are set by the user.
- Using this, the number of Amazon EC2 instances you're using increases automatically as the demand rises to maintain the performance, and decreases automatically as the demand decreases to minimize the cost.
- Auto Scaling is particularly effective for those applications that fluctuate on hourly, daily, or weekly usage.
- Auto Scaling is enabled by Amazon CloudWatch and is available at no extra cost.
- AWS CloudWatch can be used to measure CPU utilization, network traffic, etc.

Explain about Elastic Load Balancing?

- **Elastic Load Balancing** (ELB) automatically distributes incoming request traffic across multiple Amazon EC2 instances and results in achieving higher fault tolerance.
- It detects unfit instances and automatically reroutes traffic to fit instances until the unfit instances have been restored in a round-robin manner.
- However, if we need more complex routing algorithms, then choose other services like Amazon Route53.
- ELB consists of the following three components.

Load Balancer:

This includes monitoring and handling the requests incoming through the Internet/intranet and distributes them to EC2 instances registered with it.

Control Service:

This includes automatically scaling of handling capacity in response to incoming traffic by adding and removing load balancers as required. It also performs fitness check of instances.

SSL Termination:

ELB provides SSL termination that saves precious CPU cycles, encoding and decoding SSL within your EC2 instances attached to the ELB. An X.509 certificate is required to be configured within the ELB. This SSL connection in the EC2 instance is optional, we can also terminate it.

Features of ELB:

Following are the most prominent features of ELB −

- ELS is designed to handle unlimited requests per second with gradually increasing load pattern.
- We can configure EC2 instances and load balancers to accept traffic.
- We can add/remove load balancers as per requirement without affecting the overall flow of information.
- It is not designed to handle sudden increase in requests like online exams, online trading, etc.
- Customers can enable Elastic Load Balancing within a single Availability Zone or across multiple zones for even more consistent application performance.

How to Create Load Balancers?

How to Delete a Load Balancer?