MODULE 10: Automatic Scaling and Monitoring

1. Elastic Load Balancing (ELB)
2. Cloudwatch
3. Auto Scaling

---

1. Elastic Load Balancing

- ➔ It distributes incoming application or network traffic across multiple targets in single availability zones or across multiple availability zones.
- ➔ It scales your load balancer as traffic to your application changes overtime.

Types of ELB

a. Application load balancer
b. Network load balancer
c. Classic load balancer

Use cases

- ➔ High availability and fault tolerant application
- ➔ Containerised application
- ➔ Elasticity and scalability
- ➔ VPC
- ➔ Hybrid environments
- ➔ Invoke lambda over http/https

AWS monitoring

- ➔ Access logs
- ➔ CloudWatch metrics
- ➔ CloudTrail logs

2. Cloudwatch

- ➔ Monitor
- ➔ Collect and tracks
- ➔ Alarms, event
- ➔ **It create alarm based on:** static threshold, anomaly detection, metric maths expression

3. Auto Scaling

- ➔ Helps to maintain application availability.
- ➔ Auto scaling group is a collection of EC2 instances that are treated as a logical grouping for the purpose of automated scaling and management.
- ➔ **Scale out:** launch instance.
- ➔ **Scale in:** terminate instance.
- ➔ It monitors your application and automatically adjusts capacity to maintain steady predictable performance at the lower possible cost.