💡 AUTOSCALING

Autoscaling is an advanced feature of AWS which will automatically do resource management based on server load.

🖊 ***Purpose*** : Autoscaling provides users to manage resources to ensure the traffic is handling smoothly, it will be added/removed instances depending on the demand.

🖊 **Major Components**:

- 🔹 EC2 instance - Virtual server exists in #ec2 , applications are deployed through this.
- 🔹 Autoscaling group - collection of EC2 instances and policies , ads/removes instances depend on the load.
- 🔹 AMI - Amazon Machine Image - It provides all information required to launch new instances. Multiple instances can be launched from one AMI.
- 🔹 Load Balancer - It is used to increase the capacity and reliability of applications. The main function is it will divides traffic among instances.

🖊 **Types of Autoscaling**:

- 🔹 Manual scaling - Adding/Removing instances are changed manually using a CLI or console.
- 🔹 Scheduled scaling - Execution of add/remove instances are based on schedules.
- 🔹 Dynamic scaling - Mostly used when there is unpredictable traffic. The number of EC2 instances is changed automatically based on signals that are provided by a CloudWatch alarm.
- 🔹 Predictive scaling - Adding/Removing instances based on the regular pattern of traffic increases/decreases.

🖊 **Advantages**:

- 🔹 Reduce cost.
- 🔹 Enhances performance.
- 🔹 Better fault tolerance.
- 🔹 Maintain application availability.