AWS Auto Scaling Group (ASG)

What is ASG :

- When load of an application increased because of high traffic systems may not be able to handle the request, hence we need more instances. ASG help in adding and removing instances based on the application load.

It has some parameters :

- Minimum (ASG ensure to maintain the instances as per the given number)
- Desired (number of instances running previously or currently)
- Maximum (ASG can launch maximum by number of instances as per given number)
- Automatically register the new instances to the load balancer

What is the difference and in which scenario you use the ASG scaling policies ?

- **Target :** intelligently increase & decrease based on the mentioned target value (70%)
- **Simple :** Scale based on cooldown & warmup period, can't ignore alarms
- **Step :** Allow to scale step by step by ignoring the alarms
- **Schedule :** In predictive situation based on historical data or with known workload

ASG Attributes :

- Launch Configuration :
    - AMI
    - Instance type
    - User data
    - EBS volume
    - Security groups
    - SSH key pair
- Sizing (Min, Max, Desired)
- VPC and Subnets
- ELB
- Scaling policies (Scale-In / Scale-Out)
- CloudWatch Alarms
- ASG Rules (Metrics)
    - Average CPU usage
    - Number of request on ELB per instance
    - Average network in
    - Average network out

Important :

- LC cannot be updated by modifying the , so you need to replace with a new one by copying and updating
- IAM roles are inherited from ASG to instances

- Whenever any instance get terminated under ASG for any reason ASG will create a new one automatically
- If any instance will be declared as unhealthy by ELB ASG will terminate them and replace with a new one
- ASG is free

CPU Load Test :

- Increase the load :

```
# yes > /dev/null &
```

- Decrease the load :

```
# killall yes
```