**AUTO SCALING GROUP:**

* It will automatically scale up and scale down the instances based on the traffic or load.
* It helps to maintain the high availability of our application.
* It is one of the components in EC2 dashboard.
* Auto Scaling Group will not work perfectly or efficiently without load balancer.

**Types of Auto Scaling:**

* **Horizontal scaling**
* **Vertical scaling**

**Horizontal Scaling:**

In which Servers or instances are increased horizontally which means add the servers.

**Types of Horizontal scaling:**

* **Dynamic Scaling**
* **Predictive Scaling**
* **Schedule Action**

**Dynamic Scaling:**

It will scale up or scale down the instances based on CPU utilization.

**Predictive Scaling:**

It will increase or decrease the servers based on or analyze the historical data and AWS itself will take decision.

**Schedule Action:**

It will scale in or scale out the instances based on Date and Time.

**Vertical Scaling:**

It will increase the specification like CPU ,MEMORY to existing server.It will not add the servers so application will go down for some time.

**Steps to Create Auto Scaling Group:**

* Go to EC2 Dashboard and click on Auto Scaling Group.
* First we need to create Launch Template before creating Auto Scaling Group or while creating Auto Scaling Group.
* Click on Create Auto scaling group.
* Enter the name of Auto scaling group.
* Select the Launch template whatever we have created earlier.
* Click on Next option and select the VPC .
* Select the two availability zone and subnet for high availability of application.
* Select the No Load Balancer option.
* Enter the health check grace period time whatever we want.
* Enter the Desired capacity ,minimum and maximum capacity of instances.
* Select the No scaling policies which is optional.
* Select the No policy option under instance maintenance policy.
* Click on next and add tags and click on next.
* Finally click on create auto scaling group.
* Go to Ec2 check whether the instances are creaed or not.