Terraform: EC2 Autoscaling

- Auto Scaling ensures that Amazon EC2 instances are sufficient to run your application.
- When the number of requests increases the load on the servers also increases, AWS will identify and autoscale the resource as per the defined configuration.
- Autoscaling group is mandatory to Auto Scale the EC2 Instances.

- **AutoScaling Components -**
- **Groups** : Logical groups which contain the collection of EC2 instances with similar characteristics for scaling and management purpose.
- Auto-scaling group also maintains a fixed number of instances even if an instance becomes unhealthy.
- Autoscaling group keep checking the EC2 Instances using the HealthCheck.
- If any instance becomes unhealthy, the auto-scaling group terminates the unhealthy instance and launches another instance to replace it.
- Auto scaling groups can increase or decrease the number of EC2 Instances.

- **AutoScaling Components -**
- **Launch Configurations** : Launch configuration is a template used by auto scaling group to launch EC2 instances.
- User specify the Amazon Machine Image (AMI), instances type, key pair, and security groups etc.. while creating the launch configuration.
- User can also edit the launch configuration.

- **AutoScaling Components -**
- **Scaling Plan** : Scaling plans tells Auto Scaling when and how to scale.
- Maintaining Current instance level at all time - User can configure and maintain a specified number of running instances at all the time in the auto scaling group.
- If any unhealthy instance occurs, auto-scaling terminates that instance and launches new instances to replace it.
- Manual Scaling - In Manual scaling, you specify only the changes in maximum, minimum, or desired capacity of your auto scaling groups. Auto-scaling maintains the instances with updated capacity.

- **AutoScaling Components -**
- **Schedule Base Scaling** : User can create a scheduled action which tells Amazon EC2 auto-scaling to perform the scaling action based on the specific time.
- Scale based on demand : This is the most advanced scaling model, resources scales by using a scaling policy.

Will see you in Next Lecture…

See you in next lecture …