

## Quiz: Auto Scaling

Chandra Lingam, Copyright © 2019 Cloud Wave LLC. All Rights Reserved.

1. Application Load Balancer has two target groups A and B. EC2 instances are registered with both target groups. Auto scaling is configured to monitor load balancer health checks and maintain required capacity.

If a failed instance is detected, it should launch a new instance and register with both target groups.

Health check for target group A succeeds for all instances; however, health check for target group B fails for one instance.

What would auto scaling do?

- A. Auto scaling would route traffic only to healthy instances
  - B. Auto scaling would mark the instances that fail any health check as unhealthy and launch replacement instances
  - C. Auto scaling would launch a new instance in target group B
  - D. Behavior is undefined
2. If you delete an auto scaling group, what would happen to the instances that are part of the auto scaling group?
    - A. Instances are detached from the auto scaling group and continue running
    - B. Instances are terminated
    - C. Instances are stopped
    - D. You must first remove instances before you can delete an Auto scaling group
  3. You have an Auto scaling group that has a policy to add additional instances based on CloudWatch metrics. You have already reached your account limit for number of EC2 instances. When auto scaling triggers a scaling out event to add an additional EC2 instance, what would happen?
    - A. Instance is launched as limits do not apply to auto scaling actions
    - B. Instance would fail to launch
    - C. Autoscaling would terminate an unused instance and launch a new instance
  4. Your application has several EC2 instances. You would like to rapidly increase capacity when demand increases for your services while gradually reducing instances when there is not much demand.

How can you achieve this?

- A. Scaling out and scaling in events needs to be matched evenly to prevent unbalanced resources
- B. Use different policies
- C. Use fewer time periods for scaling out and larger number of time periods for scaling in events

- D. Use larger number of time periods for scaling out events and smaller number of time periods for scaling in events
5. Your application demand is fairly stable, and you are not expecting any changes in demand in the near future.  
There are six EC2 instances currently handling all the requests.  
Given this scenario would you still use auto scaling?
- A. Yes
  - B. No

Answers:

- 1. B - All health checks must pass for an instance to be considered healthy. In this case only one health check passes, so the failing instance would be replaced
- 2. B - Instances are managed as part of Auto Scaling Group and they are terminated when you delete an autoscaling group
- 3. B - You won't be able to launch any more instances in the region until you increase the limit with AWS or terminate an instance that you no longer need
- 4. B – Use two different policies. Scaling out policy with desired increase in instances and scaling in policy with desired decrease in instances
- 5. A - Auto scaling can maintain capacity by monitoring health of your instances and replacing them if they are not healthy