**Elastic Load Balancer**

1. **What is Elastic Load Balancing (ELB) in AWS?**
   * *Answer:* Elastic Load Balancing (ELB) is a managed service in AWS that automatically distributes incoming application traffic across multiple targets, such as EC2 instances, containers, and IP addresses, improving fault tolerance and scalability of applications.
2. **What are the different types of load balancers in AWS?**
   * *Answer:* AWS offers three types of load balancers: Application Load Balancer (ALB), Network Load Balancer (NLB), Gateway Load Balancer (GLB), and Classic Load Balancer (CLB). Each type caters to different application needs and protocols.
3. **How does ELB integrate with AWS Auto Scaling?**
   * *Answer:* ELB works in conjunction with Auto Scaling to dynamically adjust capacity based on demand, automatically distributing traffic across instances added or removed by Auto Scaling.
4. **What is cross-zone load balancing in ELB?**
   * *Answer:* Cross-zone load balancing in ELB evenly distributes incoming traffic across multiple Availability Zones (AZs) within a region, preventing overloading of instances in one AZ while others remain underutilized.
5. **How can you improve security when using load balancers?**
   * *Answer:* You can associate security groups, authentication, SSL/TLS decryption with ELB to provide more network security. Elastic Load Balancer associate with Amazon Virtual Private Cloud (VPC) to provide strong network security.
6. **What is connection draining in ELB?**
   * *Answer:* Connection draining is a feature in ELB that allows in-flight requests to complete before terminating unhealthy instances, ensuring that no requests are lost during the process.
7. **What is the significance of using Elastic Load Balancing with Auto Scaling for application availability?**
   * *Answer:* Combining ELB with Auto Scaling ensures that incoming traffic is distributed evenly across instances, and the Auto Scaling group adjusts capacity to maintain application availability.
8. **How can you monitor the health and performance of targets behind an Elastic Load Balancer?**
   * *Answer:* You can use CloudWatch metrics to monitor the health and performance of targets behind an Elastic Load Balancer, such as request count, latency, and error rates.
9. **What is the difference between Application Load Balancer and Network Load Balancer in AWS?**
   * *Answer:* Application Load Balancer operates at layer-7 of the OSI model and is suitable for HTTP/HTTPS traffic, while Network Load Balancer operates at layer-4 and is suitable for TCP/UDP traffic.
10. **What is the significance of sticky sessions in ELB?**
    * *Answer:* Sticky sessions in ELB ensure that a user's requests are directed to the same instance during a session, improving application performance and user experience.
11. **What is the difference between a target group and a listener in ELB?**
    * 1. Layer

ALB: Works at Layer 7 (Application Layer), routing based on HTTP/HTTPS content.  
NLB: Works at Layer 4 (Network Layer), routing based on IP addresses and ports.

* + 2. Use Case

ALB: Ideal for web apps that need content-based routing (e.g., path, headers).  
NLB: Best for high-performance, low-latency apps (e.g., TCP/UDP traffic).

* + 3. Performance

ALB: Slightly slower, good for HTTP/HTTPS traffic.  
NLB: Ultra-low latency, supports millions of requests per second.

4. Health Checks

ALB: Application-level (HTTP/HTTPS).  
NLB: Network-level (TCP).

* + 5. IP Addresses

ALB: Dynamic Ips.  
NLB: Supports static IPs or Elastic IPs.

* + 6. SSL

ALB: Supports SSL termination.  
NLB: Can pass through SSL traffic.

1. **How can you configure load balancing across multiple regions in ELB?**
   * *Answer:* You can configure load balancing across multiple regions in ELB by using Route 53's latency-based routing feature, which directs traffic to the region with the lowest latency.

**ASG**

1. **What is AWS Auto Scaling and how does it work?**
   * AWS Auto Scaling is a service that automatically adjusts the capacity of your Amazon Elastic Compute Cloud (EC2) instances, Amazon Elastic Container Service (ECS) tasks,
2. **What are the types of Auto Scaling Policies**There are three primary types of dynamic scaling policies in AWS EC2 Auto Scaling:
   * Target Tracking: Increases or decreases the current capacity of the group based on a target value for a specific metric.
   * Step Scaling: Increases or decreases the current capacity of the group based on a set of scaling adjustments, known as step adjustments, that vary based on the size of the alarm breach.
   * Simple Scaling: Increases or decreases the current capacity of the group based on a single scaling adjustment, with a cooldown period between each scaling activity
3. **What are the two main components of AWS Auto Scaling?**
   * The two main components of AWS Auto Scaling are: **Scaling Plans:** These are the resources that you want to scale, such as EC2 instances, ECS tasks, or DynamoDB tables**. Scaling Policies:** These are the rules that define how your resources should scale based on demand.
4. **How does Auto Scaling help with cost optimization?**
   * Auto Scaling helps with cost optimization by adding or removing instances based on demand, thus reducing the amount of resources needed to pay for when demand is low and avoiding running out of resources when demand is high.
5. **What is RTO and RPO in AWS?**
   * RTO (Recovery Time Objective) is the maximum time your business is willing to wait for a recovery to occur, while RPO (Recovery Point Objective) is the maximum amount of data loss your company is willing to accept as measured in time.
6. **How can you add an existing instance to a new Auto Scaling group?**
   * You can add an existing instance to a new Auto Scaling group by opening the EC2 console, selecting your instance under Instances, choosing Actions -> Instance Settings -> Attach to Auto Scaling Group, selecting a new Auto Scaling group, and attaching the group to the instance.
7. **What are some potential drawbacks and limitations of Auto Scaling?**
   * Some potential drawbacks and limitations of Auto Scaling include increased costs if not configured properly, added complexity to infrastructure management, latency during instance launches, and the need for careful monitoring and adjustment.