**1. What is an Elastic Load Balancer (ELB)?**

[Elastic Load Balancing](https://datavalley.ai/) ([ELB](https://www.datavalley.ai/aws-load-balancer-interview-questions/)) is a load-balancing service for [Amazon Web Services (AWS)](https://www.youtube.com/@datavalley-ai) deployments. ELB automatically distributes incoming application traffic and scales resources to meet traffic demand. It also helps an IT team adjust capacity according to incoming application and network traffic. Load balancing divides the amount of work that a computer has to do among multiple computers so that users, in general, get served faster. ELB offers enhanced features including:

Detection of unhealthy Elastic Compute Cloud (EC2) instances.

Spreading instances across healthy channels only.

Flexible cipher support.

Centralized management of Secure Sockets Layer (SSL) certificates.

Optional public key authentication.

Support for both IPv4 and IPv6.

**2. What are the key features provided by Elastic Load Balancer (ELB)?**

Elastic Load Balancer (ELB) is a load balancing service provided by Amazon Web Services (AWS). It automatically distributes incoming application traffic across multiple targets, such as Amazon EC2 instances, containers, and IP addresses, in one or more Availability Zones. ELB helps to increase the availability and fault tolerance of your application.

Here are some key features provided by ELB:

1. High availability: ELB automatically distributes traffic across multiple targets in multiple Availability Zones, which helps to increase the availability of your application.
2. Health checks: ELB can perform health checks on your targets and automatically route traffic only to the healthy targets.
3. Automatic scaling: ELB can scale automatically in response to incoming traffic, so you don’t have to worry about manually adding or removing targets as demand changes.
4. SSL/TLS offloading: ELB can terminate SSL/TLS connections, which frees up your targets to focus on processing requests rather than handling encryption and decryption.
5. Sticky sessions: ELB can maintain session affinity, or “stickiness,” so that requests from a particular client are always routed to the same target.
6. Cross-zone load balancing: ELB can evenly distribute traffic across all targets in all enabled Availability Zones, which helps to increase the fault tolerance of your application.
7. Customized health checks: ELB allows you to customize the health check settings for your targets, such as the ping target, the interval between pings, and the number of consecutive failures required before marking a target as unhealthy.
8. Multiple load balancer types: ELB provides three types of load balancers: Application Load Balancer, Network Load Balancer, and Classic Load Balancer. Each type is optimized for different workloads and use cases.

**3. How**[**AWS Elastic Load Balancing**](https://www.datavalley.ai/aws-load-balancer-interview-questions/)**Works?**

A load balancer accepts incoming traffic from clients and routes requests to its registered targets (such as EC2 instances) in one or more Availability Zones. The load balancer also monitors the health of its registered targets and ensures that it routes traffic only to healthy targets. When the load balancer detects an unhealthy target, it stops routing traffic to that target. It then resumes routing traffic to that target when it detects that the target is healthy again.

A listener is a process that checks for connection requests. It is configured with a protocol and port number for connections from clients to the load balancer. Likewise, it is configured with a protocol and port number for connections from the load balancer to the targets.

**4. What do you mean by a target group in**[**AWS**](https://www.datavalley.ai/aws-interview-questions/)**Load Balancing?**

A target group is a group of resources that you can load balance across in Amazon Web Services (AWS). It is used to specify a group of targets, such as Amazon EC2 instances, containers, and IP addresses, that you want to route traffic to. Target groups are used with Application Load Balancers and Network Load Balancers in AWS.

**5. What is the difference between auto-scaling and ELB?**

**Load balancing** evenly distributes load to application instances in all availability zones in a region while auto scaling makes sure instances scale up or down depending on the load.

**Auto-scaling** is a cloud computing technique for dynamically allocating computational resources. Depending on the load to a server farm or pool, the number of servers that are active will typically vary automatically as user needs fluctuate.

**Auto-scaling and load-balancing are related because an application typically scales based on load-balancing serving capacity. In other words, the serving capacity of the load balancer is one of several metrics that shape the auto-scaling policy.**

**6. How do you configure session affinity in an**[**AWS**](https://www.datavalley.ai/aws-interview-questions/)**Elastic Load Balancer?**

Session affinity, also known as sticky sessions, is a technique used to route traffic from a user to the same web server during the duration of that user’s session. This ensures that all requests from that user are handled by the same server, which can be important for maintaining state information or ensuring that a user’s data is always available.

To configure session affinity in an AWS Elastic Load Balancer, you first need to create a load balancer and add your web servers to it. Then, you need to enable session affinity for the load balancer and specify the duration of the session. Finally, you need to configure your web servers to use the load balancer as their session affinity provider

**7. Explain NLB in AWS?**

AWS Network Load Balancer (NLB) is an Amazon Web Services (AWS) tool that distributes end user traffic across multiple cloud resources to ensure low latency and high throughput for applications.When a target becomes slow or unavailable, the Network Load Balancer routes traffic to another target.

A load balancer accepts incoming traffic from clients and routes requests to its registered targets (such as EC2 instances) in one or more Availability Zones. The load balancer also monitors the health of its registered targets and ensures that it routes traffic only to healthy targets.

**8. What are the different types of load**[**balancers**](https://www.datavalley.ai/aws-load-balancer-interview-questions/)**?**

There are three main types of load balancers: hardware, software, and cloud. Hardware load balancers are physical devices that are installed in between the server and the client. Software load balancers are installed on the server and work with the server software to distribute traffic. Cloud load balancers are hosted in the cloud and work with cloud-based applications.

**9. What are the types of load balancers?**

Elastic Load Balancing supports the following types of load balancers:

Application Load Balancers – It allows a developer to configure and route incoming end-user traffic to applications based in the Amazon Web Services (AWS) public cloud. It pushes traffic across multiple targets in multiple AWS Availability Zones.

Network Load Balancers – The Network Load Balancing feature distributes traffic across several servers by using the TCP/IP networking protocol. It works by combining two or more computers that are running applications into a single virtual cluster, NLB provides reliability and performance for web servers and other mission-critical servers.

* Gateway Load Balancers – GLB enables you to deploy, scale, and manage virtual appliances, such as firewalls, intrusion detection and prevention systems, and deep packet inspection systems. A Gateway Load Balancer operates at the third layer of the Open Systems Interconnection (OSI) model, the network layer.
* Classic Load Balancers – It provides basic load balancing across multiple Amazon EC2 instances and operates at both the request level and connection level. Classic Load Balancer is intended for applications that are built within the EC2-Classic network.

**10. How can we assign a static IP address to an ELB?**

The network interfaces IP address and Windows adapter IP addresses must match for a successful network connection.

* Open the Amazon Elastic Compute Cloud (Amazon EC2) console.
* Choose the Region where you want to create your Network Load Balancer.
* Allocate Elastic IP addresses for your Network Load Balancer.
* Note: Be sure to create the same number of Elastic IP addresses as the number of subnets that you select in each Availability Zone. For more information, see Elastic IP address limit.
* Under Load Balancing, choose Load Balancers.
* Choose Create Load Balancer.
* For Network Load Balancer, choose to Create.
* Select the VPC where you plan to provide the Network Load Balancer.
* Select the check box for the desired Availability Zone. Then, select the subnet in that Availability Zone.
* By default, the Network Load Balancer assigns a public IP address. If you want to use an Elastic IP address: For an IPv4 address, choose Choose an Elastic IP.
* Select one of the Elastic IP addresses available in the specified Region. This IP address will be assigned to the Network Load Balancer’s interfaces.
* (Optional) If you have multiple Availability Zones, repeat steps 8-9 for each zone.
* Complete the remaining steps for creating a Network Load Balancer.

**11. What is**[**VPC**](https://www.datavalley.ai/aws-vpc-interview-questions/)**load balancer?**

When you provision an Elastic Load Balancer for your VPC, you can assign security groups to it. You can place ELBs into VPC subnets, and you can also use subnet ACLs(Access Control Lists). The EC2 instances that you register with the Elastic Load Balancer do not need to have public IP addresses.

**12. Which ELB component is responsible for monitoring the Load Balancers?**

**The controller service**

The load balancers monitor the traffic and handle requests and The **controller service** is responsible for monitoring the load balancers, adding and removing load balancers as required and verifying that the load balancers are functioning properly.

**13. What are the different types of load balancing options provided by Amazon Elastic Load Balancing (ELB)?**

Amazon Elastic Load Balancing (ELB) provides three types of load balancers:

1. Application Load Balancer: This type of load balancer is optimized for load balancing HTTP and HTTPS traffic. It allows you to specify routing rules based on the content of the request, such as the host, the path, and the query string.
2. Network Load Balancer: This type of load balancer is optimized for load balancing TCP traffic and is designed to handle millions of requests per second with low latency. It can be used to load balance traffic to Amazon EC2 instances, containers, and IP addresses.
3. Classic Load Balancer: This is the original type of load balancer provided by ELB and is designed to load balance HTTP, HTTPS, and TCP traffic. It is a good choice for applications that require simple load balancing with basic health checks.

**14. Can we have internal or private ELB as origin in cloudfront?**

No, it should be public as cloudfront doesn’t have a way to access private ELB.

**15. Can you explain what round-robin load balancing is?**

Round-robin load balancing is a method of load balancing that distributes traffic evenly across a number of servers. This is done by taking turns sending requests to each server in a sequence. This ensures that all servers receive an equal amount of traffic, and helps to prevent any one server from becoming overloaded.

**16. Are you using Auto Scaling for your**[**EC2**](https://www.datavalley.ai/aws-ec2-interview-questions-answers/)**instances ? How have you configured Auto scaling policy ?**

yes, you can create auto scaling policy to set max and min no of instance.

**17. What is the best way to set up a load balancer in a cloud environment like Amazon Web Services?**

The best way to set up a load balancer in a cloud environment like Amazon Web Services is to use a combination of an Elastic Load Balancer (ELB) and an Auto Scaling group. The ELB will distribute traffic evenly between the instances in the Auto Scaling group, and the Auto Scaling group will ensure that there are always enough instances available to handle the traffic.

**18. List the types of techniques that are used by load balancers?**

* Round Robin.
* Weighted Round Robin.
* Least Connection.
* Weighted Least Connection.
* Resource Based (Adaptive)
* Resource-Based (SDN Adaptive)
* Fixed Weighting.
* Weighted Response Time.

**19. What is the difference between cluster and load balancing?**

Here is a summary of the key differences between cluster and load balancing:

| **Feature** | **Cluster** | **Load Balancing** |
| --- | --- | --- |
| Purpose | Improved availability and fault tolerance | Improved performance and scalability |
| How it works | Multiple servers work together as a single system | Incoming requests are distributed across multiple servers |
| Failover | Automatically switches to a standby server in case of a failure | Automatically routes requests to healthy servers |
| Scaling | Scales vertically by adding more servers to the cluster | Scales horizontally by adding more servers to the pool |
| Load balancing algorithm | Typically uses active/passive or active/active configuration | Can use various algorithms, such as round robin, least connection, and source hash |
| Example use cases | Database clusters, web servers | Web servers, application servers |

Cluster refers to a group of servers that work together as a single system to improve the availability and fault tolerance of a service. A cluster typically consists of multiple servers configured in an active/passive or active/active configuration, with one or more servers serving as standby servers that can take over if the primary server fails. Clusters are used to improve the availability of critical services, such as databases and web servers.

Load balancing refers to the process of distributing incoming requests across multiple servers in order to improve the performance and scalability of a service. Load balancing can be used with various types of servers, such as web servers and application servers, and can use various algorithms to determine how to distribute the requests. Load balancing helps to ensure that incoming requests are handled efficiently and that the service remains available even if one or more servers fail.

**20. What’s the difference between Active and Passive Health Checks?**

Active health checks are initiated by the load balancer itself, and passive health checks are initiated by the server. Active health checks are generally more accurate, because the load balancer can directly check whether the server is responding. Passive health checks are less accurate, because the server may be responding to the load balancer but not to actual client requests.

**21. Why is it important to have fault tolerance when using a load balancer?**

A load balancer is a key component in a high availability system, and as such, it is important to have a fault tolerance strategy in place in case the load balancer itself fails. One common way to achieve this is to use a redundant array of independent load balancers (RAIL), which provides a measure of protection should one of the load balancers fail.

**22. When would you choose a Layer 4 vs a Layer 7 load balancer?**

Layer 4 load balancers work at the transport layer, which means they are only concerned with the source and destination IP addresses and port numbers of packets. This makes them much simpler and faster than Layer 7 load balancers, which have to inspect the contents of each packet to make routing decisions. However, this also means that Layer 4 load balancers cannot make decisions based on things like the contents of HTTP headers, so they are not well-suited for load balancing web traffic.

**23. Can you give me some examples of real world use cases for DNS load balancing?**

DNS load balancing can be used in a number of different ways, depending on your specific needs. For example, if you have a website that gets a lot of traffic, you might use DNS load balancing to distribute that traffic across multiple servers, in order to avoid overloading any one server. Alternatively, you might use DNS load balancing to provide redundancy in case one of your servers goes down. There are many other potential use cases for DNS load balancing, but these are two of the most common.

**24. Can you explain how DNS load balancing works? How does it differ from other forms of load balancing?**

DNS load balancing is a method of load balancing that uses the Domain Name System (DNS) to distribute traffic across a number of servers. This type of load balancing is often used because it is simple to set up and does not require any special hardware or software. DNS load balancing is different from other forms of load balancing because it relies on the DNS system to route traffic, rather than using a dedicated load balancer.

**25. What are some advantages of using a reverse proxy as opposed to a forward proxy with load balancing?**

A reverse proxy can provide a single point of contact for clients, which can simplify the configuration of firewalls and improve security. Additionally, a reverse proxy can perform caching and compression, which can improve performance.

**26. How can you improve security when using load balancers?**

One way to improve security when using load balancers is to use a reverse proxy. A reverse proxy is a server that sits between the load balancer and the web server. The reverse proxy can provide an additional layer of security by filtering traffic and only allowing requests that meet certain criteria to pass through to the web server.

**27. Is it possible to implement SSL encryption when using a load balancer? If yes, then how?**

Yes, it is possible to implement SSL encryption when using a load balancer. This can be done by setting up a secure socket layer (SSL) between the load balancer and the web server. This will ensure that all data passing between the two is encrypted and secure.

**28. What is the purpose of sticky sessions when using a load balancer?**

Sticky sessions are used to ensure that all requests from a particular user are routed to the same server. This is important because it can help to ensure that session information is not lost and that the user experience is consistent.

**29. Which of the Load balancer – Application / Network provides advanced features ? What is the use of other Load Balancer than if one provides advanced feature than other?**

Application Load Balancer is high level ( Level 7 Networking Layer ) and provides better features and flexibility than Network Load Balancer ( Level 4 Network Layer) Network Load Balancer is turn provides better performance as it regulates raw traffic.

**30. Difference between internal and external Load Balancer?**

Here is a summary of the key differences between internal and external Elastic Load Balancers (ELBs) in Amazon Web Services (AWS):

| **Feature** | **Internal Load Balancer** | **External Load Balancer** |
| --- | --- | --- |
| Purpose | Load balancing within a Virtual Private Cloud (VPC) | Load balancing between the internet and a VPC |
| Accessibility | Only accessible from within the VPC | Accessible from the internet |
| Domain name | Uses a private IP address as the domain name | Uses a public DNS name as the domain name |
| Supported protocols | HTTP, HTTPS, TCP | HTTP, HTTPS, TCP |
| Supported load balancer types | Application Load Balancer, Network Load Balancer | Application Load Balancer, Network Load Balancer, Classic Load Balancer |
| Health checks | Can perform health checks on targets within the VPC | Can perform health checks on targets within and outside the VPC |
| Example use cases | Load balancing within a VPC, connecting multiple VPCs | Load balancing between the internet and a VPC, exposing a VPC-based service to the internet |

An internal Load Balancer is a type of ELB that is used to load balance traffic within a VPC. It is only accessible from within the VPC and uses a private IP address as the domain name. Internal Load Balancers are often used to load balance traffic between private subnets within a VPC or to connect multiple VPCs.

An external Load Balancer is a type of ELB that is used to load balance traffic between the internet and a VPC. It is accessible from the internet and uses a public DNS name as the domain name. External Load Balancers are often used to expose a VPC-based service to the internet or to load balance traffic between the internet and a VPC.

## AWS [Auto Scaling](https://youtube.com/@datacademy-ai?sub_confirmation=1) Interview Questions

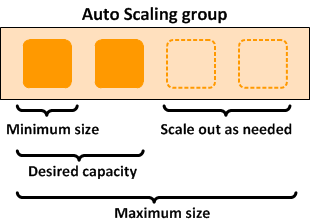
### 1. What is AWS Auto-Scaling and how does it work?

Amazon Web Services [(AWS)](https://datavalley.ai/wp-admin/post.php?post=8846&action=edit)[Auto Scaling](https://youtube.com/@datacademy-ai?sub_confirmation=1)is a service that automatically adjusts the capacity of your Amazon Elastic Compute Cloud[(EC2](https://datavalley.ai/sql-query-interview-questions/)) instances, Amazon Elastic Container Service (ECS) tasks, or Amazon DynamoDB tables in response to changes in demand for your application. With AWS Auto Scaling, you can ensure that you have the right amount of resources available to meet the demands of your application while minimizing the costs associated with idle resources.

[AWS](https://datavalley.ai/google-cloud-platform-gcp-datacademy-ai/) Auto Scaling works by monitoring the performance of your application and automatically increasing or decreasing the number of resources as needed to maintain the desired performance level. You can configure AWS Auto Scaling to scale resources based on a variety of metrics, such as CPU utilization or network bandwidth, and you can specify minimum and maximum resource limits to ensure that your application has sufficient resources to meet its needs while staying within your budget.

AWS Auto Scaling is a flexible and powerful tool that can help you optimize the performance and cost of your application. It can be especially useful for applications with variable or unpredictable workloads, as it allows you to scale resources up or down as needed to meet changing demands.

To use AWS Auto Scaling, you will need to create an Auto Scaling group and define the minimum and a maximum number of resources that you want to have available, as well as the desired capacity and any scaling policies or rules that you want to use to control the scaling process. [AWS](https://www.datavalley.ai/aws-interview-questions/)Auto Scaling will then monitor your application and automatically adjust the number of resources as needed to meet your specified performance criteria.



### 2. Does AWS Autoscaling Work?

Yes, Amazon Web Services [(AWS)](https://datavalley.ai/google-cloud-platform-gcp-datacademy-ai/) Autoscaling is a service that enables you to automatically scale resources in response to changes in demand for your application. With AWS Autoscaling, you can ensure that you have the right amount of resources available to meet the demands of your application while minimizing the costs associated with idle resources.

AWS Autoscaling works by monitoring the performance of your application and automatically increasing or decreasing the number of resources (such as Amazon Elastic Compute Cloud (EC2) instances, Amazon Elastic Container Service (ECS) tasks, or Amazon DynamoDB throughput) as needed to maintain the desired performance level. You can configure AWS Autoscaling to scale resources based on a variety of metrics, such as CPU utilization or network bandwidth, and you can specify minimum and maximum resource limits to ensure that your application has sufficient resources to meet its needs while staying within your budget.

AWS Autoscaling is a flexible and powerful tool that can help you optimize the performance and cost of your application. It can be especially useful for applications with variable or unpredictable workloads, as it allows you to scale resources up or down as needed to meet changing demands.

### 3. Define Auto Scaling.

[Auto-scaling](https://datavalley.ai/itil-interview-questions/) is a technique used to automatically adjust the capacity of a system to meet changing workloads. It is commonly used in cloud computing environments, such as Amazon Web Services (AWS), to scale resources such as Amazon Elastic Compute Cloud[(EC2)](https://datavalley.ai/itil-interview-questions/) instances, Amazon Elastic Container Service (ECS) tasks, or Amazon DynamoDB tables up or down as needed to meet the demands of an application.

With auto-scaling, you can define policies or rules that specify the conditions under which the system should scale resources up or down. For example, you might configure auto-scaling to increase the number of EC2 instances in your application when the average CPU utilization of your current instances exceeds a certain threshold. Similarly, you might configure auto-scaling to decrease the number of instances when the average CPU utilization falls below a certain threshold, in order to save costs.

Auto scaling can help you optimize the performance and cost of your application by ensuring that you have the right amount of resources available to meet the demands of your users while minimizing the costs associated with idle resources. It can be especially useful for applications with variable or unpredictable workloads, as it allows you to scale resources up or down as needed to meet changing demands.

### 4. What is the difference between a single instance Web environment and Load-Balanced Auto-Scaling?

In a single instance web environment, your application is running on a single server or virtual machine [(VM).](https://datavalley.ai/itil-interview-questions/) This means that all requests from users are directed to that single server, and the performance of your application is limited by the resources of that server. This can be a simple and cost-effective solution for small applications with low traffic, but it may not be sufficient for larger or more heavily-trafficked applications, as the single server may not be able to handle the increased workload.

On the other hand, in a load-balanced, auto-scaling environment, your application is running on multiple servers or VMs, and a load balancer is used to distribute incoming requests evenly across those servers. This allows your application to scale horizontally, meaning that you can add more servers or VMs as needed to handle increased traffic.

Load-balanced, auto-scaling environments offer several benefits over single-instance environments. They can handle higher levels of traffic and provide a better performance, as the workload is distributed across multiple servers. They can also provide better availability, as the load balancer can automatically redirect traffic to healthy servers if one of the servers goes down.

Additionally, auto-scaling allows you to automatically add or remove servers based on demand, so you can ensure that you have the right amount of resources available to meet the needs of your users. This can help you optimize the cost of your application, as you only pay for the resources you actually use.

### 5. On an EC2 instance, an application of yours is active. Once the CPU usage on your instance hits 80%, you must reduce the load on it. What strategy do you use to complete the task?

There are several strategies that you could use to reduce the load on an [EC2](https://datavalley.ai/)instance when the CPU usage hits 80%. Here are a few options:

1. **Scale-out:**One option would be to add additional EC2 instances to your application, and use a load balancer to distribute incoming requests evenly across all of the instances. This will allow you to handle more traffic and reduce the load on any individual instance.
2. **Optimize your application:**Another option would be to optimize your application to use fewer resources. This might involve improving the performance of your code, using more efficient algorithms, or reducing the number of requests that your application makes to external services.
3. **Use auto-scaling:**You could also use Amazon Web Services ([AWS](https://www.datavalley.ai/aws-s3-interview-question/)) Auto Scaling to automatically add or remove EC2 instances based on demand. For example, you could configure Auto Scaling to increase the number of instances when the CPU usage of your current instances exceeds a certain threshold, and to decrease the number of instances when the CPU usage falls below a certain threshold. This will allow you to automatically scale your application up or down as needed to meet changing demands.
4. **Use a larger instance type:**If you are using a smaller instance type, you might consider upgrading to a larger instance type with more CPU and memory resources. This will allow you to handle more traffic and reduce the load on the instance.
5. **Use a cache:**Implementing a cache (such as Amazon [ElastiCache)](https://datavalley.ai/) can help reduce the load on your EC2 instance by allowing your application to store frequently accessed data in memory, rather than having to retrieve it from a database or other external service on every request.

There is no one-size-fits-all solution for reducing the load on an [EC2](https://datavalley.ai/launch-ec2-instance-custom-ami-aws-datavalley/)instance, and the best approach will depend on your specific needs and resources. It may be helpful to try a combination of these strategies to find the most effective solution for your application.

### ****6. Explain the auto-scaling feature of EC2 along with its benefits.****

Amazon Elastic Compute Cloud[(EC2)](https://datavalley.ai/) Auto Scaling is a service that enables you to automatically scale the capacity of your EC2 instances up or down in response to changes in demand for your application. With EC2 Auto Scaling, you can ensure that you have the right amount of resources available to meet the demands of your users while minimizing the costs associated with idle resources.

Some of the benefits of using [EC2](https://datavalley.ai/launch-ec2-instance-custom-ami-aws-datavalley/)Auto Scaling include:

* **Improved performance:**By adding or removing EC2 instances as needed, EC2 Auto Scaling can help you ensure that your application has the resources it needs to meet the demands of your users. This can result in improved performance and a better user experience.
* **Better availability:**EC2 Auto Scaling can help you improve the availability of your application by automatically replacing any instances that fail or become unavailable. This can help you avoid downtime and ensure that your users can access your application when they need to.
* **Cost optimization:**EC2 Auto Scaling allows you to automatically scale your resources up or down as needed to meet changing demands. This can help you optimize your costs by only paying for the resources that you actually use.
* **Simplified resource management:**EC2 Auto Scaling makes it easy to manage the capacity of your EC2 instances, as it automatically scales your resources up or down based on your specified policies and rules. This can help you save time and reduce the complexity of managing your resources manually.

To use EC2 Auto Scaling, you will need to create an Auto Scaling group and define the minimum and a maximum number of EC2 instances that you want to have available, as well as the desired capacity and any scaling policies or rules that you want to use to control the scaling process. EC2 Auto Scaling will then monitor your application and automatically adjust the number of instances as needed to meet your specified performance criteria.

### ****7. Can you change the instance type of the instances that are running in your application tier and are also using autoscaling? If yes, then how? (Choose one of the following)****

Yes, you can change the instance type of the instances that are running in your application tier and are also using autoscaling. Here’s one way you can do this:

1. **Create a new launch configuration with the desired instance type:**First, you will need to create a new launch configuration that specifies the new instance type that you want to use. You can do this using the AWS Management Console, the AWS Command Line Interface (CLI), or the AWS [SDKs.](https://datavalley.ai/)
2. **Update the Auto Scaling group to use the new launch configuration:** Next, you will need to update your Auto Scaling group to use the new launch configuration that you created. You can do this using the AWS Management Console, the [AWS](https://www.datavalley.ai/aws-interview-questions/) CLI, or the AWS SDKs.
3. **Wait for the instances to be replaced:** Once you have updated the Auto Scaling group to use the new launch configuration, the Auto Scaling group will automatically replace the old instances with new ones using the new instance type. This process may take some time, as the Auto Scaling group will replace the instances one at a time to ensure that there is always at least one instance available to serve traffic.

It’s important to note that changing the instance type of your instances may impact the performance and cost of your application. Be sure to carefully consider the impact of any changes that you make, and test your application thoroughly after making any changes to ensure that it is performing as expected.

### ****8. I created a web application with autoscaling. I observed that the traffic on my application is the highest on Wednesdays and Fridays between 9 AM and 7 PM. What would be the best solution for me to handle the scaling?****

To handle the scaling of your web application in the most efficient way, you could use Amazon Web Services (AWS) Auto Scaling with scheduled scaling. Scheduled scaling allows you to specify when and by how much you want to scale your application, based on a schedule that you define.

To use scheduled scaling for your web application, you will need to do the following:

1. **Create an Auto Scaling group:** First, you will need to create an Auto Scaling group for your web application. This will allow you to manage the capacity of your application and specify the minimum and the maximum number of instances that you want to have available.
2. **Create a scheduled action:**Next, you will need to create a scheduled action that specifies when and by how much you want to scale your application. You can do this using the [AWS](https://www.datavalley.ai/aws-interview-questions/)Management Console, the AWS Command Line Interface (CLI), or the AWS SDKs. When creating the scheduled action, you will need to specify the desired capacity for your Auto Scaling group, as well as the start and end times for the action.
3. **Set the desired capacity for your Auto Scaling group:**Finally, you will need to set the desired capacity for your Auto Scaling group to the desired capacity specified in your scheduled activities. This will ensure that your Auto Scaling group scales your application according to your schedule.

With scheduled scaling, you can ensure that your application has the resources it needs to meet the demands of your users during the times of highest traffic while minimizing the costs associated with idle resources. This can help you optimize the performance and cost of your application, and provide a better user experience for your users.

### ****9. You have an application running on your Amazon****[EC2](https://www.datavalley.ai/restore-ec2-from-snapshot/)****instance. You want to reduce the load on your instance as soon as the CPU utilization reaches 100 percent. How will you do that?****

There are several strategies that you could use to reduce the load on an Amazon Elastic Compute Cloud (EC2) instance when the CPU utilization reaches 100 percent. Here are a few options:

1. **Scale-out:**One option would be to add additional EC2 instances to your application, and use a load balancer to distribute incoming requests evenly across all of the instances. This will allow you to handle more traffic and reduce the load on any individual instance.
2. **Optimize your application:**Another option would be to optimize your application to use fewer resources. This might involve improving the performance of your code, using more efficient algorithms, or reducing the number of requests that your application makes to external services.
3. **Use auto-scaling:**You could also use Amazon Web Services (AWS) Auto Scaling to automatically add or remove EC2 instances based on demand. For example, you could configure Auto Scaling to increase the number of instances when the CPU utilization of your current instances exceeds a certain threshold, and to decrease the number of instances when the CPU utilization falls below a certain threshold. This will allow you to automatically scale your application up or down as needed to meet changing demands.
4. **Use a larger instance type:**If you are using a smaller instance type, you might consider upgrading to a larger instance type with more CPU and memory resources. This will allow you to handle more traffic and reduce the load on the instance.
5. **Use a cache:**Implementing a cache (such as Amazon ElastiCache) can help reduce the load on your [EC2](https://www.datavalley.ai/restore-ec2-from-snapshot/)instance by allowing your application to store frequently accessed data in memory, rather than having to retrieve it from a database or other external service on every request.

There is no one-size-fits-all solution for reducing the load on an EC2 instance, and the best approach will depend on your specific needs and resources. It may be helpful to try a combination of these strategies to find the most effective solution for your application.

### ****10. You have an application running on an EC2 instance. You need to reduce the load on your instance as soon as the CPU utilization reaches 80 percent. How will you accomplish the job?****

It can be done by creating an autoscaling group to deploy more instances when the CPU utilization of the EC2 instance exceeds 80 percent and distributing traffic among instances by creating an application load balancer and registering EC2 instances as target instances.

## [AWS Auto Scaling Interview Questions](https://youtube.com/@datacademy-ai?sub_confirmation=1)

### 11. What is Amazon EC2 Auto Scaling?

Amazon Elastic Compute Cloud (EC2) Auto Scaling is a service that enables you to automatically scale the capacity of your EC2 instances up or down in response to changes in demand for your application. With EC2 Auto Scaling, you can ensure that you have the right amount of resources available to meet the demands of your users while minimizing the costs associated with idle resources.

EC2 Auto Scaling works by monitoring the performance of your application and automatically increasing or decreasing the number of EC2 instances as needed to maintain the desired performance level. You can configure EC2 Auto Scaling to scale resources based on a variety of metrics, such as CPU utilization or network bandwidth, and you can specify minimum and maximum resource limits to ensure that your application has sufficient resources to meet its needs while staying within your budget.

EC2 Auto Scaling is a flexible and powerful tool that can help you optimize the performance and cost of your application. It can be especially useful for applications with variable or unpredictable workloads, as it allows you to scale resources up or down as needed to meet changing demands.

To use EC2 Auto Scaling, you will need to create an Auto Scaling group and define the minimum and a maximum number of [EC2](https://datavalley.ai/launch-ec2-instance-custom-ami-aws-datavalley/) instances that you want to have available, as well as the desired capacity and any scaling policies or rules that you want to use to control the scaling process. EC2 Auto Scaling will then monitor your application and automatically adjust the number of instances as needed to meet your specified performance criteria.

### 12. What is the Pricing for Amazon [EC2](https://www.datavalley.ai/launch-ec2-instance-custom-ami-aws-datacademy/)Auto Scaling?

There is no additional charge for using Amazon Elastic Compute Cloud (EC2) Auto Scaling. You will only pay for the EC2 instances that you use, as well as any other resources that your application consumes, such as Elastic Load Balancers or Amazon Elastic Block Store (EBS) volumes.

The cost of your [EC2](https://www.datavalley.ai/launch-ec2-instance-custom-ami-aws-datacademy/)instances will depend on the instance type, the region in which they are deployed, and the length of time that they are running. You can choose from a variety of instance types, each with different combinations of CPU, memory, and storage resources, to meet the needs of your application. You can also choose to use on-demand instances or reserved instances, depending on your needs and budget. On-demand instances are charged based on the number of hours that they are running, while reserved instances offer a discounted rate in exchange for a commitment to use the instances for a specified period of time.

You can use the EC2 Pricing Calculator to estimate the cost of your [EC2](https://www.datavalley.ai/launch-ec2-instance-custom-ami-aws-datacademy/)instances and other resources based on your specific configuration and usage patterns. This can help you plan your budget and ensure that you have the resources you need to meet the demands of your application, without overspending.

### 13. How do Auto Scaling groups work?

You can create, access, and manage your Auto Scaling groups using any of the following interfaces:

* **AWS Management Console** – Provides a web interface that you can use to access your Auto Scaling groups. If you’ve signed up for an AWS account, you can access your Auto Scaling groups by signing into the AWS Management Console, using the search box on the navigation bar to search for **Auto Scaling groups**, and then choose **Auto Scaling groups**.
* **AWS Command Line Interface (AWS CLI)** – Provides commands for a broad set of AWS services, and is supported on Windows, macOS, and Linux. To get started.
* **AWS Tools for Windows PowerShell** – Provides commands for a broad set of AWS products for those who script in the PowerShell environment.
* **AWS SDKs** – Provides language-specific API operations and takes care of many of the connection details, such as calculating signatures, handling request retries, and handling errors.
* **Query API** – Provides low-level API actions that you call using HTTPS requests. Using the Query API is the most direct way to access AWS services. However, it requires your application to handle low-level details such as generating the hash to sign the request and handling errors.
* **AWS CloudFormation** – Supports creating Auto Scaling groups using CloudFormation templates.

To connect programmatically to an AWS service, you use an endpoint.

### 14. What are Auto Scaling components?

Auto-scaling components are the building blocks that are used to create and manage an Amazon Web Services (AWS) Auto Scaling system. These components include:

1. **Launch configurations:**A launch configuration is a template that specifies the properties of the Amazon Elastic Compute [Cloud](https://www.datavalley.ai/aws-cloudfront-interview-questions/)(EC2) instances or Amazon Elastic Container Service (ECS) tasks that you want to include in your auto-scaling group. It specifies the instance type, the AMI to use, the security groups and key pairs to use, and any other necessary configuration details.
2. **Auto-scaling groups:**An auto-scaling group (ASG) is a collection of EC2 instances or ECS tasks that are configured to run your application, along with the policies and rules that control how the group scales in response to changes in demand.
3. **Scaling policies:**A scaling policy defines how an ASG should scale in response to changes in demand. It specifies the metric to use to trigger a scale-out or scale-in event, as well as the number of instances or tasks to add or remove in response to the event.
4. **CloudWatch alarms:**CloudWatch alarms are used to monitor the performance of your application and trigger scaling events based on predefined thresholds. For example, you could create an alarm that triggers a scale-out event when the average CPU utilization of your instances exceeds 80 percent.
5. **Amazon EC2 instances or Amazon ECS tasks:** These are the resources that are used to run your application and are managed by the ASG. They can be EC2 instances or ECS tasks, depending on your needs and preferences.

By using these components together, you can create a scalable, highly available system that can handle changes in demand and ensure that your application has the resources it needs to meet the needs of your users.

### 15. What are Auto-Scaling Related Services?

Amazon Web Services (AWS) Auto Scaling is a service that enables you to automatically scale the capacity of your Amazon Elastic Compute Cloud ([EC2](https://www.datavalley.ai/launch-ec2-instance-custom-ami-aws-datacademy/)) instances or Amazon Elastic Container Service (ECS) tasks up or down in response to changes in demand for your application. There are several other AWS services that are commonly used in conjunction with Auto Scaling to create a scalable and highly available system:

1. **Amazon Elastic Load Balancer (ELB):**ELB is a service that automatically distributes incoming traffic across multiple EC2 instances or ECS tasks. It can help you improve the availability and performance of your application by distributing traffic evenly across multiple instances and automatically replacing any instances that fail or become unavailable.
2. **Amazon Elastic Block Store (EBS):**EBS is a service that provides persistent block storage for EC2 instances. It can be used to store data that needs to be persisted across instance restarts, such as database files or application logs.
3. **Amazon CloudWatch:**CloudWatch is a service that enables you to monitor the performance of your AWS resources in real time. It can be used to set alarms that trigger scaling events based on predefined thresholds, such as when the average CPU utilization of your instances exceeds a certain percentage.
4. **Amazon Simple Notification Service (SNS):** SNS is a service that enables you to send notifications when certain events occur within your AWS resources. It can be used to send notifications when an instance is added or removed from an Auto Scaling group, or when an alarm is triggered.

By using these services in conjunction with AWS Auto Scaling, you can create a scalable and highly available system that can handle changes in demand and ensure that your application has the resources it needs to meet the needs of your users.

## FAQ’s

### What is an auto scaling in AWS?

Auto Scaling is a feature in Amazon Web Services (AWS) that automatically increases or decreases the number of Amazon Elastic Compute Cloud (EC2) instances in a configuration based on predefined policies and rules. The goal of Auto Scaling is to ensure that your application has the right amount of resources available to handle changes in demand, without manual intervention.

Auto Scaling works by creating a group of Amazon EC2 instances, called an Auto Scaling group, that are created from a common Amazon Machine Image (AMI). You can then define rules and policies for when instances should be added or removed from the group. For example, you might define a policy that says to add an instance to the group when the average network latency exceeds a certain threshold, or to remove an instance when the average CPU usage falls below a certain level.

There are several benefits to using Auto Scaling:

1.Cost Savings: By adding or removing instances based on demand, Auto Scaling can help you reduce the amount of resources you need to pay for when demand is low and avoid running out of resources when demand is high.

2.High Availability: Auto Scaling can help you ensure that your application is always available by automatically replacing instances that fail or become unavailable.

3.Scalability: Auto Scaling can help you scale your application up and down automatically, so that you can handle changes in traffic and load.

Overall, Auto Scaling provides a simple and powerful way to ensure that your application has the resources it needs to handle changes in demand, without manual intervention.

### What are the two main components of AWS Auto Scaling?

The two main components of AWS Auto Scaling are the Auto Scaling group and the launch configuration.  
1. Auto Scaling group: This is the core component of Auto Scaling. It defines the group of Amazon EC2 instances that are created from a common Amazon Machine Image (AMI). It also defines the rules and policies for when instances should be added or removed from the group.  
2. Launch Configuration: It is a blueprint that describes all the settings for an instance. It contains the AMI to be used, the instance type, key pair and security group. It is essentially a template that tells Auto Scaling how to launch and configure the instances in the group.

### What is Auto Scaling in AWS interview questions?

Here are a few Auto Scaling interview questions:  
1. Can you explain what Auto Scaling is and what its main benefits are?  
2. What are the two main components of AWS Auto Scaling?  
3. How does Auto Scaling work in relation to Amazon Elastic Block Store (EBS) volumes?  
4. Can you give an example of when you would use Auto Scaling?  
5. How does Auto Scaling help with cost optimization?

### What is the difference between auto scaling and ELB?

Here is a table that compares Auto Scaling and Elastic Load Balancing (ELB) in terms of their features and use cases:

## Interview Questions on Amazon RDS

Here are some most commonly asked certification interview question related to [Amazon RDS](https://www.datavalley.ai/interview-questions-on-amazon-rds/)

### 1. What is Amazon RDS?

A: [Amazon RDS](https://datavalley.ai/interview-questions-on-amazon-rds/) (Relational Database Service) is a managed service offered by (Amazon Web Services) that enables users to easily set up, operate and scale a relational database in the cloud. It supports popular relational databases like MySQL, PostgreSQL, Oracle, SQL Server, MariaDB, and Amazon Aurora.

### 2. How does Amazon RDS differ from Amazon EC2?

A: [Amazon EC2](https://www.datavalley.ai/interview-questions-on-aws-elastic-compute-cloud/) is a virtual machine service that provides users with the ability to launch and manage their own instances of operating systems and software applications. Amazon RDS, on the other hand, is a fully managed service that allows users to launch and operate a database in the cloud without worrying about infrastructure and management tasks.

### 3. What are the benefits of using Amazon RDS?

A: The benefits of using [Amazon RDS](https://datavalley.ai/interview-questions-on-amazon-rds/) include easy database management and scaling, automatic backups and software updates, high availability, and security. Additionally, Amazon RDS offers cost savings, as users only pay for the resources they use.

### 4. Which database engines does Amazon RDS support?

A: Amazon RDS supports popular relational database engines, including [MySQL](https://datavalley.ai/spark-sql-creating-databases-and-tables/), PostgreSQL, Oracle, SQL Server, MariaDB, and Amazon Aurora.

### 5. What is Amazon Aurora, and how does it differ from other database engines?

A: [Amazon Aurora](https://datavalley.ai/aws-vpc-interview-questions/) is a [MySQL](https://datavalley.ai/what-is-mysql/)and PostgreSQL-compatible relational database engine that is designed to be highly scalable and available. It is built on top of Amazon’s cloud infrastructure and is designed to deliver high performance, reliability, and availability at a lower cost than traditional commercial databases.

### 6. Can you configure automatic backups in Amazon RDS?

A: Yes, Amazon RDS enables users to configure automatic backups of their databases. They can choose between full backups and transaction logs backups, and set up retention periods for the backups.

### 7. How can you improve the performance of a database in Amazon RDS?

A: To improve the performance of a database in Amazon RDS, users can scale the instance size, enable read replicas, optimize database queries and indexes, and use caching technologies like Amazon ElastiCache.

### 8. How does Amazon RDS ensure high availability?

A: [Amazon RDS](https://datavalley.ai/interview-questions-on-amazon-rds/) ensures high availability through its Multi-AZ deployment option. In this deployment, Amazon RDS automatically creates a standby replica of the database in a different availability zone, which is constantly synchronized with the primary database. If a failure occurs in the primary database, Amazon RDS automatically fails over to the standby replica, ensuring high availability[.](https://datavalley.ai/)

### 9. How can you secure a database in Amazon RDS?

A: To secure a database in Amazon RDS, users can implement security best practices like using strong passwords, encrypting data at rest and in transit, configuring network security, and monitoring the database for suspicious activity using Amazon Web Services CloudTrail and Amazon CloudWatch.

### 10. What is the maximum storage size limit for an Amazon RDS PostgreSQL database?

A: The maximum storage size limit for an Amazon RDS PostgreSQL database is 64TB.

### 11. How can you scale a PostgreSQL database in Amazon RDS?

A: To scale a [PostgreSQL](https://datavalley.ai/interview-questions-on-amazon-aurora/)database in [Amazon RDS](https://datavalley.ai/interview-questions-on-amazon-rds/), users can choose to increase the size of the instance or add read replicas. Additionally, they can use Amazon Aurora PostgreSQL as an alternative, which offers higher performance and scalability.

### 12. Can you use custom PostgreSQL extensions with Amazon RDS?

A: Yes, users can use custom PostgreSQL extensions with [Amazon RDS](https://datavalley.ai/interview-questions-on-amazon-rds/), as long as they are compatible with the version of PostgreSQL running in Amazon RDS and do not require superuser privileges to install or configure.

### 13. What is the difference between a snapshot and a backup in Amazon RDS PostgreSQL?

A: A snapshot is a point-in-time copy of the database, while a backup is a copy of the database’s data files. Snapshots are used for disaster recovery and are stored in [Amazon S3](https://datavalley.ai/aws-certified-data-analytics-specialty-aws-data/), while backups are used for replication and are stored on the same instance as the database.

### 14. Can you encrypt data in transit and at rest in Amazon RDS PostgreSQL?

A: Yes, users can encrypt data in transit and at rest in Amazon RDS PostgreSQL. For data in transit, users can use SSL/TLS encryption. For data at rest, users can choose to use the default encryption provided by Amazon RDS or use their own encryption keys.

[](https://youtube.com/@datavalley-ai?sub_confirmation=1)

### 15. Can you configure automated backups for an Amazon RDS PostgreSQL database?

A: Yes, users can configure automated backups for an Amazon RDS PostgreSQL database. They can specify the backup window, retention period, and whether to enable multi-AZ deployments for backups.

### 16. What is a read replica in Amazon RDS PostgreSQL?

A: A read replica is a copy of the primary database that is used for read-heavy workloads. It allows users to offload read traffic from the primary database, improving performance and scalability. Read replicas in [Amazon RDS](https://datavalley.ai/interview-questions-on-amazon-rds/)PostgreSQL use asynchronous replication.

### 17. What is the purpose of the parameter group in Amazon RDS PostgreSQL?

A: The parameter group in Amazon RDS PostgreSQL allows users to manage the configuration of the database instance. It contains settings like connection limits, buffer sizes, and log settings. Users can modify the parameter group to optimize the performance and functionality of their PostgreSQL database instance.

### 18. What is the maximum storage size limit for an Amazon RDS MySQL database?

A: The maximum storage size limit for an Amazon RDS MySQL database is 64TB.

### 19. Can you use custom MySQL extensions with Amazon RDS?

A: Yes, users can use custom MySQL extensions with Amazon RDS, as long as they are compatible with the version of MySQL running in Amazon RDS.

### 20. How can you scale a MySQL database in Amazon RDS?

A: To scale a MySQL database in Amazon RDS, users can choose to increase the size of the instance or add read replicas. Additionally, they can use Amazon Aurora MySQL as an alternative, which offers higher performance and scalability.

### 21. Can you encrypt data in transit and at rest in Amazon RDS MySQL?

A: Yes, users can encrypt data in transit and at rest in Amazon RDS [MySQL](https://www.datavalley.ai/nodejs-mysql-tutorial/). For data in transit, users can use SSL/TLS encryption. For data at rest, users can choose to use the default encryption provided by Amazon RDS or use their own encryption keys.

### 22. Can you configure automated backups for an Amazon RDS MySQL database?

A: Yes, users can configure automated backups for an [Amazon](https://datavalley.ai/aws-certified-data-analytics-specialty-amazon-s3/)RDS MySQL database. They can specify the backup window, retention period, and whether to enable multi-AZ deployments for backups.

### 23. What is a read replica in Amazon RDS MySQL?

A: A read replica is a copy of the primary database that is used for read-heavy workloads. It allows users to offload read traffic from the primary database, improving performance and scalability. Read replicas in[Amazon RDS](https://www.datavalley.ai/aws-rds-interview-questions-and-answers/) MySQL use either asynchronous or semi-synchronous replication.

### 24. What is the purpose of the parameter group in Amazon RDS MySQL?

A: The parameter group in Amazon RDS MySQL allows users to manage the configuration of the database instance. It contains settings like connection limits, buffer sizes, and log settings. Users can modify the parameter group to optimize the performance and functionality of their [MySQL](https://www.datavalley.ai/what-is-mysql/) database instance.

### 25. How can you ensure high availability for an Amazon RDS MySQL database?

A: To ensure high availability for an Amazon RDS MySQL database, users can enable multi-AZ deployments. In a multi-AZ deployment, Amazon RDS automatically creates a standby replica of the database in a different availability zone, which is constantly synchronized with the primary database. If a failure occurs in the primary database, Amazon RDS automatically fails over to the standby replica, ensuring high availability.

### 26. What is the maximum storage size limit for an Amazon RDS MariaDB database?

A: The maximum storage size limit for an [Amazon RDS](https://www.datavalley.ai/aws-rds-interview-questions-and-answers/) MariaDB database is 64TB.

### 27. Can you use custom MariaDB extensions with Amazon RDS?

A: Yes, users can use custom MariaDB extensions with Amazon RDS, as long as they are compatible with the version of MariaDB running in Amazon RDS.

### 28. How can you scale a MariaDB database in Amazon RDS?

A: To scale a MariaDB database in Amazon RDS, users can choose to increase the size of the instance or add read replicas. Additionally, they can use Amazon Aurora with MariaDB compatibility as an alternative, which offers higher performance and scalability.

### 29. Can you encrypt data in transit and at rest in Amazon RDS MariaDB?

A: Yes, users can encrypt data in transit and at rest in Amazon RDS MariaDB. For data in transit, users can use SSL/TLS encryption. For data at rest, users can choose to use the default encryption provided by Amazon RDS or use their own encryption keys.

### 30. Can you configure automated backups for an Amazon RDS MariaDB database?

A: Yes, users can configure automated backups for an Amazon RDS MariaDB database. They can specify the backup window, retention period, and whether to enable multi-AZ deployments for backups.

### 31. What is a read replica in Amazon RDS MariaDB?

A: A read replica is a copy of the primary database that is used for read-heavy workloads. It allows users to offload read traffic from the primary database, improving performance and scalability. Read replicas in Amazon RDS MariaDB use either asynchronous or semi-synchronous replication.

### 32. What is the purpose of the parameter group in Amazon RDS MariaDB?

A: The parameter group in Amazon RDS MariaDB allows users to manage the configuration of the database instance. It contains settings like connection limits, buffer sizes, and log settings. Users can modify the parameter group to optimize the performance and functionality of their MariaDB database instance.

### 33. How can you ensure high availability for an Amazon RDS MariaDB database?

A: To ensure high availability for an Amazon RDS MariaDB database, users can enable multi-AZ deployments. In a multi-AZ deployment, Amazon RDS automatically creates a standby replica of the database in a different availability zone, which is constantly synchronized with the primary database. If a failure occurs in the primary database, Amazon RDS automatically fails over to the standby replica, ensuring high availability.

[](https://youtube.com/@datavalley-ai?sub_confirmation=1)

### 34. What is the maximum storage size limit for an Amazon RDS SQL Server database?

A: The maximum storage size limit for an [Amazon RDS SQL](https://datavalley.ai/interview-questions-on-amazon-rds/) Server database is 16TB.

### 35. Can you use custom SQL Server extensions with Amazon RDS?

A: Yes, users can use custom SQL Server extensions with Amazon RDS, as long as they are compatible with the version of SQL Server running in Amazon RDS.

### 36. How can you scale a SQL Server database in Amazon RDS?

A: To scale a SQL Server database in Amazon RDS, users can choose to increase the size of the instance or add read replicas. Additionally, they can use Amazon Aurora with SQL Server compatibility as an alternative, which offers higher performance and scalability.

### 37. Can you encrypt data in transit and at rest in Amazon RDS SQL Server?

A: Yes, users can encrypt data in transit and at rest in Amazon RDS [SQL Server](https://www.datavalley.ai/sql-server-interview-questions/). For data in transit, users can use SSL/TLS encryption. For data at rest, users can choose to use the default encryption provided by Amazon RDS or use their own encryption keys.

### 38. Can you configure automated backups for an Amazon RDS SQL Server database?

A: Yes, users can configure automated backups for an [Amazon](https://datavalley.ai/big-data-analytics-using-aws-emr-athena/) RDS SQL Server database. They can specify the backup window, retention period, and whether to enable multi-AZ deployments for backups.

### 39. What is a read replica in Amazon RDS SQL Server?

A: A read replica is a copy of the primary database that is used for read-heavy workloads. It allows users to offload read traffic from the primary database, improving performance and scalability. Read replicas in Amazon RDS SQL Server use asynchronous replication.

### 40. What is the purpose of the parameter group in Amazon RDS SQL Server?

A: The parameter group in Amazon RDS [SQL Server](https://www.datavalley.ai/what-is-sql)allows users to manage the configuration of the database instance. It contains settings like connection limits, buffer sizes, and log settings. Users can modify the parameter group to optimize the performance and functionality of their SQL Server database instance.

### 41. How can you ensure high availability for an Amazon RDS SQL Server database?

A: To ensure high availability for an [Amazon RDS](https://datavalley.ai/interview-questions-on-amazon-rds/) SQL Server database, users can enable multi-AZ deployments. In a multi-AZ deployment, Amazon RDS automatically creates a standby replica of the database in a different availability zone, which is constantly synchronized with the primary database. If a failure occurs in the primary database, Amazon RDS automatically fails over to the standby replica, ensuring high availability.

### 42. What is the maximum storage size limit for an Amazon RDS Oracle database?

A: The maximum storage size limit for an Amazon RDS Oracle database is 64TB.

### 43. Can you use custom Oracle extensions with Amazon RDS?

A: Yes, users can use custom Oracle extensions with Amazon RDS, as long as they are compatible with the version of Oracle running in Amazon RDS.

### 44. How can you scale an Oracle database in Amazon RDS?

A: To scale an Oracle database in Amazon RDS, users can choose to increase the size of the instance or add read replicas. Additionally, they can use Amazon Aurora with Oracle compatibility as an alternative, which offers higher performance and scalability.

### 45. Can you encrypt data in transit and at rest in Amazon RDS Oracle?

A: Yes, users can encrypt data in transit and at rest in Amazon RDS Oracle. For data in transit, users can use SSL/TLS encryption. For data at rest, users can choose to use the default encryption provided by Amazon RDS or use their own encryption keys.

### 46. Can you configure automated backups for an Amazon RDS Oracle database?

A: Yes, users can configure automated backups for an Amazon RDS Oracle database. They can specify the backup window, retention period, and whether to enable multi-AZ deployments for backups.

What is a read replica in Amazon RDS Oracle? A: A read replica is a copy of the primary database that is used for read-heavy workloads. It allows users to offload read traffic from the primary database, improving performance and scalability. Read replicas in Amazon RDS Oracle use either asynchronous or semi-synchronous replication.

### 47. What is the purpose of the parameter group in Amazon RDS Oracle?

A: The parameter group in Amazon RDS Oracle allows users to manage the configuration of the database instance. It contains settings like connection limits, buffer sizes, and log settings. Users can modify the parameter group to optimize the performance and functionality of their Oracle database instance.

### 48. How can you ensure high availability for an Amazon RDS Oracle database?

A: To ensure high availability for an Amazon RDS Oracle database, users can enable multi-AZ deployments. In a multi-AZ deployment, Amazon RDS automatically creates a standby replica of the database in a different availability zone, which is constantly synchronized with the primary database. If a failure occurs in the primary database, Amazon RDS automatically fails over to the standby replica, ensuring high availability.

### ****1. What is AWS Lambda?****

Amazon Web Services (AWS) Lambda is a serverless compute service that runs your code in response to events and automatically manages the underlying compute resources for you. With [Lambda](https://www.datavalley.ai/aws-lambda-interview-questions/), you can run code for virtually any type of application or backend service, all with zero administration.

Lambda automatically scales and monitors your applications, so you don’t have to. You only pay for the compute time that you consume. This makes it an attractive option for building and running microservices, data processing jobs, and real-time streaming applications.

### ****2. What is Auto Scaling in Lambda?****

It is one of the features of AWS that helps one to automatically configure and spin novel instances. One of the best things about AWS Auto Scaling is that it does not require any interference at any stage. However, users can monitor everything through metrics and thresholds. To enable this task, one needs to cross a threshold, and without interference, one can see the instances scaled horizontally.

### ****3. What are the languages supported by AWS Lambda?****

Amazon Web Services ([AWS](https://www.datavalley.ai/aws-interview-questions/)) Lambda supports the following languages:

1. Node.js
2. Java
3. C#
4. Go
5. PowerShell
6. Python
7. Ruby
8. .NET Core

You can use any of these languages to write your code for AWS Lambda. You can also use the AWS Serverless Application Model (SAM) to build and deploy your serverless applications in any of these languages.

### ****4. While performing DDOS, what is the limit for execution in Lambda?****

You can now configure your AWS Lambda functions to run up to 15 minutes per execution. Previously, the maximum execution time (timeout) for a Lambda function was 5 minutes. Now, it is easier than ever to perform big data analysis, bulk data transformation, batch event processing, and statistical computations using longer-running functions.

* You can now set the timeout value for a function to any value up to 15 minutes. When the specified timeout is reached, [AWS](https://www.datavalley.ai/aws-interview-questions/)Lambda terminates the execution of your Lambda function.
* As a best practice, you should set the timeout value based on your expected execution time to prevent your function from running longer than intended.

### ****5. What do you think makes Lambda a timesaving approach?****

There can be a number of reasons behind this. One of which is that Lambda stores everything in local server memory.  Another reason can be that data is stored directly in the database without it affecting performance. In addition to these features, Lambda also has simple testing techniques; for example, integration testing can be made powerful through multiple vendors.

### ****6. What are the best practices for security in Lambda?****

Using AWS IAM (Identity Access and Management) is one of the widely used security practices in Lambda. Granting specific user access to particular roles is another effective option to enhance security. In this security practice, access can be restricted to hosts that are not trusted or authorized. In addition to this, no matter how effective and stringent the security protocols are, they should always be updated timely.

### ****7. What are the limitations of****[AWS](https://www.datavalley.ai/aws-interview-questions/)****Lambda?****

#### **Some of the limitations of AWS Lambda are as follows:**

* The default deployment package size is 50 MB.
* The ephemeral disk space is limited to 512 MB as the Lambda function will take a longer time to execute with a larger package size.
* The execution time is more when the memory allocation is less.
* The memory range is from 128 MB to 10,240 MB.
* The maximum execution timeout for a function is 15 minutes.

### ****8. What is elastic blockage storage in Lambda?****

Amazon’s elastic block storage (EBS) can be defined as a virtual storage area network where tasks can be started. It can tolerate faults easily, and users need not worry about loss of data even if the disk is damaged in the RAID. Provisioning and allocating the storage can also be done in EBS. If required, it can also be connected to an API.

### ****9. How does Lambda handle failure during event processing?****

If an AWS Lambda function fails during event processing, the function is retried. If the function continues to fail after multiple retries, it will be considered a failed execution.

By default, Lambda will retry the function for up to two times, with a delay between retries. You can configure the number of retries and the delay between retries using the AWS Management Console, the [AWS](https://www.datavalley.ai/aws-interview-questions/)CLI, or the AWS SDKs.

If the function continues to fail after the maximum number of retries, it will be considered a failed execution. You can set up notifications or alarms to be triggered when a function fails, so that you can take appropriate action to troubleshoot the issue.

### ****10. Is vertical scaling possible in Lambda?****

Yes, vertical scaling is possible in Lambda. Vertical scaling is one of the most in-demand features of AWS Lambda. This feature is generally used when a user needs to spin a larger instance. If in case, they are already using an instance, it can be paused and detached from the server. In this process, it is important to note the ID of the new device post that can continue the process.

## AWS Lambda Interview Questions

### ****11. How is a Lambda function executed?****

An [AWS](https://www.datavalley.ai/aws-vpc-interview-questions/)Lambda function is executed in response to an event. When the event occurs, Lambda assigns the function to a compute resource and begins executing the function’s code.

The function is executed in a stateless environment, which means that it does not maintain any persistent state between executions. This allows Lambda to scale the function up or down based on demand, and to efficiently allocate compute resources to handle the workload.

After the function has completed executing, the compute resources are released and the function goes back into a dormant state until the next time it is needed. This allows you to run your code without having to worry about managing the underlying infrastructure.

### ****12. Name a simple method to improve performance in AWS Lambda.****

Performance in AWS Lambda can be simply improved by using RAID, the Linux software. It also helps in increasing security.

### ****13. In how many ways can AWS Lambda be used?****

#### **One can use Lambda in the following ways:**

As an event-driven computing service, AWS [Lambda](https://datavalley.ai/aws-lambda-interview-questions/) runs code in response to events.

These events can be the changes to data in an [Amazon S3](https://www.datavalley.ai/aws-s3-interview-question/) bucket or [AWS](https://www.datavalley.ai/aws-interview-questions/) [DynamoDB](https://datavalley.ai/interview-questions-amazon-dynamodb/)

Lambda can run code in response to HTTP requests using Amazon API Gateway or API calls made using AWS SDKs.

### ****14. How does AWS Lambda secure my code?****

Lambda secures the code by encrypting it. Lambda stores the code in [Amazon S3](https://datavalley.ai/understanding-amazon-s3amazon-web-services/) buckets and encrypts it when it is resting. Further, AWS Lambda performs an additional integrity check on the code while it is running.

## Advanced AWS Lambda Interview Questions

### ****15. What are the restrictions applied to the AWS Lambda function code?****

Although AWS Lambda has very few restrictions on various kinds of operating systems activities and standard programming languages, there are a few restrictions on activities that disable instances and trace calls and inbound network connections. In addition to this, [AWS](https://www.datavalley.ai/aws-cloudfront-interview-questions/)Lambda also disables activities such as debugging systems and TCP ports. For outbound data connection, TCP or IP sockets are very supportive.

**16. How to get started with a serverless application?**

In order to get started with a serverless application, a user needs to console AWS Lambda and download the blueprint. The original file that will be downloaded should have an AWS Sam file, also known as AWS resource in the application, and a ZIP file. [AWS Cloud](https://www.datavalley.ai/aws-cloudtrail-interview-questions/) formation commands can be used for packaging and deploying serverless application codes and documentation can also be performed.

### ****17. What are the disadvantages of using the serverless approach?****

Everything in [AWS](https://www.datavalley.ai/aws-ami-interview-questions/)Lambda comes with its own merits and demerits depending on the task performed. The upper limit is strictly on the vendor control in the serverless approach and offers more downtime. Sometimes there is a loss of system functionality and the system’s limits are other issues; no dedicated hardware is available for AWS serverless approach. In most cases, customer errors can also give rise to problems.

### ****18. What is the difference between an anonymous class and the Lambda function?****

One of the biggest differences between an anonymous class and the Lambda function is the use of keywords. While the keywords in a Lambda function are used to resolve functional classes, the keywords in anonymous classes are used to resolve anonymous functional classes.

### ****19. Is Lambda Expression a nameless suspension of code?****

Yes, a lambda expression is a nameless suspension of code that can be treated as a value and passed around as a parameter or stored in a variable. It is a way to define a block of code as an anonymous function, which can be executed later.

Lambda expressions are a feature of functional programming languages, and they are often used to define callback functions or to create functional interfaces. In Java, for example, a lambda expression is a concise way to represent a functional interface (an interface with a single abstract method) as an instance.

Lambda expressions are a powerful and concise way to define and execute code, and they are widely used in modern programming languages such as Java, Python, and C#.

### ****20. What kind of code can run on AWS Lambda?****

There are many activities that can be performed on [AWS Lambda](https://datavalley.ai/aws-lambda-interview-questions/); for example, it can be used to build mobile backends from Amazon DynamoDB to retrieve and transform data. Handlers transform and compress objects as they get uploaded to Amazon S3. This is done by using Amazon Kinesis, serverless processing of streaming data. The reporting and auditing of API calls can be made to any web service of Amazon, and many other activities can also be done with the help of AWS Lambda.

### ****21. What are final variables and effectively final variables in Lambda?****

Final variables are those that cannot be modified once assigned, while effectively final variables can be changed as they are in their earlier stage and have not been assigned a value. Effectively final variables play a huge role in testing as well. If final variables are to be equipped with several additional features, this can be done through effectively final variables.

### ****22. How does AWS Lambda work?****

Many people consider AWS Lambda to be a little confusing, but it is not. It is just a simple four-step process that begins with uploading code to AWS Lambda. The next step is to set up your code to trigger from other AWS services, HTTP endpoints, or mobile apps. AWS Lambda will only run a code when it is triggered and will only use the computing resources needed to run it.

### ****23. What can one build with AWS Lambda?****

#### **The following is the list of what can one build with**[AWS](https://www.datavalley.ai/aws-ami-interview-questions/)**Lambda:**

* Real-time file processing
* Sorting real-time stream processing
* Data processing
* Data validation
* Filtering
* Third-party API requests

### ****24. Do Lambda-based functions stay available after the code or configuration is changed?****

Yes, Lambda-based functions remain available after the code or configuration is changed. When a Lambda function is updated, there is a brief period, less than a minute, when requests can be served by either the old or new version of the function.

# **What is AWS Backup?**

[**PDF**](https://docs.aws.amazon.com/pdfs/aws-backup/latest/devguide/AWSBackup-dg.pdf#whatisbackup)[**RSS**](https://docs.aws.amazon.com/aws-backup/latest/devguide/backup-dg-rss-updates.rss)

AWS Backup is a fully-managed service that makes it easy to centralize and automate data protection across AWS services, in the cloud, and on premises. Using this service, you can configure backup policies and monitor activity for your AWS resources in one place. It allows you to automate and consolidate backup tasks that were previously performed service-by-service, and removes the need to create custom scripts and manual processes. With a few clicks in the AWS Backup console, you can automate your data protection policies and schedules.

AWS Backup does not govern backups you take in your AWS environment outside of AWS Backup. Therefore, if you want a centralized, end-to-end solution for business and regulatory compliance requirements, start using AWS Backup today.

## Feature overview

AWS Backup provides many features and capabilities, including the following.

### Centralized backup management

AWS Backup provides a centralized backup console, a set of backup APIs, and the AWS Command Line Interface (AWS CLI) to manage backups across the AWS services that your applications use. With AWS Backup, you can centrally manage backup policies that meet your backup requirements. You can then apply them to your AWS resources across AWS services, enabling you to back up your application data in a consistent and compliant manner. The AWS Backup centralized backup console offers a consolidated view of your backups and backup activity logs, making it easier to audit your backups and ensure compliance.

### Policy-based backup

With AWS Backup, you can create backup policies known as backup plans. Use these backup plans to define your backup requirements and then apply them to the AWS resources that you want to protect across the AWS services that you use. You can create separate backup plans that each meet specific business and regulatory compliance requirements. This helps ensure that each AWS resource is backed up according to your requirements. Backup plans make it easy to enforce your backup strategy across your organization and across your applications in a scalable manner.

For all the configuration options for backup plans, see [Backup plan options and configuration](https://docs.aws.amazon.com/aws-backup/latest/devguide/creating-a-backup-plan.html#plan-options-and-configuration).

### Tag-based backup policies

You can use AWS Backup to apply backup plans to your AWS resources in a wide variety of ways, including tagging them. Tagging makes it easier to implement your backup strategy across all your applications and to ensure that all your AWS resources are backed up and protected. AWS tags are a great way to organize and classify your AWS resources. Integration with AWS tags enables you to quickly apply a backup plan to a group of AWS resources, so that they are backed up in a consistent and compliant manner.

For all the ways you can assign your resources to backup plans, see [Assigning resources to a backup plan](https://docs.aws.amazon.com/aws-backup/latest/devguide/assigning-resources.html).

### Lifecycle management policies

AWS Backup enables you to meet compliance requirements while minimizing backup storage costs by storing backups in a low-cost cold storage tier. You can configure lifecycle policies that automatically transition backups from warm storage to cold storage according to a schedule that you define.

For a list of resources which can be transitioned to cold storage, see [Feature availability by resource](https://docs.aws.amazon.com/aws-backup/latest/devguide/backup-feature-availability.html#features-by-resource). For steps to turn on cold storage in your backup plan, see [Lifecycle and storage tiers](https://docs.aws.amazon.com/aws-backup/latest/devguide/creating-a-backup-plan.html#backup-lifecycle).

### Cross-Region backup

Using AWS Backup, you can copy backups to multiple different AWS Regions on demand or automatically as part of a scheduled backup plan. Cross-Region backup is particularly valuable if you have business continuity or compliance requirements to store backups a minimum distance away from your production data. For more information, see [Creating backup copies across AWS Regions](https://docs.aws.amazon.com/aws-backup/latest/devguide/cross-region-backup.html).

### Cross-account management and cross-account backup

You can use AWS Backup to manage your backups across all AWS accounts inside your [AWS Organizations](https://docs.aws.amazon.com/organizations/latest/userguide/orgs_introduction.html) structure. With cross-account management, you can automatically use backup policies to apply backup plans across the AWS accounts within your organization. This makes compliance and data protection efficient at scale and reduces operational overhead. It also helps eliminate manually duplicating backup plans across individual accounts. For more information, see [Managing AWS Backup resources across multiple AWS accounts](https://docs.aws.amazon.com/aws-backup/latest/devguide/manage-cross-account.html).

You can also copy backups to multiple different AWS accounts inside your AWS Organizations management structure. This way, you can "fan in" backups to a single repository account, then "fan out" backups for greater resilience. [Creating backup copies across AWS accounts](https://docs.aws.amazon.com/aws-backup/latest/devguide/create-cross-account-backup.html).

Before you can use the cross-account management and cross-account backup features, you must have an existing organization structure configured in AWS Organizations. An organizational unit (OU) is a group of accounts that can be managed as a single entity. AWS Organizations is a list of accounts that can be grouped into organizational units and managed as a single entity.

### Auditing and reporting with AWS Backup Audit Manager

AWS Backup Audit Manager helps you simplify data governance and compliance management of your backups across AWS. AWS Backup Audit Manager provides built-in, customizable controls that you can align with your organizational requirements. You can also use these controls to automatically track your backup activities and resources.

AWS Backup Audit Manager can help you locate specific activities and resources that are not yet compliant with the controls that you defined. It also generates daily reports that you can use to demonstrate evidence of compliance with your controls over time.

To include your backup compliance alongside your overall compliance posture, you can automatically import AWS Backup Audit Manager findings into AWS Audit Manager.

### Incremental backups

AWS Backup efficiently stores your periodic backups incrementally. The first backup of an AWS resource backs up a full copy of your data. For each successive incremental backup, only the changes to your AWS resources are backed up. Incremental backups enable you to benefit from the data protection of frequent backups while minimizing storage costs.

For a list of which resources support incremental backups, see [Feature availability by resource](https://docs.aws.amazon.com/aws-backup/latest/devguide/backup-feature-availability.html#features-by-resource).

For more information on behaviors in vaults, see [Incremental backups](https://docs.aws.amazon.com/aws-backup/latest/devguide/creating-a-backup.html#incremental-backup-works).

### Full AWS Backup management

Some resource types support full AWS Backup management. The benefits of full AWS Backup management include:

* **Independent encryption**. AWS Backup automatically encrypts your backups with the KMS key of your AWS Backup vault, instead of using the same encryption key as your source resource. This increases your layers of defense. See [Encryption for backups in AWS Backup](https://docs.aws.amazon.com/aws-backup/latest/devguide/encryption.html) for more information.
* **awsbackup Amazon Resource Names (ARNs)**. Backup ARNs begin with arn:aws:backup instead of arn:aws:*source-resource*. This allows you to create access policies that apply specifically to backups and not the source resources. See [Access control](https://docs.aws.amazon.com/aws-backup/latest/devguide/access-control.html) for more information.
* **Centralized backup billing and Cost Explorer cost allocation tags.**. Charges for AWS Backup (including storage, data transfers, restores, and early deletion) appear under "Backup" in your Amazon Web Services bill, instead of appearing under each supported resource. You can also use Cost Explorer cost allocation tags to track and optimize your backup costs. See [Metering, costs, and billing for AWS Backup](https://docs.aws.amazon.com/aws-backup/latest/devguide/metering-and-billing.html) for more information.

To see which resource types are eligible for full AWS Backup management, see [Feature availability by resource](https://docs.aws.amazon.com/aws-backup/latest/devguide/backup-feature-availability.html#features-by-resource).

### Backup activity monitoring

AWS Backup provides a dashboard that makes it simple to audit backup and restore activity across AWS services. With just a few clicks on the AWS Backup console, you can view the status of recent backup jobs. You can also restore jobs across AWS services to ensure that your AWS resources are properly protected.

AWS Backup integrates with Amazon CloudWatch and Amazon EventBridge. CloudWatch allows you to track metrics and create alarms. EventBridge allows you to view and monitor AWS Backup events. For more information, see [Monitoring AWS Backup events using EventBridge](https://docs.aws.amazon.com/aws-backup/latest/devguide/eventbridge.html) and [Monitoring AWS Backup metrics with CloudWatch](https://docs.aws.amazon.com/aws-backup/latest/devguide/cloudwatch.html).

AWS Backup integrates with AWS CloudTrail. CloudTrail gives you a consolidated view of backup activity logs that make it quick and easy to audit how your resources are backed up. AWS Backup also integrates with Amazon Simple Notification Service (Amazon SNS), providing you with backup activity notifications, such as when a backup succeeds or a restore has been initiated. For more information, see [Logging AWS Backup API calls with CloudTrail](https://docs.aws.amazon.com/aws-backup/latest/devguide/logging-using-cloudtrail.html) and [Using Amazon SNS to track AWS Backup events](https://docs.aws.amazon.com/aws-backup/latest/devguide/sns-notifications.html).

### Secure your data in backup vaults

The content of each AWS Backup backup is immutable, meaning that no one can alter that content. AWS Backup further secures your backups in backup vaults, which separates them safely from their source instances. For example, your vault will retain your Amazon EC2 and Amazon EBS backups according to the lifecycle policy you choose, even if you delete the source Amazon EC2 instance and Amazon EBS volumes.

Backup vaults offer encryption and resource-based access policies that let you define who has access to your backups. You can define access policies for a backup vault that define who has access to the backups within that vault and what actions they can take. This provides a simple and secure way to control access to your backups across AWS services. To review AWS and customer managed policies for AWS Backup, see [Managed policies for AWS Backup](https://docs.aws.amazon.com/aws-backup/latest/devguide/security-iam-awsmanpol.html).

You can use AWS Backup Vault Lock to prevent anyone (including you) from deleting backups or altering their retention period. AWS Backup Vault Lock helps you enforce a write-once-read-many (WORM) model and add another layer of defense to your defense in depth. To get started, see [AWS Backup Vault Lock](https://docs.aws.amazon.com/aws-backup/latest/devguide/vault-lock.html).

## Getting started

To learn more about AWS Backup, we recommend that you start with [Getting started with AWS Backup](https://docs.aws.amazon.com/aws-backup/latest/devguide/getting-started.html).

## Supported AWS resources and applications

The following are AWS resources and third-party applications that you can back up and restore using AWS Backup. For more information, see [AWS Backup feature availability](https://docs.aws.amazon.com/aws-backup/latest/devguide/backup-feature-availability.html).

| **Service** | **Supported resource types** |
| --- | --- |
| [Amazon Elastic Compute Cloud (Amazon EC2)](https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/) | Amazon EC2 instances backed by Amazon EBS volumes |
| [Amazon Simple Storage Service (Amazon S3)](https://docs.aws.amazon.com/AmazonS3/latest/user-guide/) | Amazon S3 data |
| [Amazon Elastic Block Store (Amazon EBS)](https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/AmazonEBS.html) | Amazon EBS volumes |
| [Amazon DynamoDB](https://docs.aws.amazon.com/amazondynamodb/latest/gettingstartedguide/) | Amazon DynamoDB tables |
| [Amazon Relational Database Service (Amazon RDS)](https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/) | Amazon RDS database instances (including all database engines); Multi-Availability Zone clusters |
| [Amazon Aurora](https://docs.aws.amazon.com/AmazonRDS/latest/AuroraUserGuide/) | Aurora clusters |
| [Amazon Elastic File System (Amazon EFS)](https://docs.aws.amazon.com/efs/latest/ug/) | Amazon EFS file systems |
| [FSx for Lustre](https://docs.aws.amazon.com/fsx) | FSx for Lustre file systems |
| [FSx for Windows File Server](https://docs.aws.amazon.com/fsx) | FSx for Windows File Server file systems |
| [Amazon FSx for NetApp ONTAP](https://docs.aws.amazon.com/fsx) | FSx for ONTAP file systems |
| [Amazon FSx for OpenZFS](https://docs.aws.amazon.com/fsx) | FSx for OpenZFS file systems |
| [AWS Storage Gateway (Volume Gateway)](https://docs.aws.amazon.com/storagegateway/latest/userguide/) | AWS Storage Gateway volumes |
| [Amazon DocumentDB](https://docs.aws.amazon.com/documentdb/latest/developerguide/) | Amazon DocumentDB instance-based clusters |
| [Amazon Neptune](https://docs.aws.amazon.com/neptune/latest/userguide/) | Amazon Neptune clusters |
| [Amazon Redshift](https://docs.aws.amazon.com/redshift/latest/dg/welcome.html) | Amazon Redshift clusters |
| [Amazon Timestream](https://docs.aws.amazon.com/timestream/latest/developerguide/what-is-timestream.html) | Amazon Timestream tables |
| [VMware Cloud™ on AWS](https://aws.amazon.com/vmware) | VMware Cloud™ virtual machines on AWS |
| [VMware Cloud™ on AWS Outposts](https://aws.amazon.com/vmware/outposts/) | VMware Cloud™ virtual machines on AWS Outposts |
| [AWS CloudFormation](https://docs.aws.amazon.com/AWSCloudFormation/latest/UserGuide/stacks.html) | AWS CloudFormation stacks |
| [SAP HANA databases](https://docs.aws.amazon.com/aws-backup/latest/devguide/backup-saphana.html) | SAP HANA databases on Amazon EC2 instances |

## Pricing

With AWS Backup, you pay for backup storage, data restored, restore testing, cross-Region data transfer, and AWS Backup Audit Manager. For more information, see [AWS Backup Pricing](https://aws.amazon.com/backup/pricing/)

# **Monitoring**

[**PDF**](https://docs.aws.amazon.com/pdfs/aws-backup/latest/devguide/AWSBackup-dg.pdf#monitoring)[**RSS**](https://docs.aws.amazon.com/aws-backup/latest/devguide/backup-dg-rss-updates.rss)

AWS Backup works with other AWS tools to empower you to monitor its workloads. These tools include the following:

* [AWS Backup console dashboards](https://docs.aws.amazon.com/aws-backup/latest/devguide/backup-dashboards.html)
  + The jobs dashboard brings job health monitoring, where you can view metrics showing job successes and failures, filtered by reasons, accounts, Region, and resource type.
  + The jobs dashboard is available in Regions where AWS Backup Audit Manager is supported. See [Feature availability by AWS Region](https://docs.aws.amazon.com/aws-backup/latest/devguide/backup-feature-availability.html#features-by-region) for those Regions. All other Regions will be able to access the [CloudWatch Dashboard](https://docs.aws.amazon.com/aws-backup/latest/devguide/cloudwatch.html" \l "cloudwatch-dashboard).
* **Amazon CloudWatch** and **Amazon EventBridge** to monitor AWS Backup processes.
  + You can use CloudWatch to track metrics, create alarms, and view dashboards.
  + You can use EventBridge to view and monitor AWS Backup events.

For more information, see [Monitoring AWS Backup events using Amazon EventBridge](https://docs.aws.amazon.com/aws-backup/latest/devguide/eventbridge.html) and .

* **AWS CloudTrail** to monitor AWS Backup API calls. You can identify the time, source IP, users, and accounts making those calls. For more information, see [Logging AWS Backup API calls with CloudTrail](https://docs.aws.amazon.com/aws-backup/latest/devguide/logging-using-cloudtrail.html).
* **Amazon Simple Notification Service** (Amazon SNS) to subscribe to AWS Backup-related topics such as backup, restore, and copy events. For more information, see [Notification options with AWS Backup](https://docs.aws.amazon.com/aws-backup/latest/devguide/backup-notifications.html).

# **Security in AWS Backup**

[**PDF**](https://docs.aws.amazon.com/pdfs/aws-backup/latest/devguide/AWSBackup-dg.pdf#security-considerations)[**RSS**](https://docs.aws.amazon.com/aws-backup/latest/devguide/backup-dg-rss-updates.rss)

Cloud security at AWS is the highest priority. As an AWS customer, you benefit from a data center and network architecture that is built to meet the requirements of the most security-sensitive organizations.

Security is a shared responsibility between AWS and you. The [shared responsibility model](https://aws.amazon.com/compliance/shared-responsibility-model/) describes this as security of the cloud and security in the cloud:

* **Security of the cloud** – AWS is responsible for protecting the infrastructure that runs AWS services in the AWS Cloud. AWS also provides you with services that you can use securely. Third-party auditors regularly test and verify the effectiveness of our security as part of the [AWS compliance programs](https://aws.amazon.com/compliance/programs/). To learn about the compliance programs that apply to AWS Backup, see [AWS Services in Scope by Compliance Program](https://aws.amazon.com/compliance/services-in-scope/).
* **Security in the cloud** – Your responsibility for AWS Backup includes, but is not limited to, the following. You are also responsible for other factors including the sensitivity of your data, your organization's requirements, and applicable laws and regulations.
  + Responding to communications you receive from AWS.
  + Managing the credentials you and your team use. For more information, see [Identity and access management in AWS Backup](https://docs.aws.amazon.com/aws-backup/latest/devguide/backup-iam.html).
  + Configuring your backup plans and resource assignments to reflect your organization’s data protection policies. For more information, see [Managing backup plans](https://docs.aws.amazon.com/aws-backup/latest/devguide/getting-started.html).
  + Regularly testing your ability to find certain recovery points and restore them. For more information, see [Working with backups](https://docs.aws.amazon.com/aws-backup/latest/devguide/recovery-points.html).
  + Incorporating AWS Backup procedures in your organization’s disaster recovery and business continuity written procedures. For a start point, see [Getting started with AWS Backup](https://docs.aws.amazon.com/aws-backup/latest/devguide/getting-started.html).
  + Ensuring that your employees are familiar with and have practiced using AWS Backup along with your organizational procedures in the event of an emergency. For more information, see the [AWS Well-Architected Framework](https://docs.aws.amazon.com/wellarchitected/latest/framework/welcome.html).

This documentation helps you understand how to apply the shared responsibility model when using AWS Backup. The following topics show you how to configure AWS Backup to meet your security and compliance objectives. You also learn how to use other AWS services that help you monitor and secure your AWS Backup resources.

###### Topics

* [Compliance validation](https://docs.aws.amazon.com/aws-backup/latest/devguide/backup-compliance.html)
* [Data protection](https://docs.aws.amazon.com/aws-backup/latest/devguide/data-protection.html)
* [Identity and access management](https://docs.aws.amazon.com/aws-backup/latest/devguide/backup-iam.html)
* [Infrastructure security](https://docs.aws.amazon.com/aws-backup/latest/devguide/infrastructure-security.html)
* [Integrity](https://docs.aws.amazon.com/aws-backup/latest/devguide/backup-integrity.html)
* [Legal holds](https://docs.aws.amazon.com/aws-backup/latest/devguide/legalhold.html)
* [AWS PrivateLink](https://docs.aws.amazon.com/aws-backup/latest/devguide/backup-privatelink.html)
* [Resilience](https://docs.aws.amazon.com/aws-backup/latest/devguide/disaster-recovery-resiliency.html)