Question 1Skipped

What does this AWS CloudFormation snippet do? (Select three)

```
SecurityGroupIngress:
  - IpProtocol: tcp
    FromPort: 80
    ToPort: 80
    CidrIp: 0.0.0.0/0
  - IpProtocol: tcp
    FromPort: 22
    ToPort: 22
    CidrIp: 192.168.1.1/32
```

It only allows the IP 0.0.0.0 to reach HTTP

Correct selection

It allows any IP to pass through on the HTTP port

Correct selection

It lets traffic flow from one IP on port 22

Correct selection

It configures a security group's inbound rules

It prevents traffic from reaching on HTTP unless from the IP 192.168.1.1

It configures the inbound rules of a network access control list (network ACL)

It configures a security group's outbound rules

Overall explanation

Correct options:

It allows any IP to pass through on the HTTP port

It configures a security group's inbound rules

It lets traffic flow from one IP on port 22

A security group acts as a virtual firewall that controls the traffic for one or more instances. When you launch an instance, you can specify one or more security groups; otherwise, we use the default security group. You can add rules to each security group that allows traffic to or from its associated instances. You can modify the rules for a security group at any time; the new rules are automatically applied to all instances that are associated with the security group. When we

decide whether to allow traffic to reach an instance, we evaluate all the rules from all the security groups that are associated with the instance.

The following are the characteristics of security group rules: 1. By default, security groups allow all outbound traffic. 2. Security group rules are always permissive; you can't create rules that deny access. 3. Security groups are stateful

AWS CloudFormation provides a common language for you to model and provision AWS and third-party application resources in your cloud environment. AWS CloudFormation allows you to use programming languages or a simple text file to model and provision, in an automated and secure manner, all the resources needed for your applications across all regions and accounts. This gives you a single source of truth for your AWS and third-party resources.

Considering the given AWS CloudFormation snippet, 0.0.0.0/0 means any IP, not the IP 0.0.0.0. Ingress means traffic going into your instance, and Security Groups are different from NACL. Each "-" in our security group rule represents a different rule (YAML syntax)

Therefore the AWS CloudFormation snippet creates two Security Group inbound rules that allow any IP to pass through on the HTTP port and lets traffic flow from one source IP (192.168.1.1) on port 22.

Incorrect options:

It configures the inbound rules of a network access control list (network ACL) - A Network Access Control List (Network ACL) is an optional layer of security for your VPC that acts as a firewall for controlling traffic in and out of one or more subnets. You might set up network ACLs with rules similar to your security groups to add an additional layer of security to your VPC.

It only allows the IP 0.0.0.0 to reach HTTP

It prevents traffic from reaching on HTTP unless from the IP 192.168.1.1

It configures a security group's outbound rules

These three options contradict the description provided above. So these are incorrect.

References:

https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ec2-security-groups.html

https://aws.amazon.com/cloudformation/

Domain

Design Secure Architectures

Question 2Skipped

A development team has configured Elastic Load Balancing for host-based routing. The idea is to support multiple subdomains and different top-level domains.

The rule *.example.com matches which of the following?

example.com

example.test.com

EXAMPLE.COM

Correct answer

test.example.com

Overall explanation

Correct option:

test.example.com

You can use host conditions to define rules that route requests based on the hostname in the host header (also known as host-based routing). This enables you to support multiple subdomains and different top-level domains using a single load balancer.

A hostname is not case-sensitive, can be up to 128 characters in length, and can contain any of the following characters: 1. A–Z, a–z, 0–9 2. - . 3. * (matches 0 or more characters) 4. ? (matches exactly 1 character)

You must include at least one "." character. You can include only alphabetical characters after the final "." character.

The rule *.example.com matches test.example.com but doesn't match example.com.

Incorrect options:

example.com

example.test.com

EXAMPLE.COM

These three options contradict the explanation provided above, so these options are incorrect.

Reference:

https://docs.aws.amazon.com/elasticloadbalancing/latest/application/load-balancer-listeners.html

Domain

Design Resilient Architectures

Question 3Skipped

A company has noticed that its Amazon EBS Elastic Volume (io1) accounts for 90% of the cost and the remaining 10% cost can be attributed to the Amazon EC2 instance. The Amazon CloudWatch metrics report that both the Amazon EC2 instance and the Amazon EBS volume are under-utilized. The Amazon CloudWatch metrics also show that the Amazon EBS volume has occasional I/O bursts. The entire infrastructure is managed by AWS CloudFormation.

As a Solutions Architect, what do you propose to reduce the costs?

Keep the Amazon EBS volume to io1 and reduce the IOPS

Change the Amazon EC2 instance type to something much smaller

Correct answer

Convert the Amazon EC2 instance EBS volume to gp2

Don't use a AWS CloudFormation template to create the database as the AWS CloudFormation service incurs greater service charges

Overall explanation

Correct option:

Amazon EBS provides the various volume types, that differ in performance characteristics and price so that you can tailor your storage performance and cost to the needs of your applications. The volumes types fall into two categories:

SSD-backed volumes optimized for transactional workloads involving frequent read/write operations with small I/O size, where the dominant performance attribute is IOPS.

HDD-backed volumes optimized for large streaming workloads where throughput (measured in MiB/s) is a better performance measure than IOPS

Provisioned IOPS SSD (io1) volumes are designed to meet the needs of I/O-intensive workloads, particularly database workloads, that are sensitive to storage performance and consistency. Unlike gp2, which uses a bucket and credit model to calculate performance, an io1 volume allows you to specify a consistent IOPS rate when you create the volume, and Amazon EBS delivers the provisioned performance 99.9 percent of the time.

Convert the Amazon EC2 instance EBS volume to gp2

General Purpose SSD (gp2) volumes offer cost-effective storage that is ideal for a broad range of workloads. These volumes deliver single-digit millisecond latencies and the ability to burst to 3,000 IOPS for an extended duration. Between a minimum of 100 IOPS (at 33.33 GiB and below) and a maximum of 16,000 IOPS (at 5,334 GiB and above), baseline performance scales linearly at 3 IOPS per GiB of volume size. AWS designs gp2 volumes to deliver a provisioned performance of 99% uptime. A gp2 volume can range in size from 1 GiB to 16 TiB.

Therefore, gp2 is the right choice as it is more cost-effective than io1, and it also allows a burst in performance when needed.

Incorrect options:

Keep the Amazon EBS volume to io1 and reduce the IOPS - Keeping the Amazon EBS volume to io1 and reducing the IOPS may interfere with the burst of performance we need, so this option is ruled out.

Change the Amazon EC2 instance type to something much smaller - Changing the Amazon EC2 instance type to something much smaller won't affect 90% of the costs that are incurred, therefore this option is also incorrect.

Don't use a AWS CloudFormation template to create the database as the AWS CloudFormation service incurs greater service charges - This statement is incorrect as AWS CloudFormation is a free service to use. The resources that are invoked by CloudFormation are charged as per their utilization rates, but using AWS CloudFormation will not cost anything.

References:

https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-volume-types.html#EBSVolumeTypes_gp2

https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-volume-types.html#EBSVolumeTypes_piops

Domain

Design Cost-Optimized Architectures

Question 4Skipped

You are working as a Solutions Architect for a photo processing company that has a proprietary algorithm to compress an image without any loss in quality. Because of the efficiency of the algorithm, your clients are willing to wait for a response that carries their compressed images back. You also want to process these jobs asynchronously and scale quickly, to cater to the high demand. Additionally, you also want the job to be retried in case of failures.

Which combination of choices do you recommend to minimize cost and comply with the requirements? (Select two)

Amazon EC2 On-Demand Instances

Amazon EC2 Reserved Instances (RIs)

Correct selection

Amazon Simple Queue Service (Amazon SQS)

Amazon Simple Notification Service (Amazon SNS)

Correct selection

Amazon EC2 Spot Instances

Overall explanation

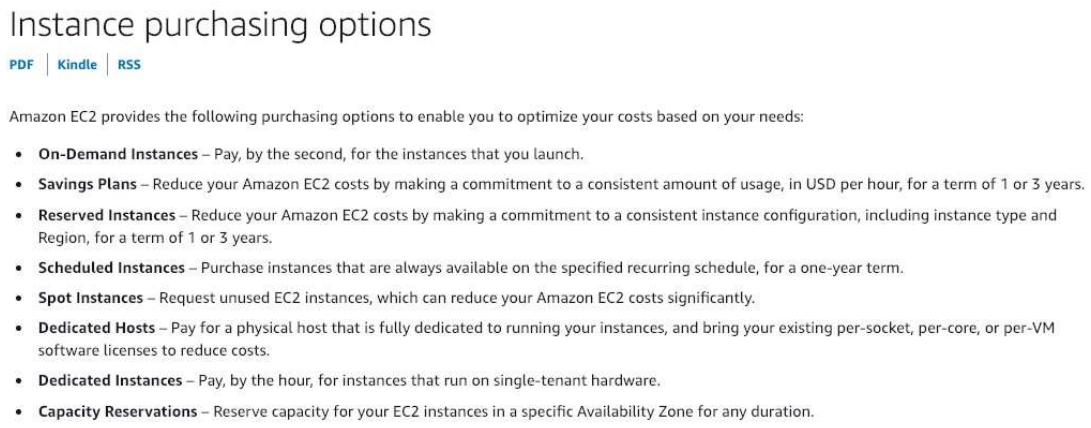
Correct options:

Amazon EC2 Spot Instances

A Spot Instance is an unused Amazon EC2 instance that is available for less than the On-Demand price. Because Spot Instances enable you to request unused Amazon EC2 instances at steep discounts, you can lower your Amazon EC2 costs significantly. The hourly price for a Spot Instance is called a Spot price. The Spot price of each instance type in each Availability Zone (AZ) is set by Amazon EC2 and adjusted gradually based on the long-term supply of and demand for Spot Instances. Your Spot Instance runs whenever capacity is available and the maximum price per hour for your request exceeds the Spot price.

To process these jobs, due to the unpredictable nature of their volume, and the desire to save on costs, spot Instances are recommended as compared to on-demand instances. As spot instances are cheaper than reserved instances and do not require long term commitment, spot instances are a better fit for the given use-case.

Amazon EC2 Instance purchasing options:



Instance purchasing options

PDF | Kindle | RSS

Amazon EC2 provides the following purchasing options to enable you to optimize your costs based on your needs:

- **On-Demand Instances** – Pay, by the second, for the instances that you launch.
- **Savings Plans** – Reduce your Amazon EC2 costs by making a commitment to a consistent amount of usage, in USD per hour, for a term of 1 or 3 years.
- **Reserved Instances** – Reduce your Amazon EC2 costs by making a commitment to a consistent instance configuration, including instance type and Region, for a term of 1 or 3 years.
- **Scheduled Instances** – Purchase instances that are always available on the specified recurring schedule, for a one-year term.
- **Spot Instances** – Request unused EC2 instances, which can reduce your Amazon EC2 costs significantly.
- **Dedicated Hosts** – Pay for a physical host that is fully dedicated to running your instances, and bring your existing per-socket, per-core, or per-VM software licenses to reduce costs.
- **Dedicated Instances** – Pay, by the hour, for instances that run on single-tenant hardware.
- **Capacity Reservations** – Reserve capacity for your EC2 instances in a specific Availability Zone for any duration.

via - https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/instance-purchasing-options.html

Amazon Simple Queue Service (Amazon SQS)

Amazon Simple Queue Service (Amazon SQS) is a fully managed message queuing service that enables you to decouple and scale microservices, distributed systems, and serverless applications. SQS offers two types of message queues. Standard queues offer maximum throughput, best-effort ordering, and at-least-once delivery. Amazon SQS FIFO (First-In-First-out) queues are designed to guarantee that messages are processed exactly once, in the exact order that they are sent.

Amazon SQS will allow you to buffer the image compression requests and process them asynchronously. It also has a direct built-in mechanism for retries and scales seamlessly.

Incorrect options:

Amazon Simple Notification Service (Amazon SNS) - Amazon Simple Notification Service (Amazon SNS) is a highly available, durable, secure, fully managed pub/sub messaging service that enables you to decouple microservices, distributed systems, and serverless applications. Amazon SNS provides topics for high-throughput, push-based, many-to-many messaging. SNS is not the right fit for this use-case, since its not a queuing mechanism.

Amazon EC2 Reserved Instances (RIs) - Reserved instances (RIs) reduce your Amazon EC2 costs by making a commitment to a consistent instance configuration, including instance type and Region, for a term of 1 or 3 years. For the given use case, this kind of annual commitment might not be a desirable option.

Amazon EC2 On-Demand Instances - With On-Demand Instances, you pay for compute capacity by the second with no long-term commitments. You have full control over its lifecycle—you decide when to launch, stop, hibernate, start, reboot, or terminate it. There is no long-term commitment required when you purchase On-Demand Instances. You pay only for the seconds that your On-Demand Instances are running. AWS recommends that you use On-Demand Instances for applications with short-term, irregular workloads that cannot be interrupted.

References:

https://aws.amazon.com/sqs/

https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-spot-instances.html

Domain

Design Cost-Optimized Architectures

Question 5Skipped

A developer in your company has set up a classic 2 tier architecture consisting of an Application Load Balancer and an Auto Scaling group (ASG) managing a fleet of Amazon EC2 instances. The Application Load Balancer is deployed in a subnet of size 10.0.1.0/24 and the Auto Scaling group is deployed in a subnet of size 10.0.4.0/22.

As a solutions architect, you would like to adhere to the security pillar of the well-architected framework. How do you configure the security group of the Amazon EC2 instances to only allow traffic coming from the Application Load Balancer?

Add a rule to authorize the security group of the Auto Scaling group

Add a rule to authorize the CIDR 10.0.1.0/24

Correct answer

Add a rule to authorize the security group of the Application Load Balancer

Add a rule to authorize the CIDR 10.0.4.0/22

Overall explanation

Correct option:

An Auto Scaling group (ASG) contains a collection of Amazon EC2 instances that are treated as a logical grouping for automatic scaling and management. An Auto Scaling group also enables you to use Amazon EC2 Auto Scaling features such as health check replacements and scaling policies. Both maintaining the number of instances in an Auto Scaling group and automatic scaling are the core functionality of the Amazon EC2 Auto Scaling service.

Add a rule to authorize the security group of the Application Load Balancer

A security group acts as a virtual firewall that controls the traffic for one or more instances. When you launch an instance, you can specify one or more security groups; otherwise, we use the default security group. You can add rules to each security group that allow traffic to or from its associated instances. You can modify the rules for a security group at any time; the new rules are automatically applied to all instances that are associated with the security group. When deciding to allow traffic to reach an instance, all the rules from all the security groups that are associated with the instance are evaluated.

The following are the characteristics of security group rules: 1. By default, security groups allow all outbound traffic. 2. Security group rules are always permissive; you can't create rules that deny access. 3. Security groups are stateful

Application Load Balancer (ALB) operates at the request level (layer 7), routing traffic to targets – Amazon EC2 instances, containers, IP addresses and Lambda functions based on the content of the request. Ideal for advanced load balancing of HTTP and HTTPS traffic, Application Load

Balancer provides advanced request routing targeted at delivery of modern application architectures, including microservices and container-based applications.

Incorrect option:

Add a rule to authorize the CIDR 10.0.4.0/22

Add a rule to authorize the security group of the Auto Scaling group

Add a rule to authorize the CIDR 10.0.1.0/24

Adding the entire CIDR of the Application Load Balancer would work, but wouldn't guarantee that only the Auto Scaling group can access the Amazon EC2 instances that are part of the Auto Scaling group. Here, the right solution is to add a rule on the Auto Scaling group security group to allow incoming traffic only from the security group configured for the Application Load Balancer.

Reference:

https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ec2-security-groups.html

Domain

Design Secure Architectures

Question 6Skipped

A media company uses Amazon ElastiCache Redis to enhance the performance of its Amazon RDS database layer. The company wants a robust disaster recovery strategy for its caching layer that guarantees minimal downtime as well as minimal data loss while ensuring good application performance.

Which of the following solutions will you recommend to address the given use-case?

Schedule daily automatic backups at a time when you expect low resource utilization for your cluster

Correct answer

Opt for Multi-AZ configuration with automatic failover functionality to help mitigate failure

Schedule manual backups using Redis append-only file (AOF)

Add read-replicas across multiple availability zones (AZs) to reduce the risk of potential data loss because of failure

Overall explanation

Correct option:

Opt for Multi-AZ configuration with automatic failover functionality to help mitigate failure

Multi-AZ is the best option when data retention, minimal downtime, and application performance are a priority.

Data-loss potential - Low. Multi-AZ provides fault tolerance for every scenario, including hardware-related issues.

Performance impact - Low. Of the available options, Multi-AZ provides the fastest time to recovery, because there is no manual procedure to follow after the process is implemented.

Cost - Low to high. Multi-AZ is the lowest-cost option. Use Multi-AZ when you can't risk losing data because of hardware failure or you can't afford the downtime required by other options in your response to an outage.

Incorrect options:

Schedule daily automatic backups at a time when you expect low resource utilization for your cluster - Data loss potential is high, almost up to a day's worth of data. Hence, this is not the right option.

Schedule manual backups using Redis append-only file (AOF) - Manual backups using AOF are retained indefinitely and are useful for testing and archiving. You can schedule manual backups to occur up to 20 times per node within any 24-hour period. Although AOF provides a measure of fault tolerance, it can't protect your data from a hardware-related cache node failure, so there is a risk of data loss.

Add read-replicas across multiple availability zones (AZs) to reduce the risk of potential data loss because of failure - To scale read capacity, Amazon ElastiCache allows you to add up to five read replicas across multiple availability zones. Read replicas are used to ease out read traffic from the primary database and cannot be used as a complete fault-tolerant solution in itself.

Reference:

https://docs.aws.amazon.com/AmazonElastiCache/latest/red-ug/FaultTolerance.html

Domain

Design Resilient Architectures

Question 7Skipped

A company has recently created a new department to handle their services workload. An IT team has been asked to create a custom VPC to isolate the resources created in this new department. They have set up the public subnet and internet gateway (IGW). However, they are not able to ping the Amazon EC2 instances with elastic IP address (EIP) launched in the newly created VPC.

As a Solutions Architect, the team has requested your help. How will you troubleshoot this scenario? (Select two)

Disable Source / Destination check on the Amazon EC2 instance

Contact AWS support to map your VPC with subnet

Correct selection

Check if the route table is configured with internet gateway

Create a secondary internet gateway to attach with public subnet and move the current internet gateway to private and write route tables

Correct selection

Check if the security groups allow ping from the source

Overall explanation

Correct options:

Check if the route table is configured with internet gateway

An internet gateway (IGW) is a horizontally scaled, redundant, and highly available VPC component that allows communication between instances in your VPC and the internet. An internet gateway serves two purposes: to provide a target in your VPC route tables for internet-routable traffic, and to perform network address translation (NAT) for instances that have been assigned public IPv4 addresses. An internet gateway supports IPv4 and IPv6 traffic.

To enable access to or from the internet for instances in a subnet in a VPC, you must do the following: 1. Attach an internet gateway to your VPC. 2. Add a route to your subnet's route table that directs internet-bound traffic to the internet gateway. 3. Ensure that instances in your subnet have a globally unique IP address 4. Ensure that your network access control lists and security group rules allow the relevant traffic to flow to and from your instance.

A route table contains a set of rules, called routes, that are used to determine where network traffic from your subnet or gateway is directed. After creating an IGW, make sure the route tables are updated. Additionally, ensure the security group allows the ICMP protocol for ping requests.

Check if the security groups allow ping from the source

A security group acts as a virtual firewall that controls the traffic for one or more instances. When you launch an instance, you can specify one or more security groups; otherwise, AWS uses the default security group. You can add rules to each security group that allow traffic to or from its associated instances. You can modify the rules for a security group at any time; the new rules are automatically applied to all instances that are associated with the security group. To decide whether to allow traffic to reach an instance, all the rules from all the security groups that are associated with the instance are evaluated.

The following are the characteristics of security group rules: 1. By default, security groups allow all outbound traffic. 2. Security group rules are always permissive; you can't create rules that deny access. 3. Security groups are stateful

Incorrect options:

Disable Source / Destination check on the Amazon EC2 instance - The Source/Destination Check attribute controls whether source/destination checking is enabled on the instance. Disabling this attribute enables an instance to handle network traffic that isn't specifically destined for the instance. For example, instances running services such as network address translation, routing, or a firewall should set this value to disabled. The default value is enabled. Source/Destination Check is not relevant to the question and it has been added as a distractor.

Create a secondary internet gateway to attach with public subnet and move the current internet gateway to private and write route tables - There is no such thing as a secondary IGW. This option is added as a distractor.

Contact AWS support to map your VPC with subnet - You cannot contact AWS support to map your VPC with the subnet.

References:

https://docs.aws.amazon.com/vpc/latest/userguide/VPC_Internet_Gateway.html

https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ec2-security-groups.html

https://docs.aws.amazon.com/vpc/latest/userguide/VPC_Route_Tables.html

https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-eni.html#change_source_dest_check

Domain

Design Secure Architectures

Question 8Skipped

For security purposes, a development team has decided to deploy the Amazon EC2 instances in a private subnet. The team plans to use VPC endpoints so that the instances can access some AWS services securely. The members of the team would like to know about the two AWS services that support Gateway Endpoints.

As a solutions architect, which of the following services would you suggest for this requirement? (Select two)

Correct selection

Amazon S3

Amazon Simple Queue Service (Amazon SQS)

Amazon Simple Notification Service (Amazon SNS)

Correct selection

Amazon DynamoDB

Amazon Kinesis

Overall explanation

Correct options:

Amazon S3

Amazon DynamoDB

A VPC endpoint enables you to privately connect your VPC to supported AWS services and VPC endpoint services powered by AWS PrivateLink without requiring an internet gateway, NAT device, VPN connection, or AWS Direct Connect connection. Instances in your VPC do not require public IP addresses to communicate with resources in the service. Traffic between your VPC and the other service does not leave the Amazon network.

Endpoints are virtual devices. They are horizontally scaled, redundant, and highly available VPC components. They allow communication between instances in your VPC and services without imposing availability risks or bandwidth constraints on your network traffic.

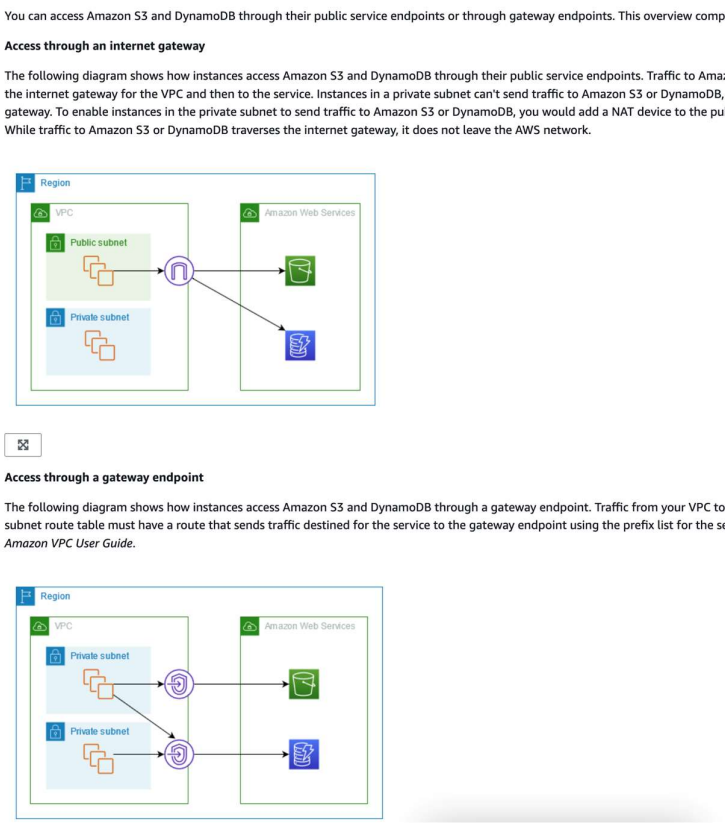
There are two types of VPC endpoints: Interface Endpoints and Gateway Endpoints. An Interface Endpoint is an Elastic Network Interface with a private IP address from the IP address range of your subnet that serves as an entry point for traffic destined to a supported service.

A Gateway Endpoint is a gateway that you specify as a target for a route in your route table for traffic destined to a supported AWS service. The following AWS services are supported: Amazon S3 and Amazon DynamoDB.

You can use two types of VPC endpoints to access Amazon S3: gateway endpoints and interface endpoints. A gateway endpoint is a gateway that you specify in your route table to access Amazon S3 from your VPC over the AWS network. Interface endpoints extend the functionality of gateway endpoints by using private IP addresses to route requests to Amazon S3 from within your VPC, on premises, or from a VPC in another AWS Region using VPC peering or AWS Transit Gateway.

You must remember that these two services use a VPC gateway endpoint. The rest of the AWS services use VPC interface endpoints.

Gateway VPC endpoints:



via - https://docs.aws.amazon.com/vpc/latest/userguide/vpce-gateway.html

Incorrect options:

Amazon Simple Queue Service (Amazon SQS)

Amazon Simple Notification Service (Amazon SNS)

Amazon Kinesis

As mentioned in the description above, these three options use interface endpoints, so these are incorrect.

Reference:

https://docs.aws.amazon.com/vpc/latest/userguide/vpc-endpoints.html

Domain

Design Secure Architectures

Question 9Skipped

A systems administrator is creating IAM policies and attaching them to IAM identities. After creating the necessary identity-based policies, the administrator is now creating resource-based policies.

Which is the only resource-based policy that the IAM service supports?

Permissions boundary

AWS Organizations Service Control Policies (SCP)

Correct answer

Trust policy

Access control list (ACL)

Overall explanation

Correct option:

You manage access in AWS by creating policies and attaching them to IAM identities (users, groups of users, or roles) or AWS resources. A policy is an object in AWS that, when associated with an identity or resource, defines their permissions. Resource-based policies are JSON policy documents that you attach to a resource such as an Amazon S3 bucket. These policies grant the specified principal permission to perform specific actions on that resource and define under what conditions this applies.

Trust policy

Trust policies define which principal entities (accounts, users, roles, and federated users) can assume the role. An IAM role is both an identity and a resource that supports resource-based policies. For this reason, you must attach both a trust policy and an identity-based policy to an IAM role. The IAM service supports only one type of resource-based policy called a role trust policy, which is attached to an IAM role.

Incorrect options:

AWS Organizations Service Control Policies (SCP) - If you enable all features of AWS organization, then you can apply service control policies (SCPs) to any or all of your accounts. SCPs are JSON policies that specify the maximum permissions for an organization or organizational unit (OU). The SCP limits permissions for entities in member accounts, including each AWS account root user. An explicit deny in any of these policies overrides the allow.

Access control list (ACL) - Access control lists (ACLs) are service policies that allow you to control which principals in another account can access a resource. ACLs cannot be used to control access for a principal within the same account. Amazon S3, AWS WAF, and Amazon VPC are examples of services that support ACLs.

Permissions boundary - AWS supports permissions boundaries for IAM entities (users or roles). A permissions boundary is an advanced feature for using a managed policy to set the maximum permissions that an identity-based policy can grant to an IAM entity. An entity's permissions boundary allows it to perform only the actions that are allowed by both its identity-based policies and its permissions boundaries.

References:

https://docs.aws.amazon.com/IAM/latest/UserGuide/access_policies.html#policies_resource-based

https://docs.aws.amazon.com/IAM/latest/UserGuide/access_policies_boundaries.html

Domain

Design Secure Architectures

Question 10Skipped

A music-sharing company uses a Network Load Balancer to direct traffic to 5 Amazon EC2 instances managed by an Auto Scaling group. When a very popular song is released, the Auto Scaling Group scales to 100 instances and the company incurs high network and compute fees.

The company wants a solution to reduce the costs without changing any of the application code. What do you recommend?

Move the songs to Amazon S3

Leverage AWS Storage Gateway

Move the songs to Amazon S3 Glacier

Correct answer

Use an Amazon CloudFront distribution

Overall explanation

Correct option:

Use an Amazon CloudFront distribution

Amazon CloudFront is a fast content delivery network (CDN) service that securely delivers data, videos, applications, and APIs to customers globally with low latency, high transfer speeds, all within a developer-friendly environment.

Amazon CloudFront points of presence (POPs) (edge locations) make sure that popular content can be served quickly to your viewers. CloudFront also has regional edge caches that bring more of your content closer to your viewers, even when the content is not popular enough to stay at a POP, to help improve performance for that content.

Regional edge caches help with all types of content, particularly content that tends to become less popular over time. Examples include user-generated content, such as video, photos, or artwork; e-commerce assets such as product photos and videos; and news and event-related content that might suddenly find new popularity.

Amazon CloudFront is the right answer because we can put it in front of our Auto Scaling group and leverage a Global Caching feature that will help us distribute the content reliably with dramatically reduced costs (the ASG won't need to scale as much).

Incorrect options:

Leverage AWS Storage Gateway - AWS Storage Gateway is a hybrid cloud storage service that gives you on-premises access to virtually unlimited cloud storage. The service provides three different types of gateways – Tape Gateway, File Gateway, and Volume Gateway – that seamlessly connect on-premises applications to cloud storage, caching data locally for low-latency access. AWS Storage Gateway cannot be used for distributing files to end-users, so this option is ruled out.

Move the songs to Amazon S3 - Amazon Simple Storage Service (Amazon S3) is an object storage service that offers industry-leading scalability, data availability, security, and performance. Using Amazon S3 would imply changing the application code, so this option is ruled out.

Move the songs to Amazon S3 Glacier - Amazon S3 Glacier and S3 Glacier Deep Archive are secure, durable, and extremely low-cost Amazon S3 cloud storage classes for data archiving and long-term backup. They are designed to deliver 99.999999999% durability and provide comprehensive security and compliance capabilities that can help meet even the most stringent regulatory requirements. Amazon Glacier is not applicable as the files are frequently requested (Glacier has retrieval times ranging from a few minutes to hours), so this option is also ruled out.

References:

https://aws.amazon.com/cloudfront/

https://aws.amazon.com/storagegateway/

Domain

Design High-Performing Architectures

Question 11Skipped

A company has developed a popular photo-sharing website using a serverless pattern on the AWS Cloud using Amazon API Gateway and AWS Lambda. The backend uses an Amazon RDS PostgreSQL database. The website is experiencing high read traffic and the AWS Lambda functions are putting an increased read load on the Amazon RDS database.

The architecture team is planning to increase the read throughput of the database, without changing the application's core logic. As a Solutions Architect, what do you recommend?

Use Amazon RDS Multi-AZ feature

Use Amazon ElastiCache

Correct answer

Use Amazon RDS Read Replicas

Use Amazon DynamoDB

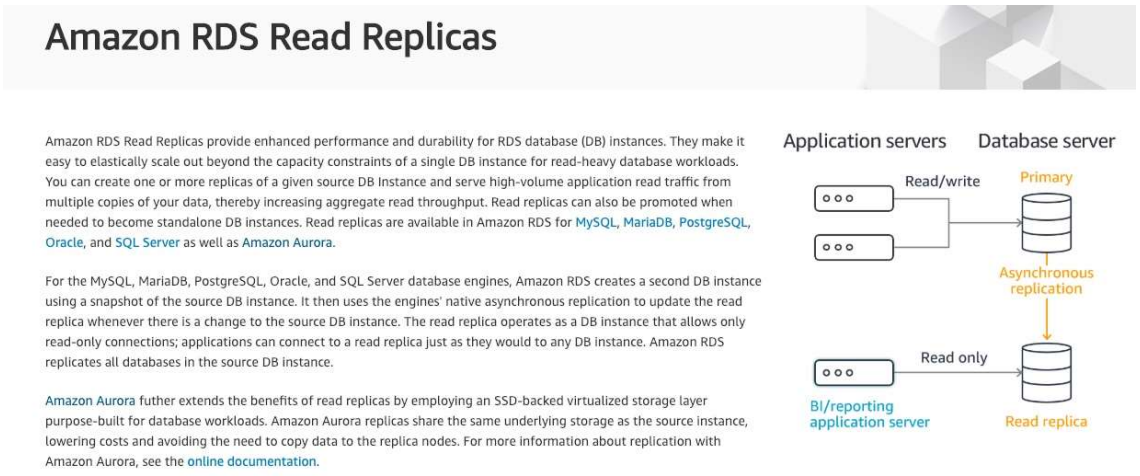
Overall explanation

Correct option:

Use Amazon RDS Read Replicas

Amazon RDS Read Replicas provide enhanced performance and durability for RDS database (DB) instances. They make it easy to elastically scale out beyond the capacity constraints of a single DB instance for read-heavy database workloads. You can create one or more replicas of a given source DB Instance and serve high-volume application read traffic from multiple copies of your data, thereby increasing aggregate read throughput. Read replicas can also be promoted when needed to become standalone DB instances.

More on Amazon RDS Read Replicas:



via - https://aws.amazon.com/rds/features/read-replicas/

Incorrect options:

Use Amazon RDS Multi-AZ feature - Amazon RDS Multi-AZ deployments provide enhanced availability and durability for RDS database (DB) instances, making them a natural fit for production database workloads. When you provision a Multi-AZ DB Instance, Amazon RDS automatically creates a primary DB Instance and synchronously replicates the data to a standby instance in a different Availability Zone (AZ). Each AZ runs on its own physically distinct, independent infrastructure, and is engineered to be highly reliable. In case of an infrastructure failure, Amazon RDS performs an automatic failover to the standby (or to a read replica in the case of Amazon Aurora), so that you can resume database operations as soon as the failover is complete.

Use Amazon ElastiCache - Amazon ElastiCache allows you to seamlessly set up, run, and scale popular open-Source compatible in-memory data stores in the cloud. Build data-intensive apps or boost the performance of your existing databases by retrieving data from high throughput and low latency in-memory data stores. Amazon ElastiCache is a popular choice for

real-time use cases like Caching, Session Stores, Gaming, Geospatial Services, Real-Time Analytics, and Queuing.

Use Amazon DynamoDB - Amazon DynamoDB is a key-value and document database that delivers single-digit millisecond performance at any scale. It's a fully managed, multi-Region, multi-master, durable database with built-in security, backup and restore, and in-memory caching for internet-scale applications. Amazon DynamoDB can handle more than 10 trillion requests per day and can support peaks of more than 20 million requests per second.

Amazon RDS Multi-AZ helps with disaster recovery in case of an AZ failure. Amazon ElastiCache would definitely help with the read load, but would require a refactor of the application's core logic. Amazon DynamoDB with DAX would also probably help with the read load, but once again it would require a refactor of the application's core logic. Here, our only option to scale reads is to use Amazon RDS Read Replicas.

References:

https://aws.amazon.com/rds/features/multi-az/

https://aws.amazon.com/rds/features/read-replicas/

Domain

Design Resilient Architectures

Question 12Skipped

As a Solutions Architect, you are tasked to design a distributed application that will run on various Amazon EC2 instances. This application needs to have the highest performance local disk to cache data. Also, data is copied through an Amazon EC2 to EC2 replication mechanism. It is acceptable if the instance loses its data when stopped or terminated.

Which storage solution do you recommend?

Amazon Elastic File System (Amazon EFS)

Correct answer

Instance Store

Amazon Elastic Block Store (EBS)

Amazon Simple Storage Service (Amazon S3)

Overall explanation

Correct option:

Instance Store

An instance store provides temporary block-level storage for your instance. This storage is located on disks that are physically attached to the host computer. Instance store is ideal for the temporary storage of information that changes frequently, such as buffers, caches, scratch data, and other temporary content, or for data that is replicated across a fleet of instances, such as a load-balanced pool of web servers.

Instance store volumes are included as part of the instance's usage cost. Some instance types use NVMe or SATA-based solid-state drives (SSD) to deliver high random I/O performance. This is a good option when you need storage with very low latency, but you don't need the data to persist when the instance terminates.

Incorrect options:

Amazon Elastic Block Store (EBS) - Amazon Elastic Block Store (EBS) is an easy to use, high-performance block storage service designed for use with Amazon Elastic Compute Cloud (EC2) for both throughput and transaction-intensive workloads at any scale.

Amazon Elastic File System (Amazon EFS) - Amazon Elastic File System (Amazon EFS) provides a simple, scalable, fully managed elastic NFS file system for use with AWS Cloud services and on-premises resources. It is built to scale on-demand to petabytes without disrupting applications, growing and shrinking automatically as you add and remove files, eliminating the need to provision and manage capacity to accommodate growth.

Amazon Simple Storage Service (Amazon S3) - Amazon Simple Storage Service (Amazon S3) is an object storage service that offers industry-leading scalability, data availability, security, and performance. Your applications can easily achieve thousands of transactions per second in request performance when uploading and retrieving storage from Amazon S3.

Instance Store will have the highest disk performance but you would lose the storage if the instance is terminated, which is acceptable in this case. Amazon EBS volumes would provide good performance as far as disk goes, but not as good as Instance Store. Amazon EBS data survives instance termination or reboots. Amazon EFS is a network drive and it would not match the performance of the Instance Store. Finally, Amazon S3 cannot be mounted as a local disk (natively).

Reference:

https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/InstanceStorage.html

Domain

Design Resilient Architectures

Question 13Skipped

A retail company uses AWS Cloud to manage its technology infrastructure. The company has deployed its consumer-focused web application on Amazon EC2-based web servers and uses Amazon RDS PostgreSQL database as the data store. The PostgreSQL database is set up in a private subnet that allows inbound traffic from selected Amazon EC2 instances. The database also uses AWS Key Management Service (AWS KMS) for encrypting data at rest.

Which of the following steps would you recommend to facilitate end-to-end security for the data-in-transit while accessing the database?

Create a new security group that blocks SSH from the selected Amazon EC2 instances into the database

Correct answer

Configure Amazon RDS to use SSL for data in transit

Use IAM authentication to access the database instead of the database user's access credentials

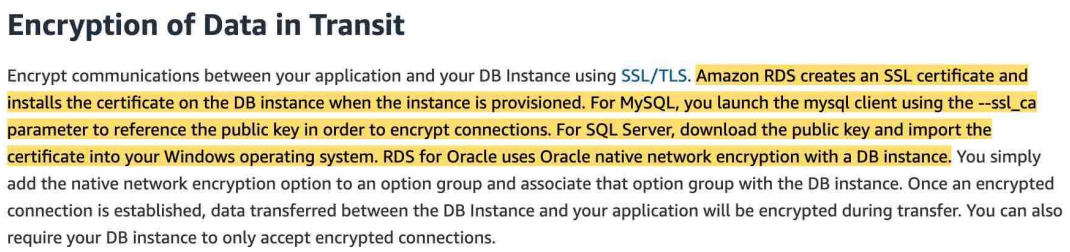
Create a new network access control list (network ACL) that blocks SSH from the entire Amazon EC2 subnet into the database

Overall explanation

Correct option:

Configure Amazon RDS to use SSL for data in transit

You can use Secure Socket Layer / Transport Layer Security (SSL/TLS) connections to encrypt data in transit. Amazon RDS creates an SSL certificate and installs the certificate on the DB instance when the instance is provisioned. For MySQL, you launch the MySQL client using the --ssl_ca parameter to reference the public key to encrypt connections. Using SSL, you can encrypt a PostgreSQL connection between your applications and your PostgreSQL DB instances. You can also force all connections to your PostgreSQL DB instance to use SSL.



Encryption of Data in Transit

Encrypt communications between your application and your DB Instance using SSL/TLS. Amazon RDS creates an SSL certificate and installs the certificate on the DB instance when the instance is provisioned. For MySQL, you launch the mysql client using the --ssl_ca parameter to reference the public key in order to encrypt connections. For SQL Server, download the public key and import the certificate into your Windows operating system. RDS for Oracle uses Oracle native network encryption with a DB instance. You simply add the native network encryption option to an option group and associate that option group with the DB instance. Once an encrypted connection is established, data transferred between the DB Instance and your application will be encrypted during transfer. You can also require your DB instance to only accept encrypted connections.

via - https://aws.amazon.com/rds/features/security/

Incorrect options:

Use IAM authentication to access the database instead of the database user's access credentials - You can authenticate to your database instance using AWS Identity and Access Management (IAM) database authentication. IAM database authentication works with MySQL and PostgreSQL. With this authentication method, you don't need to use a password when you connect to a database instance. Instead, you use an authentication token.

IAM authentication is just another way to authenticate the user's credentials while accessing the database. It would not significantly enhance the security in a way that enabling SSL does by facilitating the in-transit encryption for the database.

Create a new security group that blocks SSH from the selected Amazon EC2 instances into the database

Create a new network access control list (network ACL) that blocks SSH from the entire Amazon EC2 subnet into the database

Both these options are added as distractors. You cannot SSH into an Amazon RDS database instance.

References:

https://aws.amazon.com/rds/features/security/

https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/CHAP_MySQL.html#MySQL.Concepts.SSLSupport

https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/CHAP_PostgreSQL.html#PostgreSQL.Concepts.General.SSL

https://aws.amazon.com/blogs/database/using-iam-authentication-to-connect-with-pgadmin-amazon-aurora-postgresql-or-amazon-rds-for-postgresql/

Domain

Design Secure Architectures

Question 14Skipped

An Internet of Things (IoT) company would like to have a streaming system that performs real-time analytics on the ingested IoT data. Once the analytics is done, the company would like to send notifications back to the mobile applications of the IoT device owners.

As a solutions architect, which of the following AWS technologies would you recommend to send these notifications to the mobile applications?

Amazon Kinesis with Amazon Simple Email Service (Amazon SES)

Correct answer

Amazon Kinesis with Amazon Simple Notification Service (Amazon SNS)

Amazon Simple Queue Service (Amazon SQS) with Amazon Simple Notification Service (Amazon SNS)

Amazon Kinesis with Amazon Simple Queue Service (Amazon SQS)

Overall explanation

Correct option:

Amazon Kinesis with Amazon Simple Notification Service (Amazon SNS)

Amazon Kinesis makes it easy to collect, process, and analyze real-time, streaming data so you can get timely insights and react quickly to new information. Amazon Kinesis offers key capabilities to cost-effectively process streaming data at any scale, along with the flexibility to choose the tools that best suit the requirements of your application.

With Amazon Kinesis, you can ingest real-time data such as video, audio, application logs, website clickstreams, and IoT telemetry data for machine learning, analytics, and other applications. Amazon Kinesis enables you to process and analyze data as it arrives and respond instantly instead of having to wait until all your data is collected before the processing can begin.

Amazon Kinesis will be great for event streaming from the IoT devices, but not for sending notifications as it doesn't have such a feature.

Amazon Simple Notification Service (Amazon SNS) is a highly available, durable, secure, fully managed pub/sub messaging service that enables you to decouple microservices, distributed systems, and serverless applications. Amazon SNS provides topics for high-throughput, push-

based, many-to-many messaging. Amazon SNS is a notification service and will be perfect for this use case.

Streaming data with Amazon Kinesis and using Amazon SNS to send the response notifications is the optimal solution for the current scenario.

Incorrect options:

Amazon Simple Queue Service (Amazon SQS) with Amazon Simple Notification Service (Amazon SNS) - Amazon Simple Queue Service (SQS) is a fully managed message queuing service that enables you to decouple and scale microservices, distributed systems, and serverless applications. Amazon SQS eliminates the complexity and overhead associated with managing and operating message-oriented middleware and empowers developers to focus on differentiating work. Using SQS, you can send, store, and receive messages between software components at any volume, without losing messages or requiring other services to be available. Kinesis is better for streaming data since queues aren't meant for real-time streaming of data.

Amazon Kinesis with Amazon Simple Email Service (Amazon SES) - Amazon Simple Email Service (Amazon SES) is a cloud-based email sending service designed to help digital marketers and application developers send marketing, notification, and transactional emails. It is a reliable, cost-effective service for businesses of all sizes that use email to keep in contact with their customers. It is an email service and not a notification service as is the requirement in the current use case.

Amazon Kinesis with Amazon Simple Queue Service (Amazon SQS) - As explained above, Amazon Kinesis works well for streaming real-time data. Amazon SQS is a queuing service that helps decouple system architecture by offering flexibility and ease of maintenance. It cannot send notifications. Amazon SQS is paired with SNS to provide this functionality.

References:

https://aws.amazon.com/sns/

https://aws.amazon.com/kinesis/

https://aws.amazon.com/ses/

https://aws.amazon.com/sqs/

Domain

Design Resilient Architectures

Question 15Skipped

A company wants to adopt a hybrid cloud infrastructure where it uses some AWS services such as Amazon S3 alongside its on-premises data center. The company wants a dedicated private connection between the on-premise data center and AWS. In case of failures though, the company needs to guarantee uptime and is willing to use the public internet for an encrypted connection.

What do you recommend? (Select two)

Use AWS Direct Connect connection as a backup connection

Use AWS Site-to-Site VPN as a primary connection

Correct selection

Use AWS Site-to-Site VPN as a backup connection

Use Egress Only Internet Gateway as a backup connection

Correct selection

Use AWS Direct Connect connection as a primary connection

Overall explanation

Correct options:

Use AWS Direct Connect connection as a primary connection

AWS Direct Connect lets you establish a dedicated network connection between your network and one of the AWS Direct Connect locations. Using industry-standard 802.1q VLANs, this dedicated connection can be partitioned into multiple virtual interfaces. AWS Direct Connect does not involve the Internet; instead, it uses dedicated, private network connections between your intranet and Amazon VPC.

Use AWS Site-to-Site VPN as a backup connection

AWS Site-to-Site VPN enables you to securely connect your on-premises network or branch office site to your Amazon Virtual Private Cloud (Amazon VPC). You can securely extend your data center or branch office network to the cloud with an AWS Site-to-Site VPN connection. A VPC VPN Connection utilizes IPSec to establish encrypted network connectivity between your intranet and Amazon VPC over the Internet. VPN Connections can be configured in minutes and are a good solution if you have an immediate need, have low to modest bandwidth requirements, and can tolerate the inherent variability in Internet-based connectivity.

AWS Direct Connect as a primary connection guarantees great performance and security (as the connection is private). Using Direct Connect as a backup solution would work but probably carries a risk it would fail as well. As we don't mind going over the public internet (which is reliable, but less secure as connections are going over the public route), we should use a Site to Site VPN which offers an encrypted connection to handle failover scenarios.

Incorrect options:

Use Egress Only Internet Gateway as a backup connection - An Egress-Only Internet Gateway is a horizontally scaled, redundant, and highly available VPC component that allows outbound communication over IPv6 from instances in your VPC to the Internet, and prevents the Internet from initiating an IPv6 connection with your instances. Egress-Only Internet Gateway cannot be used to connect on-premises data centers to AWS Cloud.

Use AWS Site-to-Site VPN as a primary connection - AWS Site-to-Site VPN as a primary connection is not advisable since the use of internet-based connection is only for failover scenarios, as stated in the problem.

Use AWS Direct Connect connection as a backup connection - AWS Direct Connect connection is a highly secure, physical connection. It is also a costly solution and hence does not make much sense to set up the connection and keep it only as a backup.

References:

https://aws.amazon.com/directconnect/

https://aws.amazon.com/vpn/

Domain

Design Secure Architectures

Question 16Skipped

A company has grown from a small startup to an enterprise employing over 1000 people. As the team size has grown, the company has recently observed some strange behavior, with Amazon S3 buckets settings being changed regularly.

How can you figure out what's happening without restricting the rights of the users?

Use Amazon S3 access logs to analyze user access using Athena

Implement an IAM policy to forbid users to change Amazon S3 bucket settings

Correct answer

Use AWS CloudTrail to analyze API calls

Implement a bucket policy requiring AWS Multi-Factor Authentication (AWS MFA) for all operations

Overall explanation

Correct option:

Use AWS CloudTrail to analyze API calls

AWS CloudTrail is a service that enables governance, compliance, operational auditing, and risk auditing of your AWS account. With AWS CloudTrail, you can log, continuously monitor, and retain account activity related to actions across your AWS infrastructure. AWS CloudTrail provides event history of your AWS account activity, including actions taken through the AWS Management Console, AWS SDKs, command-line tools, and other AWS services.

In general, to analyze any API calls made within an AWS account, AWS CloudTrail is used. You can record the actions that are taken by users, roles, or AWS services on Amazon S3 resources and maintain log records for auditing and compliance purposes. To do this, you can use server access logging, AWS CloudTrail logging, or a combination of both. AWS recommends that you use AWS CloudTrail for logging bucket and object-level actions for your Amazon S3 resources.

Incorrect options:

Implement an IAM policy to forbid users to change Amazon S3 bucket settings - You manage access in AWS by creating policies and attaching them to IAM identities (users, groups of users, or roles) or AWS resources. A policy is an object in AWS that, when associated with an identity or resource, defines their permissions. AWS evaluates these policies when an IAM principal (user or role) makes a request. Permissions in the policies determine whether the request is allowed or denied. Most policies are stored in AWS as JSON documents. AWS supports six types

of policies: identity-based policies, resource-based policies, permissions boundaries, AWS Organizations service control policy (SCP), access control list (ACL), and session policies.

Implementing an IAM policy to forbid users would be disruptive and wouldn't go unnoticed.

Use Amazon S3 access logs to analyze user access using Athena - Amazon S3 server access logging provides detailed records for the requests that are made to a bucket. Server access logs are useful for many applications. For example, access log information can be useful in security and access audits. It can also help you learn about your customer base and understand your Amazon S3 bill. AWS recommends that you use AWS CloudTrail for logging bucket and object-level actions for your Amazon S3 resources, as it provides more options to store, analyze and act on the log information.

Implement a bucket policy requiring AWS Multi-Factor Authentication (AWS MFA) for all operations - Amazon S3 supports MFA-protected API access, a feature that can enforce multi-factor authentication (MFA) for access to your Amazon S3 resources. Multi-factor authentication provides an extra level of security that you can apply to your AWS environment. It is a security feature that requires users to prove the physical possession of an MFA device by providing a valid MFA code. Changing the bucket policy to require MFA would not go unnoticed.

References:

https://docs.aws.amazon.com/AmazonS3/latest/userguide/logging-with-S3.html

https://aws.amazon.com/cloudtrail/

https://docs.aws.amazon.com/IAM/latest/UserGuide/access_policies.html

https://docs.aws.amazon.com/AmazonS3/latest/dev/ServerLogs.html

https://docs.aws.amazon.com/AmazonS3/latest/dev/example-bucket-policies.html#example-bucket-policies-use-case-7

Domain

Design Resilient Architectures

Question 17Skipped

A company wants to grant access to an Amazon S3 bucket to users in its own AWS account as well as to users in another AWS account. Which of the following options can be used to meet this requirement?

Correct answer

Use a bucket policy to grant permission to users in its account as well as to users in another account

Use a user policy to grant permission to users in its account as well as to users in another account

Use permissions boundary to grant permission to users in its account as well as to users in another account

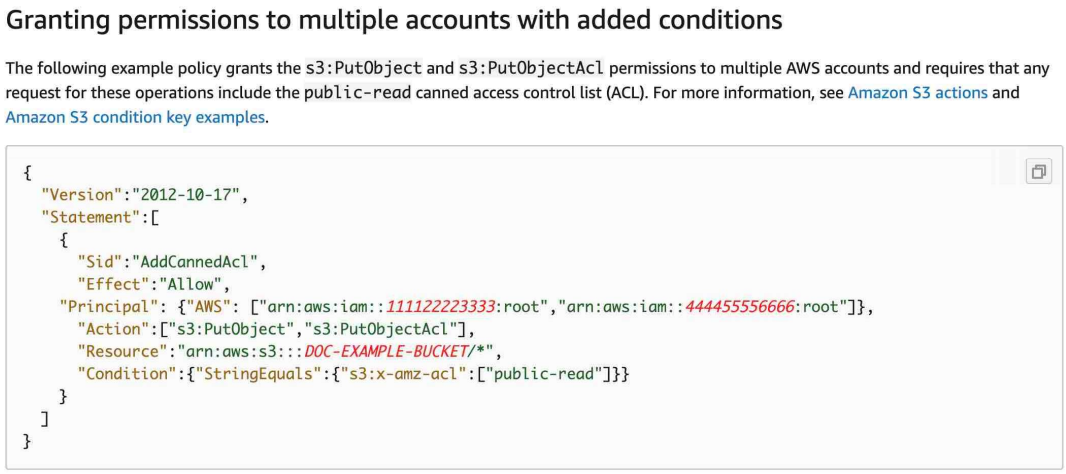
Use either a bucket policy or a user policy to grant permission to users in its account as well as to users in another account

Overall explanation

Correct option:

Use a bucket policy to grant permission to users in its account as well as to users in another account

A bucket policy is a type of resource-based policy that can be used to grant permissions to the principal that is specified in the policy. Principals can be in the same account as the resource or in other accounts. For cross-account permissions to other AWS accounts or users in another account, you must use a bucket policy.



Granting permissions to multiple accounts with added conditions

The following example policy grants the `s3:PutObject` and `s3:PutObjectAcl` permissions to multiple AWS accounts and requires that any request for these operations include the `public-read` canned access control list (ACL). For more information, see Amazon S3 actions and Amazon S3 condition key examples.

```
{
  "Version":"2012-10-17",
  "Statement":[
    {
      "Sid":"AddCannedAcl",
      "Effect":"Allow",
    "Principal": {"AWS": ["arn:aws:iam::111122223333:root","arn:aws:iam::444455556666:root"]},
      "Action":["s3:PutObject","s3:PutObjectAcl"],
      "Resource":"arn:aws:s3:::DOC-EXAMPLE-BUCKET/*",
      "Condition":{"StringEquals":{"s3:x-amz-acl":["public-read"]}}
    }
  ]
}
```

via - https://docs.aws.amazon.com/AmazonS3/latest/userguide/example-bucket-policies.html

Incorrect options:

Use either a bucket policy or a user policy to grant permission to users in its account as well as to users in another account

Use a user policy to grant permission to users in its account as well as to users in another account

If an AWS account that owns a bucket wants to grant permission to users in its own AWS account, it can use either a bucket policy or a user policy. The user policies are for managing permissions for users in their own AWS account and NOT for users in other AWS accounts. Therefore both these options are incorrect.

Use permissions boundary to grant permission to users in its account as well as to users in another account - Use a managed policy as the permissions boundary for an IAM entity (user or role). That policy defines the maximum permissions that the identity-based policies can grant to an entity, but does not grant permissions, so this option is not correct for the given use case.

Policy types

The following policy types, listed in order of frequency, are available for use in AWS. For more details, see the sections below for each policy type.

- **Identity-based policies** – Attach managed and inline policies to IAM identities (users, groups to which users belong, or roles). Identity-based policies grant permissions to an identity.
- **Resource-based policies** – Attach inline policies to resources. The most common examples of resource-based policies are Amazon S3 bucket policies and IAM role trust policies. Resource-based policies grant permissions to the principal that is specified in the policy. Principals can be in the same account as the resource or in other accounts.
- **Permissions boundaries** – Use a managed policy as the permissions boundary for an IAM entity (user or role). That policy defines the maximum permissions that the identity-based policies can grant to an entity, but does not grant permissions. Permissions boundaries do not define the maximum permissions that a resource-based policy can grant to an entity.
- **Organizations SCPs** – Use an AWS Organizations service control policy (SCP) to define the maximum permissions for account members of an organization or organizational unit (OU). SCPs limit permissions that identity-based policies or resource-based policies grant to entities (users or roles) within the account, but do not grant permissions.
- **Access control lists (ACLs)** – Use ACLs to control which principals in other accounts can access the resource to which the ACL is attached. ACLs are similar to resource-based policies, although they are the only policy type that does not use the JSON policy document structure. ACLs are cross-account permissions policies that grant permissions to the specified principal. ACLs cannot grant permissions to entities within the same account.
- **Session policies** – Pass advanced session policies when you use the AWS CLI or AWS API to assume a role or a federated user. Session policies limit the permissions that the role or user's identity-based policies grant to the session. Session policies limit permissions for a created session, but do not grant permissions. For more information, see Session Policies.

via - https://docs.aws.amazon.com/IAM/latest/UserGuide/access_policies.html

References:

https://docs.aws.amazon.com/AmazonS3/latest/userguide/example-bucket-policies.html

https://docs.aws.amazon.com/IAM/latest/UserGuide/access_policies.html

Domain

Design Secure Architectures

Question 18Skipped

A mobile gaming company is experiencing heavy read traffic to its Amazon Relational Database Service (Amazon RDS) database that retrieves player's scores and stats. The company is using an Amazon RDS database instance type that is not cost-effective for their budget. The company would like to implement a strategy to deal with the high volume of read traffic, reduce latency, and also downsize the instance size to cut costs.

Which of the following solutions do you recommend?

Correct answer

Setup Amazon ElastiCache in front of Amazon RDS

Move to Amazon Redshift

Setup Amazon RDS Read Replicas

Switch application code to AWS Lambda for better performance

Overall explanation

Correct option:

Setup Amazon ElastiCache in front of Amazon RDS

Amazon ElastiCache is an ideal front-end for data stores such as Amazon RDS, providing a high-performance middle tier for applications with extremely high request rates and/or low latency requirements. The best part of caching is that it's minimally invasive to implement and by doing so, your application performance regarding both scale and speed is dramatically improved.

Incorrect options:

Setup Amazon RDS Read Replicas - Adding read replicas would further add to the database costs and will not help in reducing latency when compared to a caching solution. So this option is ruled out.

Move to Amazon Redshift - Amazon Redshift is optimized for datasets ranging from a few hundred gigabytes to a petabyte or more. If the company is looking at cost-cutting, moving to Amazon Redshift from Amazon RDS is not an option.

Switch application code to AWS Lambda for better performance - AWS Lambda can help in running data processing workflows. But, data still needs to be read from RDS and hence we need a solution to speed up the data reads and not before/after processing.

Reference:

https://aws.amazon.com/caching/database-caching/

Domain

Design Cost-Optimized Architectures

Question 19Skipped

A ride-sharing company wants to improve the ride-tracking system that stores GPS coordinates for all rides. The engineering team at the company is looking for a NoSQL database that has single-digit millisecond latency, can scale horizontally, and is serverless, so that they can perform high-frequency lookups reliably.

As a Solutions Architect, which database do you recommend for their requirements?

Correct answer

Amazon DynamoDB

Amazon Neptune

Amazon ElastiCache

Amazon Relational Database Service (Amazon RDS)

Overall explanation

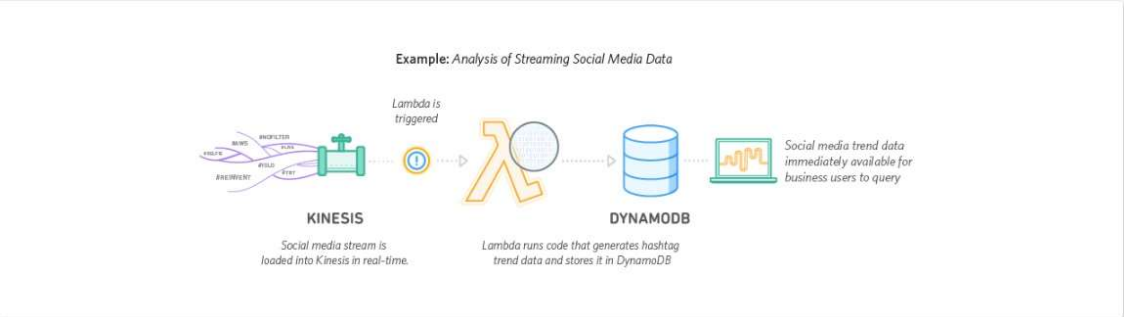
Correct option:

Amazon DynamoDB

Amazon DynamoDB is a key-value and document database that delivers single-digit millisecond performance at any scale. It's a fully managed, multi-Region, multi-master, durable NoSQL database with built-in security, backup and restore, and in-memory caching for internet-scale applications. DynamoDB can handle more than 10 trillion requests per day and can support peaks of more than 20 million requests per second. DynamoDB is serverless, has single-digit

millisecond latency and scales horizontally. This is the correct choice for the given requirements.

Sample Amazon DynamoDB solution for Real time applications:

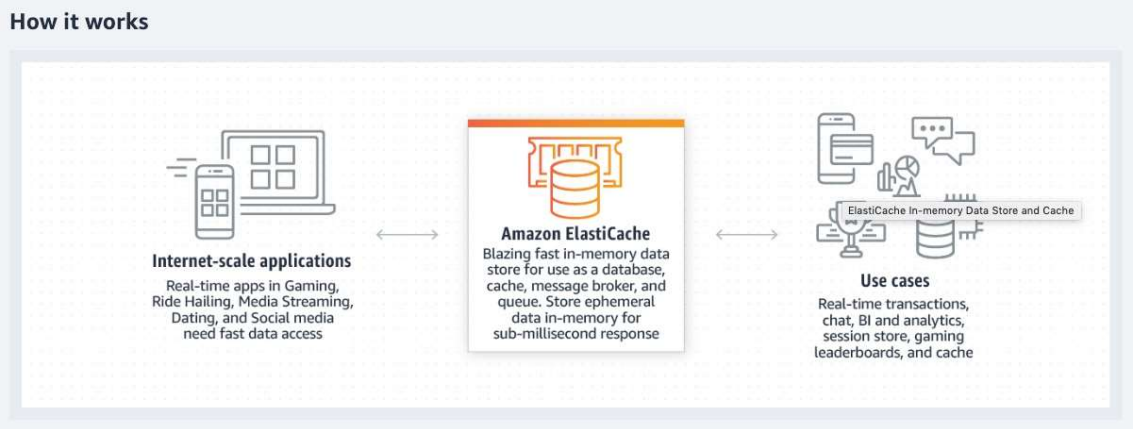


via - https://aws.amazon.com/dynamodb/

Incorrect options:

Amazon ElastiCache - Amazon ElastiCache allows you to seamlessly set up, run, and scale popular open-Source compatible in-memory data stores, compatible with Redis or Memcached. Build data-intensive apps or boost the performance of your existing databases by retrieving data from high throughput and low latency in-memory data stores. Amazon ElastiCache is a popular choice for real-time use cases like Caching, Session Stores, Gaming, Geospatial Services, Real-Time Analytics, and Queuing. The primary use-case for ElastiCache is that of a caching service and it should not be used as the main database.

How Amazon ElastiCache works:

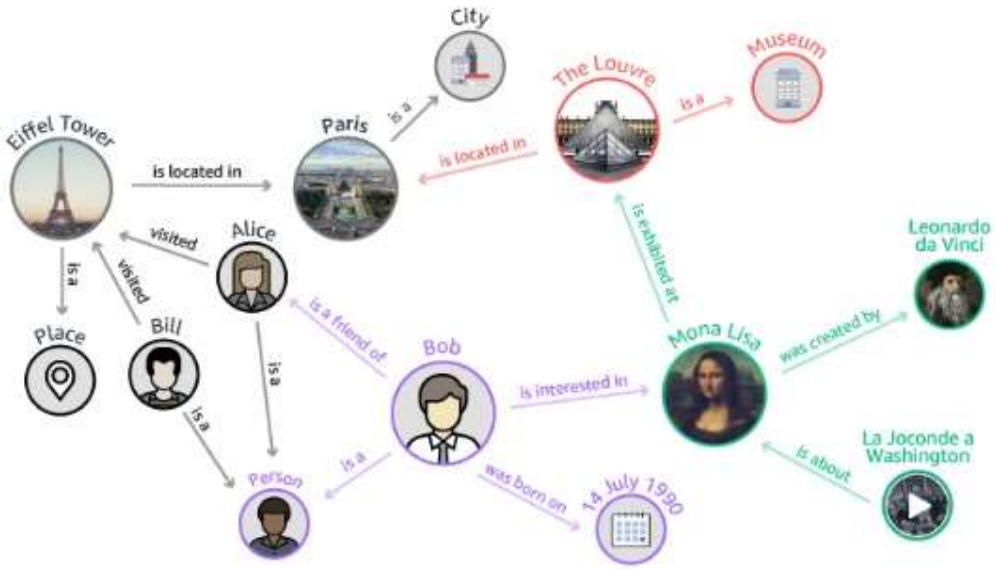


via - https://aws.amazon.com/elasticache/

Amazon Relational Database Service (Amazon RDS) - Amazon Relational Database Service (Amazon RDS) makes it easy to set up, operate, and scale a relational database in the cloud. It provides cost-efficient and resizable capacity while automating time-consuming administration tasks such as hardware provisioning, database setup, patching and backups. Relational databases are not NoSQL databases and these cannot provide the millisecond latency that the current use case needs, hence it's an incorrect choice.

Amazon Neptune - Amazon Neptune is a fast, reliable, fully-managed graph database service that makes it easy to build and run applications that work with highly connected datasets. The core of Amazon Neptune is a purpose-built, high-performance graph database engine optimized for storing billions of relationships and querying the graph with milliseconds latency. Neptune powers graph use cases such as recommendation engines, fraud detection, knowledge graphs, drug discovery, and network security.

Example Use cases of Amazon Neptune:



via - https://aws.amazon.com/neptune/

References:

https://aws.amazon.com/dynamodb/

https://aws.amazon.com/elasticache/

https://aws.amazon.com/neptune/

Domain

Design Resilient Architectures

Question 20Skipped

The development team at a social media company wants to handle some complicated queries such as "What are the number of likes on the videos that have been posted by friends of a user A?".

As a solutions architect, which of the following AWS database services would you suggest as the BEST fit to handle such use cases?

Amazon Aurora

Amazon Redshift

Correct answer

Amazon Neptune

Amazon OpenSearch Service

Overall explanation

Correct option:

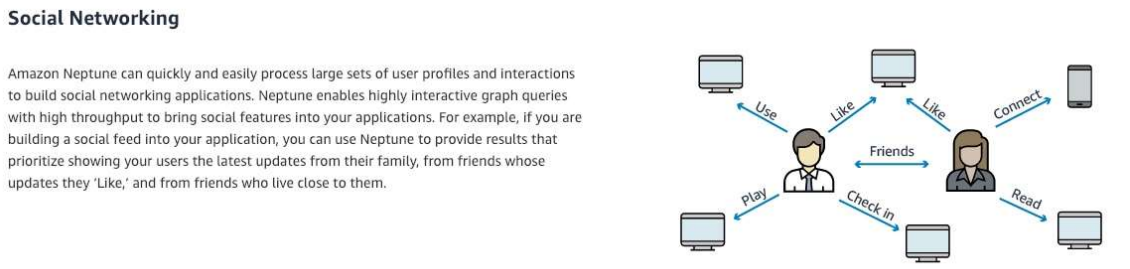
Amazon Neptune

Amazon Neptune is a fast, reliable, fully managed graph database service that makes it easy to build and run applications that work with highly connected datasets. The core of Amazon Neptune is a purpose-built, high-performance graph database engine optimized for storing billions of relationships and querying the graph with milliseconds latency. Neptune powers graph use cases such as recommendation engines, fraud detection, knowledge graphs, drug discovery, and network security.

Amazon Neptune is highly available, with read replicas, point-in-time recovery, continuous backup to Amazon S3, and replication across Availability Zones. Neptune is secure with support for HTTPS encrypted client connections and encryption at rest. Neptune is fully managed, so you no longer need to worry about database management tasks such as hardware provisioning, software patching, setup, configuration, or backups.

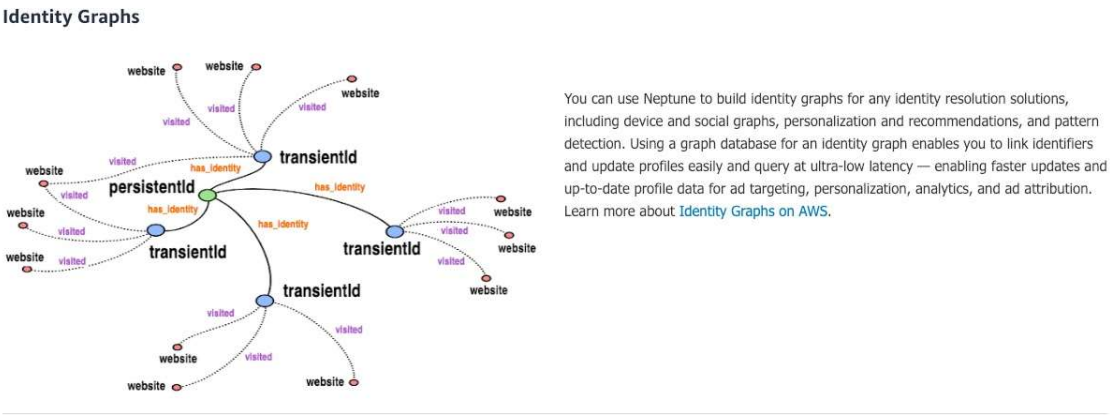
Amazon Neptune can quickly and easily process large sets of user-profiles and interactions to build social networking applications. Neptune enables highly interactive graph queries with high throughput to bring social features into your applications. For example, if you are building a social feed into your application, you can use Neptune to provide results that prioritize showing your users the latest updates from their family, from friends whose updates they 'Like,' and from friends who live close to them.

Social Networking example with Amazon Neptune:



via - https://aws.amazon.com/neptune/

Identity graphs example with Amazon Neptune:



via - https://aws.amazon.com/neptune/

Incorrect options:

Amazon OpenSearch Service - Amazon OpenSearch Service is a managed service that makes it easy for you to perform interactive log analytics, real-time application monitoring, website search, and more. OpenSearch is an open source, distributed search and analytics suite derived from Elasticsearch. Amazon OpenSearch Service offers the latest versions of OpenSearch, support for 19 versions of Elasticsearch (1.5 to 7.10 versions), as well as visualization capabilities powered by OpenSearch Dashboards and Kibana (1.5 to 7.10 versions). Amazon OpenSearch Service currently has tens of thousands of active customers with hundreds of thousands of clusters under management processing trillions of requests per month.

Amazon Redshift - Amazon Redshift is a fully-managed petabyte-scale cloud-based data warehouse product designed for large scale data set storage and analysis. The given use-case is not about data warehousing, so this is not a correct option.

Amazon Aurora - Amazon Aurora is a MySQL and PostgreSQL-compatible relational database built for the cloud, that combines the performance and availability of traditional enterprise databases with the simplicity and cost-effectiveness of open source databases. Amazon Aurora features a distributed, fault-tolerant, self-healing storage system that auto-scales up to 64 terabytes per database instance. Aurora is not an in-memory database. Here, we need a graph database due to the highly connected datasets and queries, therefore Neptune is the best answer.

Reference:

https://aws.amazon.com/neptune/

Domain

Design High-Performing Architectures

Question 21Skipped

You are working for a software as a service (SaaS) company as a solutions architect and help design solutions for the company's customers. One of the customers is a bank and has a requirement to whitelist a public IP when the bank is accessing external services across the internet.

Which architectural choice do you recommend to maintain high availability, support scaling-up to 10 instances and comply with the bank's requirements?

Use a Classic Load Balancer with an Auto Scaling Group

Use an Auto Scaling Group with Dynamic Elastic IPs attachment

Correct answer

Use a Network Load Balancer with an Auto Scaling Group

Use an Application Load Balancer with an Auto Scaling Group

Overall explanation

Correct option:

Use a Network Load Balancer with an Auto Scaling Group

Network Load Balancer is best suited for use-cases involving low latency and high throughput workloads that involve scaling to millions of requests per second. Network Load Balancer operates at the connection level (Layer 4), routing connections to targets - Amazon EC2 instances, microservices, and containers – within Amazon Virtual Private Cloud (Amazon VPC) based on IP protocol data. A Network Load Balancer functions at the fourth layer of the Open Systems Interconnection (OSI) model. It can handle millions of requests per second.

Network Load Balancers expose a fixed IP to the public web, therefore allowing your application to be predictably reached using this IP, while allowing you to scale your application behind the Network Load Balancer using an ASG.

Incorrect options:

Classic Load Balancers and Application Load Balancers use the private IP addresses associated with their Elastic network interfaces as the source IP address for requests forwarded to your web servers.

These IP addresses can be used for various purposes, such as allowing the load balancer traffic on the web servers and for request processing. It's a best practice to use security group referencing on the web servers for whitelisting load balancer traffic from Classic Load Balancers or Application Load Balancers.

However, because Network Load Balancers don't support security groups, based on the target group configurations, the IP addresses of the clients or the private IP addresses associated with the Network Load Balancers must be allowed on the web server's security group.

Use a Classic Load Balancer with an Auto Scaling Group - Classic Load Balancer 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 were built within the Amazon EC2-Classic network.

Use an Application Load Balancer with an Auto Scaling Group - Application Load Balancer operates at the request level (layer 7), routing traffic to targets – Amazon EC2 instances, containers, IP addresses and AWS Lambda functions based on the content of the request. Ideal for advanced load balancing of HTTP and HTTPS traffic, Application Load Balancer provides

advanced request routing targeted at the delivery of modern application architectures, including microservices and container-based applications.

Application and Classic Load Balancers expose a fixed DNS (=URL) rather than the IP address. So these are incorrect options for the given use-case.

Use an Auto Scaling Group with Dynamic Elastic IPs attachment - The option "Use an Auto Scaling Group (ASG) with Dynamic Elastic IPs attachment" has been added as a distractor. ASG does not have a dynamic Elastic IPs attachment feature.

References:

https://docs.aws.amazon.com/elasticloadbalancing/latest/network/introduction.html

https://aws.amazon.com/premiumsupport/knowledge-center/elb-find-load-balancer-IP/

https://docs.aws.amazon.com/elasticloadbalancing/latest/classic/elb-internet-facing-load-balancers.html

Domain

Design Resilient Architectures

Question 22Skipped

A CRM web application was written as a monolith in PHP and is facing scaling issues because of performance bottlenecks. The CTO wants to re-engineer towards microservices architecture and expose their application from the same load balancer, linked to different target groups with different URLs: checkout.mycorp.com, www.mycorp.com, yourcorp.com/profile and yourcorp.com/search. The CTO would like to expose all these URLs as HTTPS endpoints for security purposes.

As a solutions architect, which of the following would you recommend as a solution that requires MINIMAL configuration effort?

Use a wildcard Secure Sockets Layer certificate (SSL certificate)

Correct answer

Use Secure Sockets Layer certificate (SSL certificate) with SNI

Change the Elastic Load Balancing (ELB) SSL Security Policy

Use an HTTP to HTTPS redirect

Overall explanation

Correct option:

Use Secure Sockets Layer certificate (SSL certificate) with SNI

You can host multiple TLS secured applications, each with its own TLS certificate, behind a single load balancer. To use SNI, all you need to do is bind multiple certificates to the same secure listener on your load balancer. ALB will automatically choose the optimal TLS certificate for each client.

ALB's smart certificate selection goes beyond SNI. In addition to containing a list of valid domain names, certificates also describe the type of key exchange and cryptography that the server supports, as well as the signature algorithm (SHA2, SHA1, MD5) used to sign the certificate.

With SNI support AWS makes it easy to use more than one certificate with the same ALB. The most common reason you might want to use multiple certificates is to handle different domains with the same load balancer. It's always been possible to use wildcard and subject-alternate-name (SAN) certificates with ALB, but these come with limitations. Wildcard certificates only work for related subdomains that match a simple pattern and while SAN certificates can support many different domains, the same certificate authority has to authenticate each one. That means you have to reauthenticate and reprovision your certificate every time you add a new domain.

Incorrect options:

Use a wildcard Secure Sockets Layer certificate (SSL certificate) - As the use case requires different domain names, so you cannot use a wildcard SSL certificate.

Use an HTTP to HTTPS redirect - This will not provide multiple secure endpoints for different URLs such as checkout.mycorp.com or www.mycorp.com, therefore it is incorrect for the given use-case.

Change the Elastic Load Balancing (ELB) SSL Security Policy - Elastic Load Balancing (ELB) SSL Security Policy will not provide multiple secure endpoints for different URLs such as checkout.mycorp.com or www.mycorp.com, therefore it is incorrect for the given use-case.

References:

https://aws.amazon.com/blogs/aws/new-application-load-balancer-sni/

https://docs.aws.amazon.com/elasticloadbalancing/latest/classic/elb-security-policy-table.html

Domain

Design Secure Architectures

Question 23Skipped

A financial services firm has traditionally operated with an on-premise data center and would like to create a disaster recovery strategy leveraging the AWS Cloud.

As a Solutions Architect, you would like to ensure that a scaled-down version of a fully functional environment is always running in the AWS cloud, and in case of a disaster, the recovery time is kept to a minimum. Which disaster recovery strategy is that?

Correct answer

Warm Standby

Backup and Restore

Pilot Light

Multi Site

Overall explanation

Correct option:

Warm Standby

The term warm standby is used to describe a DR scenario in which a scaled-down version of a fully functional environment is always running in the cloud. A warm standby solution extends the pilot light elements and preparation. It further decreases the recovery time because some services are always running. By identifying your business-critical systems, you can fully duplicate these systems on AWS and have them always on.

Incorrect options:

Backup and Restore - In most traditional environments, data is backed up to tape and sent off-site regularly. If you use this method, it can take a long time to restore your system in the event of a disruption or disaster. Amazon S3 is an ideal destination for backup data that might be needed quickly to perform a restore. Transferring data to and from Amazon S3 is typically done through the network, and is therefore accessible from any location. Many commercial and open-source backup solutions integrate with Amazon S3.

Pilot Light - The term pilot light is often used to describe a DR scenario in which a minimal version of an environment is always running in the cloud. The idea of the pilot light is an analogy that comes from the gas heater. In a gas heater, a small flame that's always on can quickly ignite the entire furnace to heat up a house. This scenario is similar to a backup-and-restore scenario. For example, with AWS you can maintain a pilot light by configuring and running the most critical core elements of your system in AWS. When the time comes for recovery, you can rapidly provision a full-scale production environment around the critical core.

Multi Site - A multi-site solution runs in AWS as well as on your existing on-site infrastructure, in an active-active configuration. The data replication method that you employ will be determined by the recovery point that you choose.

References:

https://d1.awsstatic.com/whitepapers/aws-disaster-recovery.pdf

https://d1.awsstatic.com/asset-repository/products/CloudEndure/CloudEndure_Affordable_Enterprise-Grade_Disaster_Recovery_Using_AWS.pdf

Domain

Design Resilient Architectures

Question 24Skipped

The engineering team at a social media company has recently migrated to AWS Cloud from its on-premises data center. The team is evaluating Amazon CloudFront to be used as a CDN for its flagship application. The team has hired you as an AWS Certified Solutions Architect – Associate to advise on Amazon CloudFront capabilities on routing, security, and high availability.

Which of the following would you identify as correct regarding Amazon CloudFront? (Select three)

Amazon CloudFront can route to multiple origins based on the price class

Correct selection

Use an origin group with primary and secondary origins to configure Amazon CloudFront for high-availability and failover

Use geo restriction to configure Amazon CloudFront for high-availability and failover

Correct selection

Amazon CloudFront can route to multiple origins based on the content type

Use AWS Key Management Service (AWS KMS) encryption in Amazon CloudFront to protect sensitive data for specific content

Correct selection

Use field level encryption in Amazon CloudFront to protect sensitive data for specific content

Overall explanation

Correct options:

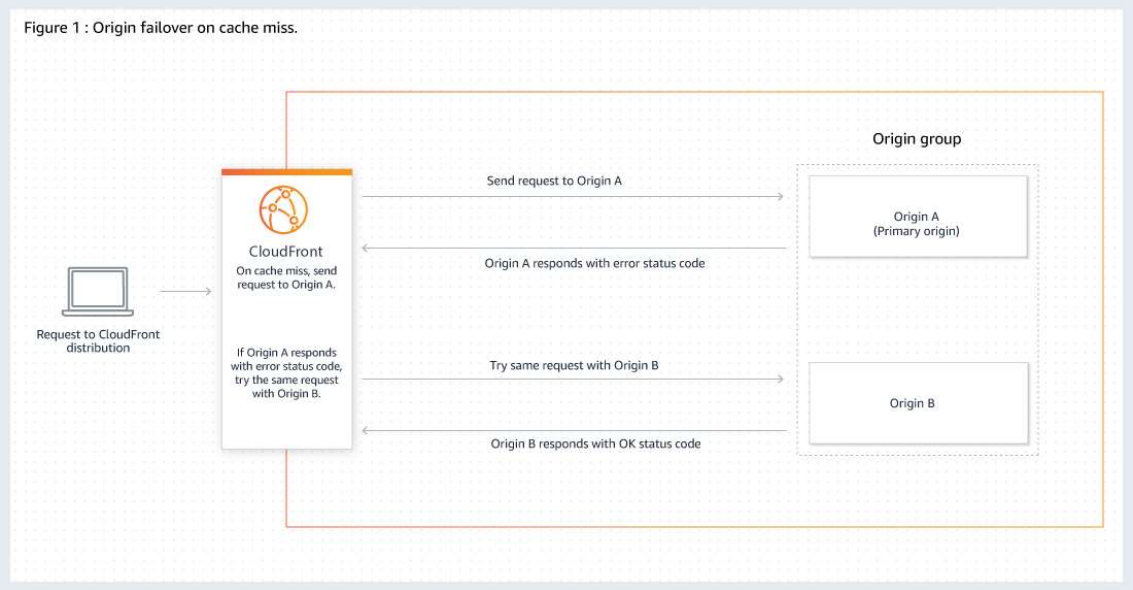
Amazon CloudFront can route to multiple origins based on the content type

You can configure a single Amazon CloudFront web distribution to serve different types of requests from multiple origins. For example, if you are building a website that serves static content from an Amazon Simple Storage Service (Amazon S3) bucket and dynamic content from a load balancer, you can serve both types of content from a Amazon CloudFront web distribution.

Use an origin group with primary and secondary origins to configure Amazon CloudFront for high-availability and failover

You can set up Amazon CloudFront with origin failover for scenarios that require high availability. To get started, you create an origin group with two origins: a primary and a secondary. If the primary origin is unavailable or returns specific HTTP response status codes that indicate a failure, CloudFront automatically switches to the secondary origin.

To set up origin failover, you must have a distribution with at least two origins. Next, you create an origin group for your distribution that includes two origins, setting one as the primary. Finally, you create or update a cache behavior to use the origin group.



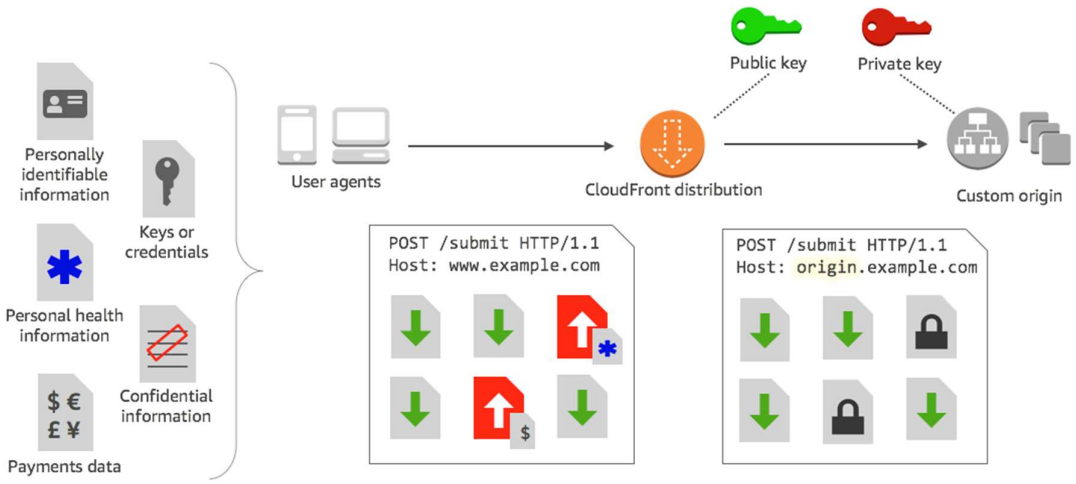
via

- https://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/high_availability_origin_failover.html

Use field level encryption in Amazon CloudFront to protect sensitive data for specific content

Field-level encryption allows you to enable your users to securely upload sensitive information to your web servers. The sensitive information provided by your users is encrypted at the edge, close to the user, and remains encrypted throughout your entire application stack. This encryption ensures that only applications that need the data—and have the credentials to decrypt it—are able to do so.

To use field-level encryption, when you configure your Amazon CloudFront distribution, specify the set of fields in POST requests that you want to be encrypted, and the public key to use to encrypt them. You can encrypt up to 10 data fields in a request. (You can't encrypt all of the data in a request with field-level encryption; you must specify individual fields to encrypt.)



via - https://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/field-level-encryption.html

Incorrect options:

Use AWS Key Management Service (AWS KMS) encryption in Amazon CloudFront to protect sensitive data for specific content - This option has been added as a distractor. You can use field level encryption in Amazon CloudFront to protect sensitive data for specific content.

Use geo restriction to configure Amazon CloudFront for high-availability and failover - You can use geo restriction, also known as geo blocking, to prevent users in specific geographic locations from accessing content that you're distributing through a Amazon CloudFront distribution. Geo restriction is not used to configure Amazon CloudFront for high availability and failover.

Amazon CloudFront can route to multiple origins based on the price class - Amazon CloudFront edge locations are grouped into geographic regions, and AWS has grouped regions into price classes. The default price class includes all regions. Another price class includes most regions (the United States; Canada; Europe; Hong Kong, Philippines, South Korea, Taiwan, and Singapore; Japan; India; South Africa; and Middle East regions) but excludes the most expensive regions. A third price class includes only the least expensive regions (the United States, Canada, and Europe regions). CloudFront can only route to multiple origins based on content type and not on the basis of the price class.

References:

https://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/high_availability_origin_failover.html

https://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/field-level-encryption.html

https://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/PriceClass.html

Domain

Design Resilient Architectures

Question 25Skipped

A leading e-commerce company runs its IT infrastructure on AWS Cloud. The company has a batch job running at 7AM daily on an Amazon RDS database. It processes shipping orders for the past day, and usually gets around 2000 records that need to be processed sequentially in a batch job via a shell script. The processing of each record takes about 3 seconds.

What platform do you recommend to run this batch job?

AWS Glue

Correct answer

Amazon Elastic Compute Cloud (Amazon EC2)

Amazon Kinesis Data Streams

AWS Lambda

Overall explanation

Correct option:

Amazon Elastic Compute Cloud (Amazon EC2)

Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides secure, resizable compute capacity in the cloud. It is designed to make web-scale cloud computing easier for developers. Amazon EC2's simple web service interface allows you to obtain and configure capacity with minimal friction. It provides you with complete control of your computing resources and lets you run on Amazon's proven computing environment. AWS Batch can be used to plan, schedule, and execute your batch computing workloads on Amazon EC2 Instances. Amazon EC2 is the right choice as it can accommodate batch processing and run customized scripts, as is the needed requirement.

Incorrect options:

AWS Glue - AWS Glue is a fully managed extract, transform, and load (ETL) service that makes it easy for customers to prepare and load their data for analytics. AWS Glue job is meant to be used for batch ETL data processing. Glue is for performing ETL, but cannot run custom shell scripts and hence not the right choice here.

Amazon Kinesis Data Streams - Amazon Kinesis Data Streams (KDS) is a massively scalable and durable real-time data streaming service. KDS can continuously capture gigabytes of data per second from hundreds of thousands of sources such as website clickstreams, database event streams, financial transactions, social media feeds, IT logs, and location-tracking events. The throughput of an Amazon Kinesis data stream is designed to scale without limits via increasing the number of shards within a data stream. However, Kinesis works great with real-time data, we are looking at batch processing, so Kinesis is not an option.

AWS Lambda - AWS Lambda lets you run code without provisioning or managing servers. AWS Lambda functions can be configured to run up to 15 minutes per execution. You can set the timeout to any value between 1 second and 15 minutes. The total runtime for the given use-case is 100 minutes (2000*3=6000 seconds = 100 minutes) but the Lambda would time out after 15 minutes, so this option is incorrect.

References:

https://aws.amazon.com/ec2/features/

https://aws.amazon.com/lambda/faqs/

Domain

Design High-Performing Architectures

Question 26Skipped

Your company runs a web portal to match developers to clients who need their help. As a solutions architect, you've designed the architecture of the website to be fully serverless with Amazon API Gateway and AWS Lambda. The backend uses Amazon DynamoDB table. You would like to automatically congratulate your developers on important milestones, such as - their first paid contract. All the contracts are stored in Amazon DynamoDB.

Which Amazon DynamoDB feature can you use to implement this functionality such that there is LEAST delay in sending automatic notifications?

Amazon DynamoDB DAX + Amazon API Gateway

Amazon Simple Queue Service (Amazon SQS) + AWS Lambda

Amazon EventBridge events + AWS Lambda

Correct answer

Amazon DynamoDB Streams + AWS Lambda

Overall explanation

Correct option:

Amazon DynamoDB is a key-value and document database that delivers single-digit millisecond performance at any scale. It's a fully managed, multi-Region, multi-master, durable database with built-in security, backup and restore, and in-memory caching for internet-scale applications.

Amazon DynamoDB Streams + AWS Lambda

Amazon DynamoDB stream is an ordered flow of information about changes to items in Amazon DynamoDB table. When you enable a stream on a table, DynamoDB captures information about every modification to data items in the table. Whenever an application creates, updates, or deletes items in the table, DynamoDB Streams writes a stream record with the primary key attributes of the items that were modified. A stream record contains information about a data modification to a single item in a DynamoDB table.

Amazon DynamoDB Streams will contain a stream of all the changes that happen to an Amazon DynamoDB table. It can be chained with an AWS Lambda function that will be triggered to react to these changes, one of which is the developer's milestone. Therefore, this is the correct option.

Incorrect options:

Amazon DynamoDB DAX + Amazon API Gateway - Amazon DynamoDB Accelerator (DAX) is a fully managed, highly available, in-memory cache for DynamoDB that delivers up to a 10x performance improvement – from milliseconds to microseconds – even at millions of requests per second. DAX does all the heavy lifting required to add in-memory acceleration to your DynamoDB tables, without requiring developers to manage cache invalidation, data population, or cluster management.

DAX is a caching layer and Amazon API Gateway is used to deploy APIs at scale, so this won't help.

Amazon Simple Queue Service (Amazon SQS) + AWS Lambda - Amazon Simple Queue Service (SQS) is a fully managed message queuing service that enables you to decouple and scale microservices, distributed systems, and serverless applications. Amazon SQS offers two types of message queues. Standard queues offer maximum throughput, best-effort ordering, and at-least-once delivery. Amazon SQS FIFO queues are designed to guarantee that messages are processed exactly once, in the exact order that they are sent.

Amazon SQS and AWS Lambda could also work, but one would need to write extra logic to send messages to SQS, whereas our data already lives on Amazon DynamoDB so Amazon DynamoDB Streams is a much better choice.

Amazon EventBridge events + AWS Lambda - You cannot use Amazon DynamoDB as a target for an Amazon EventBridge event, so this option is ruled out.

References:

https://aws.amazon.com/dynamodb/dax/

https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/Streams.html

https://docs.aws.amazon.com/AmazonCloudWatch/latest/events/WhatIsCloudWatchEvents.html

Domain

Design High-Performing Architectures

Question 27Skipped

An IT company runs a high-performance computing (HPC) workload on AWS. The workload requires high network throughput and low-latency network performance along with tightly coupled node-to-node communications. The Amazon EC2 instances are properly sized for compute and storage capacity and are launched using default options.

Which of the following solutions can be used to improve the performance of the workload?

Select an Elastic Inference accelerator while launching Amazon EC2 instances

Select the appropriate capacity reservation while launching Amazon EC2 instances

Select dedicated instance tenancy while launching Amazon EC2 instances

Correct answer

Select a cluster placement group while launching Amazon EC2 instances

Overall explanation

Correct option:

Select a cluster placement group while launching Amazon EC2 instances

When you launch a new Amazon EC2 instance, the EC2 service attempts to place the instance in such a way that all of your instances are spread out across underlying hardware to minimize correlated failures. You can use placement groups to influence the placement of a group of interdependent instances to meet the needs of your workload.

Cluster placement group packs instances close together inside an Availability Zone (AZ). This strategy enables workloads to achieve the low-latency network performance necessary for tightly coupled node-to-node communication that is typical of HPC applications.

More on Cluster placement groups:

Cluster placement groups

A cluster placement group is a logical grouping of instances within a single Availability Zone. A cluster placement group can span peered VPCs in the same Region. Instances in the same cluster placement group enjoy a higher per-flow throughput limit for TCP/IP traffic and are placed in the same high-bisection bandwidth segment of the network.

The following image shows instances that are placed into a cluster placement group.

Cluster placement groups are recommended for applications that benefit from low network latency, high network throughput, or both. They are also recommended when the majority of the network traffic is between the instances in the group. To provide the lowest latency and the highest packet-per-second network performance for your placement group, choose an instance type that supports enhanced networking. For more information, see Enhanced Networking.

We recommend that you launch your instances in the following way:

- Use a single launch request to launch the number of instances that you need in the placement group.
- Use the same instance type for all instances in the placement group.

If you try to add more instances to the placement group later, or if you try to launch more than one instance type in the placement group, you increase your chances of getting an insufficient capacity error.

via - https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/placement-groups.html

Incorrect options:

Select the appropriate capacity reservation while launching Amazon EC2 instances - Capacity Reservations enable you to reserve compute capacity for your Amazon EC2 instances in a specific Availability Zone (AZ) for any duration. This gives you the ability to create and manage Capacity Reservations independently from the billing discounts offered by Savings Plans or regional Reserved Instances (RIs). By creating Capacity Reservations, you ensure that you always have access to Amazon EC2 capacity when you need it, for as long as you need it. Capacity Reservations cannot impact the performance of the underlying Amazon EC2 instances.

Select dedicated instance tenancy while launching Amazon EC2 instances - Dedicated Instances are Amazon EC2 instances that run in a VPC on hardware that's dedicated to a single customer. Your Dedicated instances are physically isolated at the host hardware level from instances that belong to other AWS accounts. Dedicated Instances are useful from a compliance perspective, but are not meant for performance improvement.

Select an Elastic Inference accelerator while launching Amazon EC2 instances - Amazon Elastic Inference accelerators are GPU-powered hardware devices that are designed to work with any Amazon EC2 instance, Amazon Sagemaker instance, or Amazon ECS task to accelerate deep learning inference workloads at a low cost. They are for workloads that need

deep learning. Also, AWS PrivateLink VPC Endpoints are needed for Elastic Inference accelerators, which makes it unsuitable for the current scenario.

Reference:

https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/placement-groups.html

Domain

Design High-Performing Architectures

Question 28Skipped

An enterprise has decided to move its secondary workloads such as backups and archives to AWS cloud. The CTO wishes to move the data stored on physical tapes to Cloud, without changing their current tape backup workflows. The company holds petabytes of data on tapes and needs a cost-optimized solution to move this data to cloud.

What is an optimal solution that meets these requirements while keeping the costs to a minimum?

Use AWS Direct Connect, a cloud service solution that makes it easy to establish a dedicated network connection from on-premises to AWS to transfer data. Once this is done, Amazon S3 can be used to store data at lesser costs

Use AWS VPN connection between the on-premises datacenter and your Amazon VPC. Once this is established, you can use Amazon Elastic File System (Amazon EFS) to get a scalable, fully managed elastic NFS file system for use with AWS Cloud services and on-premises resources

Correct answer

Use Tape Gateway, which can be used to move on-premises tape data onto AWS Cloud. Then, Amazon S3 archiving storage classes can be used to store data cost-effectively for years

Use AWS DataSync, which makes it simple and fast to move large amounts of data online between on-premises storage and AWS Cloud. Data moved to Cloud can then be stored cost-effectively in Amazon S3 archiving storage classes

Overall explanation

Correct option:

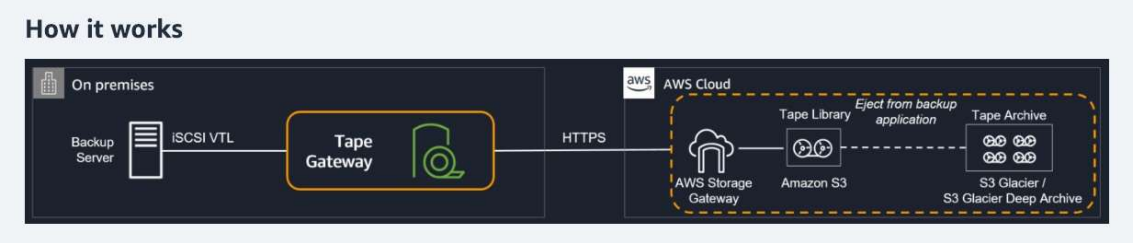
Use Tape Gateway, which can be used to move on-premises tape data onto AWS Cloud. Then, Amazon S3 archiving storage classes can be used to store data cost-effectively for years

Tape Gateway enables you to replace using physical tapes on-premises with virtual tapes in AWS without changing existing backup workflows. Tape Gateway supports all leading backup applications and caches virtual tapes on-premises for low-latency data access. Tape Gateway encrypts data between the gateway and AWS for secure data transfer and compresses data while transitioning virtual tapes between Amazon S3 and Amazon S3 Glacier, or Amazon S3 Glacier Deep Archive, to minimize storage costs.

Tape Gateway compresses and stores archived virtual tapes in the lowest-cost Amazon S3 storage classes, Amazon S3 Glacier and Amazon S3 Glacier Deep Archive. This makes it feasible for you to retain long-term data in the AWS Cloud at a very low cost. With Tape Gateway, you only pay for what you consume, with no minimum commitments and no upfront fees.

Tape Gateway stores your virtual tapes in S3 buckets managed by the AWS Storage Gateway service, so you don't have to manage your own Amazon S3 storage. Tape Gateway integrates with all leading backup applications allowing you to start using cloud storage for on-premises backup and archive without any changes to your backup and archive workflows.

Tape Gateway Overview:



via - https://aws.amazon.com/storagegateway/vtl/

Incorrect options:

Use AWS DataSync, which makes it simple and fast to move large amounts of data online between on-premises storage and AWS Cloud. Data moved to Cloud can then be stored cost-effectively in Amazon S3 archiving storage classes - AWS DataSync supports only NFS and SMB file types and hence is not the right choice for the given use case.

Use AWS Direct Connect, a cloud service solution that makes it easy to establish a dedicated network connection from on-premises to AWS to transfer data. Once this is done, Amazon S3 can be used to store data at lesser costs - AWS Direct Connect is used when customers need to retain on-premises structure because of compliance reasons and have moved the rest of the architecture to AWS Cloud. These businesses generally have an on-going requirement for low latency access to AWS Cloud and hence are willing to spend on installing the physical lines needed for this connection. The given use-case needs a cost-optimized solution and they do not have an ongoing requirement for high availability bandwidth.

Use AWS VPN connection between the on-premises datacenter and your Amazon VPC. Once this is established, you can use Amazon Elastic File System (Amazon EFS) to get a scalable, fully managed elastic NFS file system for use with AWS Cloud services and on-premises resources - VPN connection is used when businesses have an on-going requirement for connectivity from the on-premises data center to AWS Cloud. Amazon EFS is a managed file system by AWS and cannot be used for archiving on-premises tape data onto AWS Cloud.

References:

https://aws.amazon.com/storagegateway/vtl/

https://aws.amazon.com/storagegateway/faqs/

Domain

Design High-Performing Architectures

Question 29Skipped

Amazon Route 53 is configured to route traffic to two Network Load Balancer nodes belonging to two Availability Zones (AZs): AZ-A and AZ-B. Cross-zone load balancing is disabled. AZ-A has four targets and AZ-B has six targets.

Which of the below statements is true about traffic distribution to the target instances from Amazon Route 53?

Each of the four targets in AZ-A receives 8% of the traffic

Each of the four targets in AZ-A receives 10% of the traffic

Each of the six targets in AZ-B receives 10% of the traffic

Correct answer

Each of the four targets in AZ-A receives 12.5% of the traffic

Overall explanation

Correct option:

Each of the four targets in AZ-A receives 12.5% of the traffic

The nodes for your load balancer distribute requests from clients to registered targets. When cross-zone load balancing is enabled, each load balancer node distributes traffic across the registered targets in all enabled Availability Zones (AZs). When cross-zone load balancing is disabled, each load balancer node distributes traffic only across the registered targets in its Availability Zone (AZ).

Amazon Route 53 will distribute traffic such that each load balancer node receives 50% of the traffic from the clients.

If cross-zone load balancing is disabled: 1. Each of the four targets in AZ-A receives 12.5% of the traffic. 2. Each of the six targets in AZ-B receives 8.3% of the traffic.

This is because each load balancer node can route its 50% of the client traffic only to targets in its Availability Zone (AZ).

Incorrect options:

Each of the six targets in AZ-B receives 10% of the traffic - As mentioned above in the correct explanation, each of the six targets in AZ-B receives 8.3% of the traffic.

Each of the four targets in AZ-A receives 8% of the traffic - As mentioned above in the correct explanation, each of the four targets in AZ-A receives 12.5% of the traffic.

Each of the four targets in AZ-A receives 10% of the traffic - As mentioned above in the correct explanation, each of the four targets in AZ-A receives 12.5% of the traffic.

Reference:

https://docs.aws.amazon.com/elasticloadbalancing/latest/userguide/how-elastic-load-balancing-works.html

Domain

Design High-Performing Architectures

Question 30Skipped

A Big Data analytics company writes data and log files in Amazon S3 buckets. The company now wants to stream the existing data files as well as any ongoing file updates from Amazon S3 to Amazon Kinesis Data Streams.

As a Solutions Architect, which of the following would you suggest as the fastest possible way of building a solution for this requirement?

Amazon S3 bucket actions can be directly configured to write data into Amazon Simple Notification Service (Amazon SNS). Amazon SNS can then be used to send the updates to Amazon Kinesis Data Streams

Configure Amazon EventBridge events for the bucket actions on Amazon S3. An AWS Lambda function can then be triggered from the Amazon EventBridge event that will send the necessary data to Amazon Kinesis Data Streams

Leverage Amazon S3 event notification to trigger an AWS Lambda function for the file create event. The AWS Lambda function will then send the necessary data to Amazon Kinesis Data Streams

Correct answer

Leverage AWS Database Migration Service (AWS DMS) as a bridge between Amazon S3 and Amazon Kinesis Data Streams

Overall explanation

Correct option:

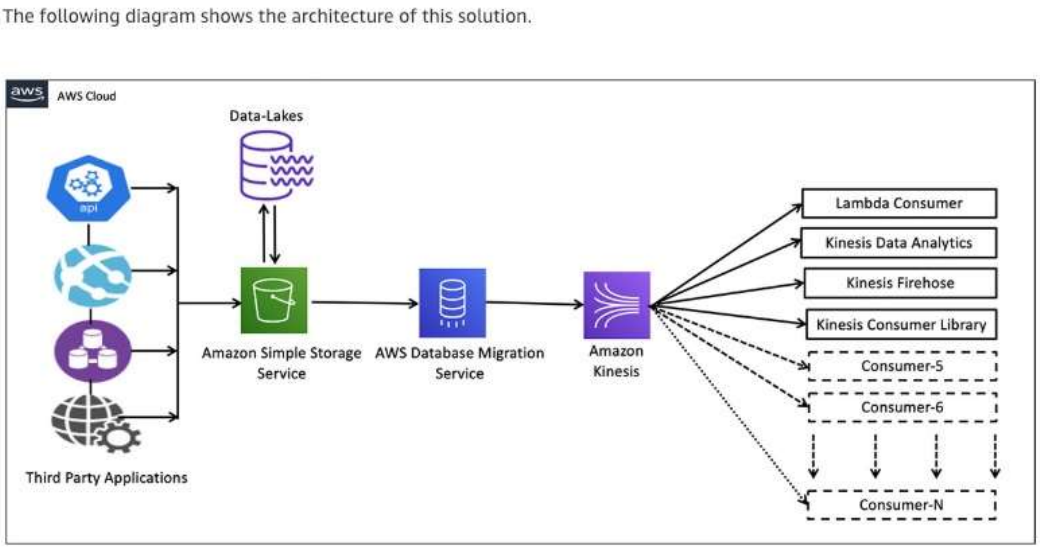
Leverage AWS Database Migration Service (AWS DMS) as a bridge between Amazon S3 and Amazon Kinesis Data Streams

You can achieve this by using AWS Database Migration Service (AWS DMS). AWS DMS enables you to seamlessly migrate data from supported sources to relational databases, data warehouses, streaming platforms, and other data stores in AWS cloud.

The given requirement needs the functionality to be implemented in the least possible time. You can use AWS DMS for such data-processing requirements. AWS DMS lets you expand the existing application to stream data from Amazon S3 into Amazon Kinesis Data Streams for real-time analytics without writing and maintaining new code. AWS DMS supports specifying Amazon S3 as the source and streaming services like Kinesis and Amazon Managed Streaming of Kafka (Amazon MSK) as the target. AWS DMS allows migration of full and change data capture (CDC) files to these services. AWS DMS performs this task out of box without any complex configuration or code development. You can also configure an AWS DMS replication instance to scale up or down depending on the workload.

AWS DMS supports Amazon S3 as the source and Kinesis as the target, so data stored in an S3 bucket is streamed to Kinesis. Several consumers, such as AWS Lambda, Amazon Kinesis Data Firehose, Amazon Kinesis Data Analytics, and the Kinesis Consumer Library (KCL), can consume the data concurrently to perform real-time analytics on the dataset. Each AWS service in this architecture can scale independently as needed.

Architecture of the proposed solution:



via - https://aws.amazon.com/blogs/big-data/streaming-data-from-amazon-s3-to-amazon-kinesis-data-streams-using-aws-dms/

Incorrect options:

Configure Amazon EventBridge events for the bucket actions on Amazon S3. An AWS Lambda function can then be triggered from the Amazon EventBridge event that will send the necessary data to Amazon Kinesis Data Streams - You will need to enable AWS Cloudtrail trail to use object-level actions as a trigger for Amazon EventBridge events. Also, using AWS Lambda functions would require significant custom development to write the data into Amazon Kinesis Data Streams, so this option is not the right fit.

Leverage Amazon S3 event notification to trigger an AWS Lambda function for the file create event. The AWS Lambda function will then send the necessary data to Amazon Kinesis Data Streams - Using AWS Lambda functions would require significant custom development to write the data into Amazon Kinesis Data Streams, so this option is not the right fit.

Amazon S3 bucket actions can be directly configured to write data into Amazon Simple Notification Service (Amazon SNS). Amazon SNS can then be used to send the updates to Amazon Kinesis Data Streams - Amazon S3 cannot directly write data into Amazon SNS, although it can certainly use Amazon S3 event notifications to send an event to Amazon SNS. Also, Amazon SNS cannot directly send messages to Amazon Kinesis Data Streams. So this option is incorrect.

Reference:

https://aws.amazon.com/blogs/big-data/streaming-data-from-amazon-s3-to-amazon-kinesis-data-streams-using-aws-dms/

Domain

Design High-Performing Architectures

Question 31Skipped

An IT company has a large number of clients opting to build their application programming interface (API) using Docker containers. To facilitate the hosting of these containers, the company is looking at various orchestration services available with AWS.

As a Solutions Architect, which of the following solutions will you suggest? (Select two)

Use Amazon EMR for serverless orchestration of the containerized services

Correct selection

Use Amazon Elastic Kubernetes Service (Amazon EKS) with AWS Fargate for serverless orchestration of the containerized services

Correct selection

Use Amazon Elastic Container Service (Amazon ECS) with AWS Fargate for serverless orchestration of the containerized services

Use Amazon Elastic Container Service (Amazon ECS) with Amazon EC2 for serverless orchestration of the containerized services

Use Amazon SageMaker for serverless orchestration of the containerized services

Overall explanation

Correct options:

Use Amazon Elastic Kubernetes Service (Amazon EKS) with AWS Fargate for serverless orchestration of the containerized services

Use Amazon Elastic Container Service (Amazon ECS) with AWS Fargate for serverless orchestration of the containerized services

Building APIs with Docker containers has been gaining momentum over the years. For hosting and exposing these container-based APIs, they need a solution which supports HTTP requests routing, autoscaling, and high availability. In some cases, user authorization is also needed.

For this purpose, many organizations are orchestrating their containerized services with Amazon Elastic Container Service (Amazon ECS) or Amazon Elastic Kubernetes Service (Amazon EKS), while hosting their containers on Amazon EC2 or AWS Fargate. Then, they can add scalability and high availability with Service Auto Scaling (in Amazon ECS) or Horizontal Pod Auto Scaler (in Amazon EKS), and they expose the services through load balancers.

When you use Amazon ECS as an orchestrator (with EC2 or Fargate launch type), you also have the option to expose your services with Amazon API Gateway and AWS Cloud Map instead of a load balancer. AWS Cloud Map is used for service discovery: no matter how Amazon ECS tasks scale, AWS Cloud Map service names would point to the right set of Amazon ECS tasks. Then, API Gateway HTTP APIs can be used to define API routes and point them to the corresponding AWS Cloud Map services.

Incorrect options:

Use Amazon Elastic Container Service (Amazon ECS) with Amazon EC2 for serverless orchestration of the containerized services - Amazon EC2 can be used to host the container services. Amazon EC2 needs hosting and management of the instance, hence does not come under serverless solution. Fargate can be used for serverless container solutions.

Use Amazon EMR for serverless orchestration of the containerized services - Amazon EMR is a web service that enables businesses, researchers, data analysts, and developers to easily and cost-effectively process vast amounts of data. It utilizes a hosted Hadoop framework running on the web-scale infrastructure of Amazon EC2 and Amazon S3. EMR is not a docker orchestration service, as required for the use case.

Use Amazon SageMaker for serverless orchestration of the containerized services - Amazon SageMaker helps data scientists and developers to prepare, build, train, and deploy high-quality machine learning (ML) models quickly by bringing together a broad set of capabilities purpose-built for ML. A powerful tool, SageMaker is not a docker orchestration service, as required for the use case.

Reference:

https://aws.amazon.com/blogs/architecture/field-notes-serverless-container-based-apis-with-amazon-ecs-and-amazon-api-gateway/

Domain

Design Resilient Architectures

Question 32Skipped

A retail company is using AWS Site-to-Site VPN connections for secure connectivity to its AWS cloud resources from its on-premises data center. Due to a surge in traffic across the VPN connections to the AWS cloud, users are experiencing slower VPN connectivity.

Which of the following options will maximize the VPN throughput?

Use AWS Global Accelerator for the VPN connection to maximize the throughput

Use Transfer Acceleration for the VPN connection to maximize the throughput

Correct answer

Create an AWS Transit Gateway with equal cost multipath routing and add additional VPN tunnels

Create a virtual private gateway with equal cost multipath routing and multiple channels

Overall explanation

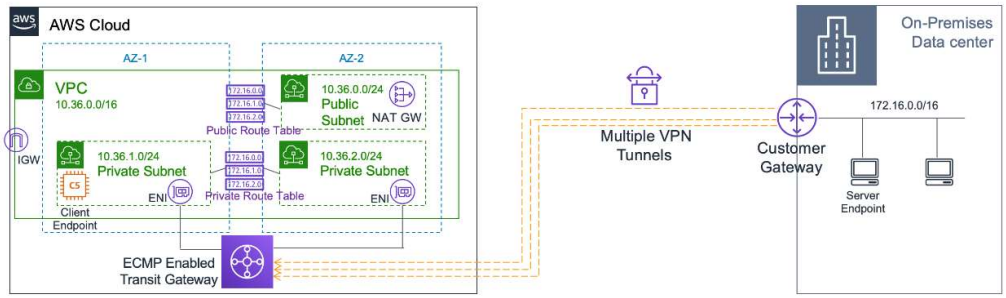
Correct option:

Create an AWS Transit Gateway with equal cost multipath routing and add additional VPN tunnels

VPN connection is a secure connection between your on-premises equipment and your VPCs. Each VPN connection has two VPN tunnels which you can use for high availability. A VPN tunnel is an encrypted link where data can pass from the customer network to or from AWS. The following diagram shows the high-level connectivity with virtual private gateways.

With AWS Transit Gateway, you can simplify the connectivity between multiple VPCs and also connect to any VPC attached to AWS Transit Gateway with a single VPN connection. AWS Transit Gateway also enables you to scale the IPsec VPN throughput with equal cost multi-path (ECMP) routing support over multiple VPN tunnels. A single VPN tunnel still has a maximum throughput of 1.25 Gbps. If you establish multiple VPN tunnels to an ECMP-enabled transit gateway, it can scale beyond the default maximum limit of 1.25 Gbps. You also must enable the dynamic routing option on your transit gateway to be able to take advantage of ECMP for scalability.

Transit gateway with ECMP over multiple VPN connections



via - https://aws.amazon.com/premiumsupport/knowledge-center/transit-gateway-ecmp-multiple-tunnels/

Incorrect options:

Use Transfer Acceleration for the VPN connection to maximize the throughput - Transfer Acceleration is an Amazon S3 bucket-level feature that enables fast, easy, and secure transfers of files over long distances between your client and an S3 bucket. Transfer Acceleration is designed to optimize transfer speeds from across the world into S3 buckets. Transfer Acceleration takes advantage of the globally distributed edge locations in Amazon CloudFront.

This option has been added as a distractor as it is not relevant to AWS VPN connections.

Use AWS Global Accelerator for the VPN connection to maximize the throughput - AWS Global Accelerator is a networking service that improves the performance of your users' traffic by up to 60% using the global network infrastructure of AWS. When the internet is congested, AWS Global Accelerator optimizes the path to your application to keep packet loss, jitter, and latency consistently low. With Global Accelerator, you are provided two global static public IPs that act as a fixed entry point to your application, improving availability. Global Accelerator automatically re-routes your traffic to your nearest healthy available endpoint to mitigate endpoint failure.

AWS Global Accelerator can be used to optimize the network path, using the congestion-free AWS global network to route traffic to the endpoint that provides the best application performance . You can use an accelerated VPN connection to avoid network disruptions that might occur when traffic is routed over the public internet. AWS Global Accelerator will not maximize the VPN throughput, so it is not the best fit for the given use case.

Create a virtual private gateway with equal cost multipath routing and multiple channels - A virtual private gateway is the VPN endpoint on the Amazon side of your Site-to-Site VPN connection that can be attached to a single VPC. A virtual private gateway does not support equal cost multi-path (ECMP) routing, so this option is incorrect.

References:

https://docs.aws.amazon.com/AmazonS3/latest/userguide/transfer-acceleration.html

https://aws.amazon.com/premiumsupport/knowledge-center/transit-gateway-ecmp-multiple-tunnels/

Domain

Design High-Performing Architectures

Question 33Skipped

You have an Amazon S3 bucket that contains files in two different folders - s3://my-bucket/images and s3://my-bucket/thumbnails. When an image is first uploaded and new, it is viewed several times. But after 45 days, analytics prove that image files are on average rarely requested, but the thumbnails still are. After 180 days, you would like to archive the image files and the thumbnails. Overall you would like the solution to remain highly available to prevent disasters happening against a whole Availability Zone (AZ).

How can you implement an efficient cost strategy for your Amazon S3 bucket? (Select two)

Create a Lifecycle Policy to transition objects to Amazon S3 One Zone IA using a prefix after 45 days

Create a Lifecycle Policy to transition all objects to Amazon S3 Standard IA after 45 days

Correct selection

Create a Lifecycle Policy to transition all objects to Amazon S3 Glacier after 180 days

Create a Lifecycle Policy to transition objects to Amazon S3 Glacier using a prefix after 180 days

Correct selection

Create a Lifecycle Policy to transition objects to Amazon S3 Standard IA using a prefix after 45 days

Overall explanation

Correct options:

To manage your S3 objects, so they are stored cost-effectively throughout their lifecycle, configure their Amazon S3 Lifecycle. An S3 Lifecycle configuration is a set of rules that define actions that Amazon S3 applies to a group of objects. There are two types of actions:

Transition actions — Define when objects transition to another storage class. For example, you might choose to transition objects to the S3 Standard-IA storage class 30 days after you created them, or archive objects to the S3 Glacier storage class one year after creating them.

Expiration actions — Define when objects expire. Amazon S3 deletes expired objects on your behalf.

Create a Lifecycle Policy to transition objects to Amazon S3 Standard IA using a prefix after 45 days

Amazon S3 Standard-IA is for data that is accessed less frequently but requires rapid access when needed. Amazon S3 Standard-IA offers high durability, high throughput, and low latency of S3 Standard, with a low per gigabyte storage price and per gigabyte retrieval fee. This combination of low cost and high performance makes Amazon S3 Standard-IA ideal for long-term storage, backups, and as a data store for disaster recovery files. The minimum storage duration charge is 30 days.

As the use-case mentions that after 45 days, image files are rarely requested, but the thumbnails still are. So you need to use a prefix while configuring the Lifecycle Policy so that only objects in the s3://my-bucket/images are transitioned to Standard IA and not all the objects in the bucket.

Create a Lifecycle Policy to transition all objects to Amazon S3 Glacier after 180 days

Amazon S3 Glacier and S3 Glacier Deep Archive are secure, durable, and extremely low-cost Amazon S3 cloud storage classes for data archiving and long-term backup. They are designed to deliver 99.999999999% durability, and provide comprehensive security and compliance capabilities that can help meet even the most stringent regulatory requirements.

Incorrect options:

Create a Lifecycle Policy to transition all objects to Amazon S3 Standard IA after 45 days - As discussed above, you need to use a prefix while configuring the Lifecycle Policy so that only objects in the s3://my-bucket/images are transitioned to Amazon S3 Standard IA and not all the objects in the bucket.

Create a Lifecycle Policy to transition objects to Amazon S3 Glacier using a prefix after 180 days - After 180 days, you can move all the objects to Amazon S3 Glacier storage as per the use case. Glacier doesn't need prefixes for the given use-case.

Create a Lifecycle Policy to transition objects to Amazon S3 One Zone IA using a prefix after 45 days - Amazon S3 Standard-IA is for data that is accessed less frequently but requires rapid access when needed. Amazon S3 Standard-IA offers high durability, high throughput, and low latency of Amazon S3 Standard, with a low per GB storage price and per GB retrieval fee. This combination of low cost and high performance makes S3 Standard-IA ideal for long-term storage, backups, and as a data store for disaster recovery files. The minimum storage duration charge is 30 days.

Amazon S3 One Zone-IA is for data that is accessed less frequently but requires rapid access when needed. Unlike other S3 Storage Classes which store data in a minimum of three Availability Zones (AZs), S3 One Zone-IA stores data in a single AZ and costs 20% less than S3 Standard-IA. The minimum storage duration charge is 30 days.

Finally, Amazon S3 One Zone IA will not achieve the necessary availability in case an Availability Zone (AZ) goes down.

References:

https://aws.amazon.com/s3/storage-classes/

https://docs.aws.amazon.com/AmazonS3/latest/dev/object-lifecycle-mgmt.html

Domain

Design Cost-Optimized Architectures

Question 34Skipped

A niche social media application allows users to connect with sports athletes. As a solutions architect, you've designed the architecture of the application to be fully serverless using Amazon API Gateway and AWS Lambda. The backend uses an Amazon DynamoDB table. Some of the star athletes using the application are highly popular, and therefore Amazon DynamoDB has increased the read capacity units (RCUs). Still, the application is experiencing a hot partition problem.

What can you do to improve the performance of Amazon DynamoDB and eliminate the hot partition problem without a lot of application refactoring?

Use Amazon ElastiCache

Use Amazon DynamoDB Streams

Use Amazon DynamoDB Global Tables

Correct answer

Use Amazon DynamoDB DAX

Overall explanation

Correct option:

Use Amazon DynamoDB DAX

Amazon DynamoDB is a key-value and document database that delivers single-digit millisecond performance at any scale. It's a fully managed, multi-Region, multi-master, durable database with built-in security, backup and restore, and in-memory caching for internet-scale applications.

Amazon DynamoDB Accelerator (DAX) is a fully managed, highly available, in-memory cache for DynamoDB that delivers up to a 10x performance improvement – from milliseconds to microseconds – even at millions of requests per second. DAX does all the heavy lifting required to add in-memory acceleration to your DynamoDB tables, without requiring developers to manage cache invalidation, data population, or cluster management.

DAX will be transparent and won't require an application refactoring, and will cache the "hotkeys". Therefore, this is the correct option.

Incorrect options:

Use Amazon DynamoDB Global Tables - Amazon DynamoDB Global Tables builds upon DynamoDB's global footprint to provide you with a fully managed, multi-region, and multi-master database that provides fast, local, read and write performance for massively scaled,

global applications. Global Tables replicates your Amazon DynamoDB tables automatically across your choice of AWS regions. But Global Tables cannot address the hotkey issue.

Use Amazon DynamoDB Streams - Amazon DynamoDB stream is an ordered flow of information about changes to items in a DynamoDB table. When you enable a stream on a table, DynamoDB captures information about every modification to data items in the table. Whenever an application creates, updates, or deletes items in the table, DynamoDB Streams writes a stream record with the primary key attributes of the items that were modified. A stream record contains information about a data modification to a single item in a DynamoDB table. DynamoDB Streams cannot address the hotkey issue.

Use Amazon ElastiCache - Amazon ElastiCache allows you to seamlessly set up, run, and scale popular open-Source compatible in-memory data stores in the cloud. Build data-intensive apps or boost the performance of your existing databases by retrieving data from high throughput and low latency in-memory data stores. Amazon ElastiCache is a popular choice for real-time use cases like Caching, Session Stores, Gaming, Geospatial Services, Real-Time Analytics, and Queuing. ElastiCache could also be a solution, but it will require a lot of refactoring work on the AWS Lambda side.

References:

https://aws.amazon.com/dynamodb/dax/

https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/Streams.html

Domain

Design High-Performing Architectures

Question 35Skipped

A ride-sharing company wants to use an Amazon DynamoDB table for data storage. The table will not be used during the night hours whereas the read and write traffic will often be unpredictable during day hours. When traffic spikes occur they will happen very quickly.

Which of the following will you recommend as the best-fit solution?

Set up Amazon DynamoDB table with a global secondary index

Set up Amazon DynamoDB global table in the provisioned capacity mode

Set up Amazon DynamoDB table in the provisioned capacity mode with auto-scaling enabled

Correct answer

Set up Amazon DynamoDB table in the on-demand capacity mode

Overall explanation

Correct option:

Set up Amazon DynamoDB table in the on-demand capacity mode

Amazon DynamoDB has two read/write capacity modes for processing reads and writes on your tables:

On-demand

Provisioned (default, free-tier eligible)

Amazon DynamoDB on-demand is a flexible billing option capable of serving thousands of requests per second without capacity planning. DynamoDB on-demand offers pay-per-request pricing for read and write requests so that you pay only for what you use.

The on-demand mode is a good option if any of the following are true:

You create new tables with unknown workloads.

You have unpredictable application traffic.

You prefer the ease of paying for only what you use.

If you choose provisioned mode, you specify the number of reads and writes per second that you require for your application. You can use auto-scaling to adjust your table's provisioned capacity automatically in response to traffic changes. This helps you govern your DynamoDB use to stay at or below a defined request rate to obtain cost predictability.

Provisioned mode is a good option if any of the following are true:

You have predictable application traffic.

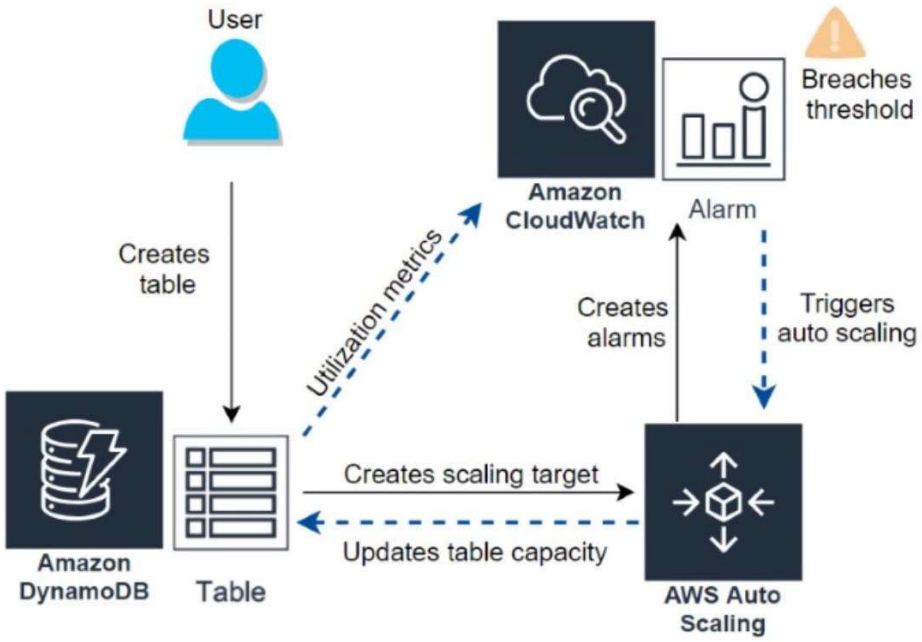
You run applications whose traffic is consistent or ramps gradually.

You can forecast capacity requirements to control costs.

Background: How DynamoDB auto scaling works

When you create a DynamoDB table, auto scaling is the default capacity setting, but you can also enable auto scaling on any table that does not have it active. Behind the scenes, as illustrated in the following diagram, DynamoDB auto scaling uses a scaling policy in Application Auto Scaling. To configure auto scaling in DynamoDB, you set the minimum and maximum levels of read and write capacity in addition to the target utilization percentage. Auto scaling uses Amazon CloudWatch to monitor a table's read and write capacity metrics. To do so, it creates CloudWatch alarms that track consumed capacity.

The upper threshold alarm is triggered when consumed reads or writes breach the target utilization percent for two consecutive minutes. The lower threshold alarm is triggered after traffic falls below the target utilization minus 20 percent for 15 consecutive minutes. When an alarm is triggered, CloudWatch initiates auto scaling activity, which checks the consumed capacity and updates the provisioned capacity of the table. For example, if you set the read capacity units (RCUs) at 100 and the target utilization to 80 percent, auto scaling increases the provisioned capacity when utilization exceeds 80 RCUs for two consecutive minutes, or decreases capacity when consumption falls below 60 RCUs (80 percent minus 20 percent) for fifteen minutes.



via - https://aws.amazon.com/blogs/database/amazon-dynamodb-auto-scaling-performance-and-cost-optimization-at-any-scale/

With on-demand, Amazon DynamoDB instantly allocates capacity as it is needed. There is no concept of provisioned capacity, and there is no delay waiting for Amazon CloudWatch thresholds or the subsequent table updates. On-demand is ideal for bursty, new, or unpredictable workloads whose traffic can spike in seconds or minutes, and when underprovisioned capacity would impact the user experience. On-demand is a perfect solution if your team is moving to a NoOps or serverless environment.

The given use case clearly states that when the traffic spikes occur they happen very quickly, thereby implying an unpredictable traffic pattern, therefore the on-demand capacity mode is the correct option for the given use case.

Incorrect options:

Set up Amazon DynamoDB table in the provisioned capacity mode with auto-scaling enabled - As mentioned in the explanation above, you should use the provisioned capacity mode (even with auto-scaling) only when you have predictable application traffic.

When you create Amazon DynamoDB table, auto-scaling is the default capacity setting, but you can also enable auto-scaling on any table that does not have it active. Behind the scenes, as illustrated in the following diagram, DynamoDB auto scaling uses a scaling policy in Application Auto Scaling. To configure auto-scaling in DynamoDB, you set the minimum and maximum levels of read and write capacity in addition to the target utilization percentage. Auto-scaling uses Amazon CloudWatch to monitor a table's read and write capacity metrics. To do so, it creates CloudWatch alarms that track consumed capacity.

Set up Amazon DynamoDB table with a global secondary index - A global secondary index (GSI) is an index with a partition key and a sort key that can be different from those on the base table. A global secondary index is considered "global" because queries on the index can span all of the data in the base table, across all partitions. A global secondary index is stored in its own partition space away from the base table and scales separately from the base table. GSI cannot be used to handle an unpredictable load on a DynamoDB table.

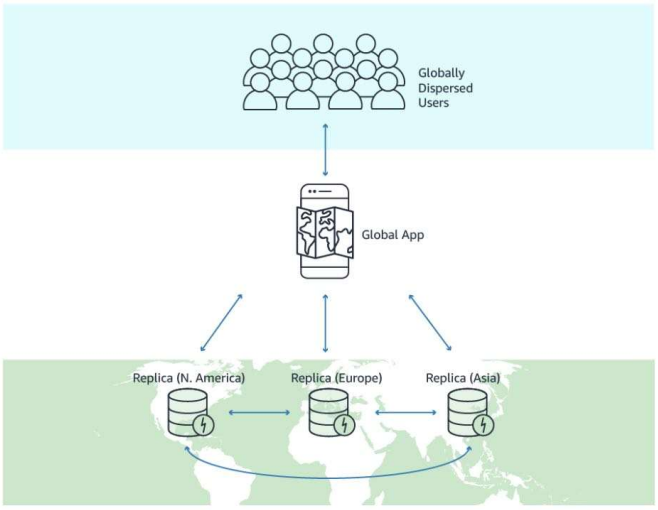
Set up Amazon DynamoDB global table in the provisioned capacity mode - Global tables build on the global Amazon DynamoDB footprint to provide you with a fully managed, multi-Region, and multi-active database that delivers fast, local, read and write performance for massively scaled, global applications. Global tables replicate your DynamoDB tables automatically across your choice of AWS Regions.

Global tables eliminate the difficult work of replicating data between Regions and resolving update conflicts, enabling you to focus on your application's business logic. In addition, global tables enable your applications to stay highly available even in the unlikely event of isolation or degradation of an entire Region.

Amazon DynamoDB global tables:

How it works

When you create a DynamoDB global table, it consists of multiple replica tables (one per AWS Region) that DynamoDB treats as a single unit. Every replica has the same table name and the same primary key schema. When an application writes data to a replica table in one Region, DynamoDB propagates the write to the other replica tables in the other AWS Regions automatically.



For example, suppose that you have a large customer base spread across three geographic areas—the US East Coast, the US West Coast, and Western Europe. Customers can update their profile information by using your application. Without a managed replication solution, you could write code to replicate data changes among the tables for each of these Regions. However, doing this would be a time-consuming and labor-intensive effort.

Instead of writing your own code, you could create a global table referencing your three Region tables and DynamoDB would then automatically replicate data changes among those tables so that changes to one Region would seamlessly propagate to the other Regions. In addition, if one of the AWS Regions were to become temporarily unavailable, your customers could still access the same data in the other Regions.

via - https://aws.amazon.com/dynamodb/global-tables/

Amazon DynamoDB global table cannot be used to handle an unpredictable load on a Amazon DynamoDB table.

References:

https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/HowItWorks.ReadWriteCapacityMode.html

https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/SecondaryIndexes.html

https://aws.amazon.com/blogs/database/amazon-dynamodb-auto-scaling-performance-and-cost-optimization-at-any-scale/

https://aws.amazon.com/dynamodb/global-tables/

Domain

Design High-Performing Architectures

Question 36Skipped

A Pharmaceuticals company is looking for a simple solution to connect its VPCs and on-premises networks through a central hub.

As a Solutions Architect, which of the following would you suggest as the solution that requires the LEAST operational overhead?

Use Transit VPC Solution to connect the Amazon VPCs to the on-premises networks

Correct answer

Use AWS Transit Gateway to connect the Amazon VPCs to the on-premises networks

Partially meshed VPC peering can be used to connect the Amazon VPCs to the on-premises networks

Fully meshed VPC peering can be used to connect the Amazon VPCs to the on-premises networks

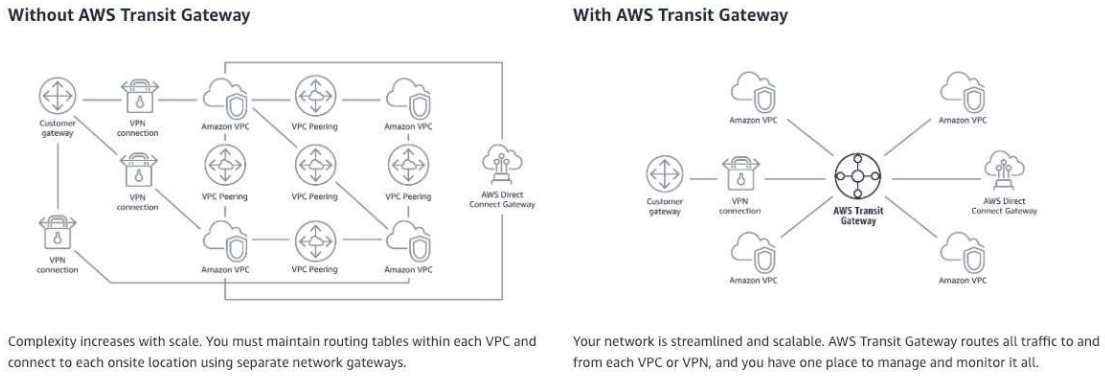
Overall explanation

Correct option:

Use AWS Transit Gateway to connect the Amazon VPCs to the on-premises networks

The AWS Transit Gateway allows customers to connect their Amazon VPCs and their on-premises networks to a single gateway. As your number of workloads running on AWS increases, you need to be able to scale your networks across multiple accounts and Amazon VPCs to keep up with the growth. With AWS Transit Gateway, you only have to create and manage a single connection from the central gateway into each Amazon VPC, on-premises data center, or remote office across your network. AWS Transit Gateway acts as a hub that controls how traffic is routed among all the connected networks, which act like spokes. This hub and spoke model simplifies management and reduces operational costs because each network only has to connect to the Transit Gateway and not to every other network.

AWS Transit Gateway:



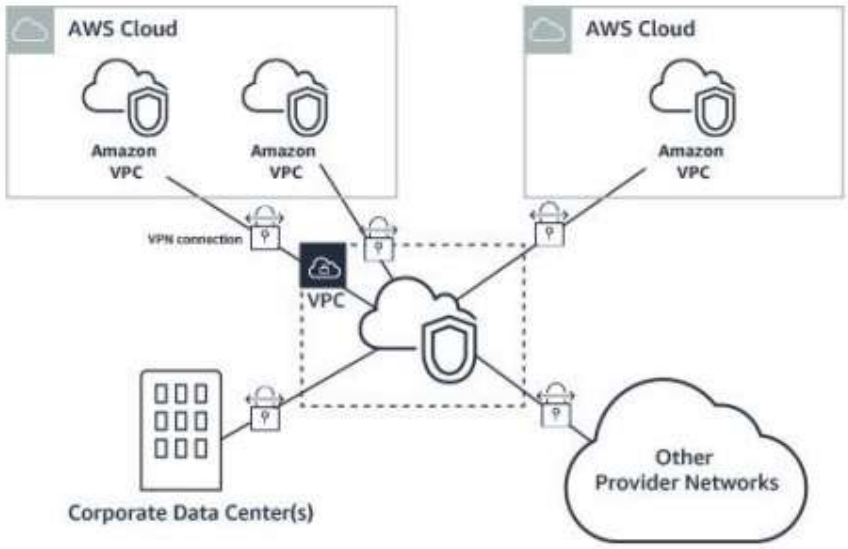
via - https://aws.amazon.com/transit-gateway/

Incorrect options:

Use Transit VPC Solution to connect the Amazon VPCs to the on-premises networks - The Transit VPC can be used to enable connectivity between various VPC's in different regions and customer data centers. You can use this to connect multiple VPCs that are geographically disparate and/or running in separate AWS accounts, to a common VPC that serves as a global

network transit center. This network topology simplifies network management and minimizes the number of connections that you need to set up.

Transit VPC:



via - https://aws.amazon.com/transit-gateway/

Transit VPC is not the right solution for this use-case as Transit Gateway provides several advantages over Transit VPC: 1. Transit Gateway abstracts away the complexity of maintaining VPN connections with hundreds of VPCs. 2. Transit Gateway removes the need to manage and scale Amazon EC2 based software appliances. AWS is responsible for managing all resources needed to route traffic. 3. Transit Gateway removes the need to manage high availability by providing a highly available and redundant Multi-AZ infrastructure. 4. Transit Gateway improves bandwidth for inter-VPC communication to burst speeds of 50 Gbps per Availability Zone (AZ). 5. Transit Gateway streamlines user costs to a simple per hour per/GB transferred model. 6. Transit Gateway decreases latency by removing Amazon EC2 proxies and the need for VPN encapsulation.

Partially meshed VPC peering can be used to connect the Amazon VPCs to the on-premises networks

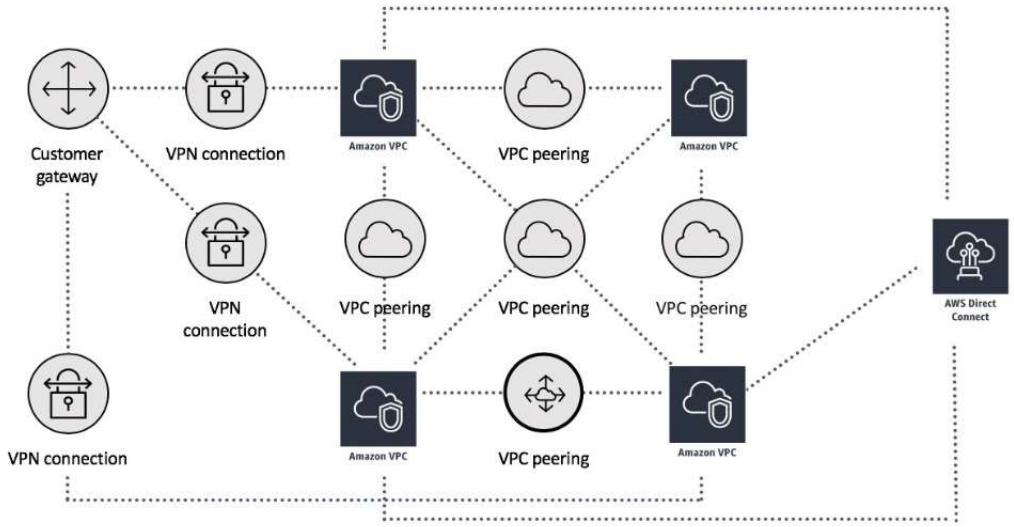
Fully meshed VPC peering can be used to connect the Amazon VPCs to the on-premises networks

The simplest way to connect two VPCs is to use VPC Peering. In this setup, a connection enables full bidirectional connectivity between the VPCs. This peering connection is used to route traffic between the VPCs. VPCs across accounts and AWS Regions can also be peered together. VPC peering only incurs costs for traffic traveling over the connection (there is no hourly infrastructure fee).

VPC peering is point-to-point connectivity, and it does not support transitive routing. If you are using VPC peering, on-premises connectivity (VPN and/or Direct Connect) must be made to each VPC. Resources in a VPC cannot reach on-premises using the hybrid connectivity of a peered VPC. VPC peering is best used when resources in one VPC must communicate with

resources in another VPC, the environment of both VPCs is controlled and secured, and the number of VPCs to be connected is less than 10 (to allow for the individual management of each connection). VPC peering offers the lowest overall cost when compared to other options for inter-VPC connectivity.

Network setup using VPC Peering:



via - https://docs.aws.amazon.com/whitepapers/latest/building-scalable-secure-multi-vpc-network-infrastructure/vpc-peering.html

You cannot use VPC Peering to establish on-premises connectivity with AWS Cloud, so both these options are incorrect.

References:

https://docs.aws.amazon.com/whitepapers/latest/building-scalable-secure-multi-vpc-network-infrastructure/transit-gateway-vs-transit-vpc.html

https://docs.aws.amazon.com/whitepapers/latest/building-scalable-secure-multi-vpc-network-infrastructure/vpc-peering.html

Domain

Design High-Performing Architectures

Question 37Skipped

A Big Data processing company has created a distributed data processing framework that performs best if the network performance between the processing machines is high. The application has to be deployed on AWS, and the company is only looking at performance as the key measure.

As a Solutions Architect, which deployment do you recommend?

Use Spot Instances

Optimize the Amazon EC2 kernel using EC2 User Data

Use a Spread placement group

Correct answer

Use a Cluster placement group

Overall explanation

Correct option:

When you launch a new Amazon EC2 instance, the EC2 service attempts to place the instance in such a way that all of your instances are spread out across underlying hardware to minimize correlated failures. You can use placement groups to influence the placement of a group of interdependent instances to meet the needs of your workload. Depending on the type of workload, you can create a placement group using one of the following placement strategies:

Cluster – packs instances close together inside an Availability Zone (AZ). This strategy enables workloads to achieve the low-latency network performance necessary for tightly-coupled node-to-node communication that is typical of HPC applications.

Partition – spreads your instances across logical partitions such that groups of instances in one partition do not share the underlying hardware with groups of instances in different partitions. This strategy is typically used by large distributed and replicated workloads, such as Hadoop, Cassandra, and Kafka.

Spread – strictly places a small group of instances across distinct underlying hardware to reduce correlated failures.

There is no charge for creating a placement group.

Use a Cluster placement group

A cluster placement group is a logical grouping of instances within a single Availability Zone (AZ). A cluster placement group can span peered VPCs in the same Region. Instances in the same cluster placement group enjoy a higher per-flow throughput limit of up to 10 Gbps for TCP/IP traffic and are placed in the same high-bisection bandwidth segment of the network.

Cluster placement groups are recommended for applications that benefit from low network latency, high network throughput, or both. They are also recommended when the majority of the network traffic is between the instances in the group. To provide the lowest latency and the highest packet-per-second network performance for your placement group, choose an instance type that supports enhanced networking.

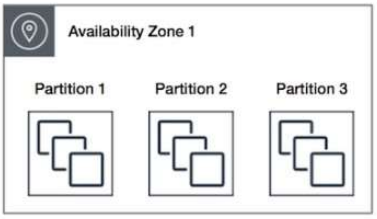
Image of Cluster placement group:

The following image shows instances that are placed into a cluster placement group.

via - https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/placement-groups.html

Image of Partition placement group:

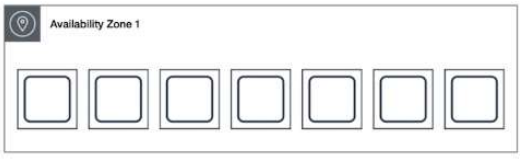
The following image is a simple visual representation of a partition placement group in a single Availability Zone. It shows instances that are placed into a partition placement group with three partitions—**Partition 1**, **Partition 2**, and **Partition 3**. Each partition comprises multiple instances. The instances in a partition do not share racks with the instances in the other partitions, allowing you to contain the impact of a single hardware failure to only the associated partition.



via - https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/placement-groups.html

Image of Spread placement group:

The following image shows seven instances in a single Availability Zone that are placed into a spread placement group. The seven instances are placed on seven different racks.



via - https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/placement-groups.html

Incorrect options:

Use Spot Instances - A Spot Instance is an unused Amazon EC2 instance that is available for less than the On-Demand price. Because Spot Instances enable you to request unused Amazon EC2 instances at steep discounts, you can lower your Amazon EC2 costs significantly. Spot

Instances are a cost-effective choice if you can be flexible about when your applications run and if your applications can be interrupted. Since performance is the key criteria, this is not the right choice.

Optimize the Amazon EC2 kernel using EC2 User Data - Optimizing the Amazon EC2 kernel won't help with network performance as it's bounded by the EC2 instance type mainly. Therefore, this option is incorrect.

Use a Spread placement group - A spread placement group is a group of instances that are each placed on distinct racks, with each rack having its own network and power source. The instances are placed across distinct underlying hardware to reduce correlated failures. A spread placement group can span multiple Availability Zones (AZs) in the same Region. You can have a maximum of seven running instances per Availability Zone (AZ) per group.

Reference:

https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/placement-groups.html

Domain

Design High-Performing Architectures

Question 38Skipped

You started a new job as a solutions architect at a company that has both AWS experts and people learning AWS. Recently, a developer misconfigured a newly created Amazon RDS database which resulted in a production outage.

How can you ensure that Amazon RDS specific best practices are incorporated into a reusable infrastructure template to be used by all your AWS users?

Store your recommendations in a custom AWS Trusted Advisor rule

Correct answer

Use AWS CloudFormation to manage Amazon RDS databases

Attach an IAM policy to interns preventing them from creating an Amazon RDS database

Create an AWS Lambda function which sends emails when it finds misconfigured Amazon RDS databases

Overall explanation

Correct option:

Use AWS CloudFormation to manage Amazon RDS databases

AWS CloudFormation provides a common language for you to model and provision AWS and third-party application resources in your cloud environment. AWS CloudFormation allows you to use programming languages or a simple text file to model and provision, in an automated and secure manner, all the resources needed for your applications across all regions and accounts. This gives you a single source of truth for your AWS and third-party resources.

AWS CloudFormation allows you to keep your infrastructure as code and re-use the best practices around your company for configuration parameters. Therefore, this is the correct option for the given use-case.

Incorrect options:

Store your recommendations in a custom AWS Trusted Advisor rule - AWS Trusted Advisor is an online tool that provides you real-time guidance to help you provision your resources following AWS best practices. Whether establishing new workflows, developing applications, or as part of ongoing improvement, take advantage of the recommendations provided by AWS Trusted Advisor regularly to help keep your solutions provisioned optimally. AWS Trusted Advisor just provides recommendations rather than creating reusable infrastructure templates.

Create an AWS Lambda function which sends emails when it finds misconfigured Amazon RDS databases - Using an AWS Lambda function to scan for a misconfigured Amazon RDS database is a reactive mechanism. It does not help in creating reusable infrastructure templates.

Attach an IAM policy to interns preventing them from creating an Amazon RDS database - Using an IAM policy to prevent interns from creating an Amazon RDS database does not solve the problem of allowing any user to create resources by leveraging reusable infrastructure templates. So, this option is ruled out.

References:

https://aws.amazon.com/premiumsupport/technology/trusted-advisor/

https://aws.amazon.com/cloudformation/

Domain

Design High-Performing Architectures

Question 39Skipped

An e-commerce company has copied 1 petabyte of data from its on-premises data center to an Amazon S3 bucket in the us-west-1 Region using an AWS Direct Connect link. The company now wants to set up a one-time copy of the data to another Amazon S3 bucket in the us-east-1 Region. The on-premises data center does not allow the use of AWS Snowball.

As a Solutions Architect, which of the following options can be used to accomplish this goal? (Select two)

Correct selection

Set up Amazon S3 batch replication to copy objects across Amazon S3 buckets in another Region using S3 console and then delete the replication configuration

Set up Amazon S3 Transfer Acceleration (Amazon S3TA) to copy objects across Amazon S3 buckets in different Regions using S3 console

Use AWS Snowball Edge device to copy the data from one Region to another Region

Correct selection

Copy data from the source bucket to the destination bucket using the aws S3 sync command

Copy data from the source Amazon S3 bucket to a target Amazon S3 bucket using the S3 console

Overall explanation

Correct options:

Copy data from the source bucket to the destination bucket using the aws S3 sync command

The aws S3 sync command uses the CopyObject APIs to copy objects between Amazon S3 buckets. The sync command lists the source and target buckets to identify objects that are in the source bucket but that aren't in the target bucket. The command also identifies objects in the source bucket that have different LastModified dates than the objects that are in the target bucket. The sync command on a versioned bucket copies only the current version of the object—previous versions aren't copied. By default, this preserves object metadata, but the access control lists (ACLs) are set to FULL_CONTROL for your AWS account, which removes any additional ACLs. If the operation fails, you can run the sync command again without duplicating previously copied objects.

You can use the command like so:

aws s3 sync s3://DOC-EXAMPLE-BUCKET-SOURCE s3://DOC-EXAMPLE-BUCKET-TARGET

Set up Amazon S3 batch replication to copy objects across Amazon S3 buckets in another Region using S3 console and then delete the replication configuration

Amazon S3 Batch Replication provides you a way to replicate objects that existed before a replication configuration was in place, objects that have previously been replicated, and objects that have failed replication. This is done through the use of a Batch Operations job.

You should note that batch replication differs from live replication which continuously and automatically replicates new objects across Amazon S3 buckets. You cannot directly use the AWS S3 console to configure cross-Region replication for existing objects. By default, replication only supports copying new Amazon S3 objects after it is enabled using the AWS S3 console. Replication enables automatic, asynchronous copying of objects across Amazon S3 buckets. Buckets that are configured for object replication can be owned by the same AWS account or by different accounts. Object may be replicated to a single destination bucket or multiple destination buckets. Destination buckets can be in different AWS Regions or within the same Region as the source bucket. Once done, you can delete the replication configuration, as it ensures that batch replication is only used for this one-time data copy operation.

If you want to enable live replication for existing objects for your bucket, you must contact AWS Support and raise a support ticket. This is required to ensure that replication is configured correctly.

Incorrect options:

Use AWS Snowball Edge device to copy the data from one Region to another Region - As the given requirement is about copying the data from one AWS Region to another AWS Region, so AWS Snowball Edge cannot be used here. AWS Snowball Edge Storage Optimized is the optimal

data transfer choice if you need to securely and quickly transfer terabytes to petabytes of data to AWS. You can use AWS Snowball Edge Storage Optimized if you have a large backlog of data to transfer or if you frequently collect data that needs to be transferred to AWS and your storage is in an area where high-bandwidth internet connections are not available or cost-prohibitive. AWS Snowball Edge can operate in remote locations or harsh operating environments, such as factory floors, oil and gas rigs, mining sites, hospitals, and on moving vehicles.

Copy data from the source Amazon S3 bucket to a target Amazon S3 bucket using the S3 console - AWS S3 console cannot be used to copy 1 petabytes of data from one bucket to another as it's not feasible. You should note that this option is different from using the replication options on the AWS console, since here you are using the copy and paste options provided on the AWS console, which is suggested for small or medium data volume. You should use S3 sync for the requirement of one-time copy of data.

Set up Amazon S3 Transfer Acceleration (Amazon S3TA) to copy objects across Amazon S3 buckets in different Regions using S3 console - Amazon S3 Transfer Acceleration (Amazon S3TA) is a bucket-level feature that enables fast, easy, and secure transfers of files over long distances between your client and an Amazon S3 bucket. You cannot use Transfer Acceleration to copy objects across Amazon S3 buckets in different Regions using Amazon S3 console.

References:

https://aws.amazon.com/premiumsupport/knowledge-center/move-objects-s3-bucket/

https://aws.amazon.com/snowball/faqs/

https://docs.aws.amazon.com/AmazonS3/latest/userguide/replication.html

Domain

Design Resilient Architectures

Question 40Skipped

As an e-sport tournament hosting company, you have servers that need to scale and be highly available. Therefore you have deployed an Elastic Load Balancing (ELB) with an Auto Scaling group (ASG) across 3 Availability Zones (AZs). When e-sport tournaments are running, the servers need to scale quickly. And when tournaments are done, the servers can be idle. As a general rule, you would like to be highly available, have the capacity to scale and optimize your costs.

What do you recommend? (Select two)

Correct selection

Set the minimum capacity to 2

Use Dedicated hosts for the minimum capacity

Set the minimum capacity to 3

Set the minimum capacity to 1

Correct selection

Use Reserved Instances (RIs) for the minimum capacity

Overall explanation

Correct options:

Set the minimum capacity to 2

An Auto Scaling group contains a collection of Amazon EC2 instances that are treated as a logical grouping for automatic scaling and management. An Auto Scaling group also enables you to use Amazon EC2 Auto Scaling features such as health check replacements and scaling policies. Both maintaining the number of instances in an Auto Scaling group and automatic scaling are the core functionality of the Amazon EC2 Auto Scaling service.

You configure the size of your Auto Scaling group by setting the minimum, maximum, and desired capacity. The minimum and maximum capacity are required to create an Auto Scaling group, while the desired capacity is optional. If you do not define your desired capacity upfront, it defaults to your minimum capacity.

An Auto Scaling group is elastic as long as it has different values for minimum and maximum capacity. All requests to change the Auto Scaling group's desired capacity (either by manual scaling or automatic scaling) must fall within these limits.

Here, even though our ASG is deployed across 3 Availability Zones (AZs), the minimum capacity to be highly available is 2. When we specify 2 as the minimum capacity, the ASG would create these 2 instances in separate Availability Zones (AZs). If demand goes up, the ASG would spin up a new instance in the third Availability Zone (AZ). Later as the demand subsides, the ASG would scale-in and the instance count would be back to 2.

Use Reserved Instances (RIs) for the minimum capacity

Reserved Instances (RIs) provide you with significant savings on your Amazon EC2 costs compared to On-Demand Instance pricing. Reserved Instances are not physical instances, but rather a billing discount applied to the use of On-Demand Instances in your account. These On-Demand Instances must match certain attributes, such as instance type and Region, to benefit from the billing discount. Since minimum capacity will always be maintained, it is cost-effective to choose reserved instances than any other option.

In case of an Availability Zone (AZ) outage, the instance in that Availability Zone (AZ) would go down however the other instance would still be available. The ASG would provision the replacement instance in the third Availability Zone (AZ) to keep the minimum count to 2.

Incorrect options:

Set the minimum capacity to 1 - This is not failure proof, since only one instance will be maintained consistently and this will be from only one Availability Zone (AZ).

Set the minimum capacity to 3 - This is not a cost-effective option, as two instances in two different Availability Zones (AZs) are enough to make the architecture disaster-proof.

Use Dedicated hosts for the minimum capacity - As there is no use-case to utilize existing per-socket, per-core, or per-VM software licenses or to run the instance on a dedicated physical host, so the option to use dedicated hosts for the minimum capacity is ruled out.

References:

https://docs.aws.amazon.com/autoscaling/ec2/userguide/AutoScalingGroup.html

https://docs.aws.amazon.com/autoscaling/ec2/userguide/asg-capacity-limits.html

Domain

Design Cost-Optimized Architectures

Question 41Skipped

A digital media company needs to manage uploads of around 1 terabyte each from an application being used by a partner company.

As a Solutions Architect, how will you handle the upload of these files to Amazon S3?

Use AWS Direct Connect to provide extra bandwidth

Correct answer

Use multi-part upload feature of Amazon S3

Use AWS Snowball

Use Amazon S3 Versioning

Overall explanation

Correct option:

Use multi-part upload feature of Amazon S3

Multi-part upload allows you to upload a single object as a set of parts. Each part is a contiguous portion of the object's data. You can upload these object parts independently and in any order. If transmission of any part fails, you can retransmit that part without affecting other parts. After all parts of your object are uploaded, Amazon S3 assembles these parts and creates the object.

AWS recommends that you use multi-part uploading in the following ways: 1. If you're uploading large objects over a stable high-bandwidth network, use multi-part uploading to maximize the use of your available bandwidth by uploading object parts in parallel for multi-threaded performance. 2. If you're uploading over a spotty network, use multi-part uploading to increase resiliency to network errors by avoiding upload restarts. When using multi-part uploading, you need to retry uploading only parts that are interrupted during the upload. You don't need to restart uploading your object from the beginning.

In general, when your object size reaches 100 megabytes, you should consider using multipart uploads instead of uploading the object in a single operation. If the file is greater than 5 gigabytes in size, you must use multi-part upload to upload that file to Amazon S3.

Incorrect options:

Use Amazon S3 Versioning - Amazon S3 Versioning is a means of keeping multiple variants of an object in the same bucket. You can use versioning to preserve, retrieve, and restore every version of every object stored in your Amazon S3 bucket. With versioning, you can easily recover from both unintended user actions and application failures. When you enable versioning for a bucket, if Amazon S3 receives multiple write requests for the same object simultaneously, it

stores all of the objects. If you overwrite an object, it results in a new object version in the bucket. You can always restore the previous version.

Use AWS Direct Connect to provide extra bandwidth - AWS Direct Connect is a cloud service solution that makes it easy to establish a dedicated network connection from your premises to AWS. Using AWS Direct Connect, you can establish private connectivity between AWS and your datacenter, office, or colocation environment, which in many cases can reduce your network costs, increase bandwidth throughput, and provide a more consistent network experience than Internet-based connections.

AWS Direct Connect lets you establish a dedicated network connection between your network and one of the AWS Direct Connect locations. This dedicated connection can be partitioned into multiple virtual interfaces. This allows you to use the same connection to access public resources such as objects stored in Amazon S3 using public IP address space, and private resources such as Amazon EC2 instances running within an Amazon Virtual Private Cloud (VPC) using private IP space, while maintaining network separation between the public and private environments. Virtual interfaces can be reconfigured at any time to meet your changing needs. This is a physical connection that takes at least a month to set up.

Use AWS Snowball - AWS Snowball Edge Storage Optimized is the optimal choice if you need to securely and quickly transfer dozens of terabytes to petabytes of data to AWS. It provides up to 80 terabytes of usable HDD storage, 40 vCPUs, 1 TB of SATA SSD storage, and up to 40 gigabytes network connectivity to address large scale data transfer and pre-processing use cases.

(The original AWS Snowball devices were transitioned out of service and AWS Snowball Edge Storage Optimized are now the primary devices used for data transfer. You may see the Snowball device on the exam, just remember that the original AWS Snowball device had 80 terabytes of storage space).

References:

https://docs.aws.amazon.com/AmazonS3/latest/dev/uploadobjusingmpu.html

https://docs.aws.amazon.com/AmazonS3/latest/dev/UploadingObjects.html

Domain

Design Resilient Architectures

Question 42Skipped

A company has migrated its application from a monolith architecture to a microservices based architecture. The development team has updated the Amazon Route 53 simple record to point "myapp.mydomain.com" from the old Load Balancer to the new one.

The users are still not redirected to the new Load Balancer. What has gone wrong in the configuration?

Correct answer

The Time To Live (TTL) is still in effect

The Alias Record is misconfigured

The health checks are failing

The CNAME Record is misconfigured

Overall explanation

Correct option:

Amazon Route 53 is a highly available and scalable cloud Domain Name System (DNS) web service. Amazon Route 53 effectively connects user requests to infrastructure running in AWS – such as Amazon EC2 instances, Elastic Load Balancing load balancers, or Amazon S3 buckets – and can also be used to route users to infrastructure outside of AWS.

You can use Amazon Route 53 to configure DNS health checks to route traffic to healthy endpoints or to independently monitor the health of your application and its endpoints. Amazon Route 53 Traffic Flow makes it easy for you to manage traffic globally through a variety of routing types, including Latency Based Routing, Geo DNS, Geoproximity, and Weighted Round Robin—all of which can be combined with DNS Failover to enable a variety of low-latency, fault-tolerant architectures.

The Time To Live (TTL) is still in effect

TTL (time to live), is the amount of time, in seconds, that you want DNS recursive resolvers to cache information about a record. If you specify a longer value (for example, 172800 seconds, or two days), you reduce the number of calls that DNS recursive resolvers must make to Amazon Route 53 to get the latest information for the record. This has the effect of reducing latency and reducing your bill for Route 53 service.

However, if you specify a longer value for TTL, it takes longer for changes to the record (for example, a new IP address) to take effect because recursive resolvers use the values in their cache for longer periods before they ask Route 53 for the latest information. If you're changing settings for a domain or subdomain that's already in use, AWS recommends that you initially specify a shorter value, such as 300 seconds, and increase the value after you confirm that the new settings are correct.

For this use-case, the most likely issue is that the TTL is still in effect so you have to wait until it expires for the new request to perform another DNS query and get the value for the new Load Balancer.

Incorrect options:

The CNAME Record is misconfigured - A CNAME record can redirect DNS queries to any DNS record. For example, you can create a CNAME record that redirects queries from acme.example.com to zenith.example.com or to acme.example.org. You don't need to use Amazon Route 53 as the DNS service for the domain that you're redirecting queries to.

The Alias Record is misconfigured - Amazon Route 53 also offers alias records, which are an Amazon Route 53-specific extension to DNS. Alias records let you route traffic to selected AWS resources, such as Amazon CloudFront distributions and Amazon S3 buckets. They also let you route traffic from one record in a hosted zone to another record. Unlike a CNAME record, you can create an alias record at the top node of a DNS namespace, also known as the zone apex. For example, if you register the DNS name example.com, the zone apex is example.com. You

can't create a CNAME record for example.com, but you can create an alias record for example.com that routes traffic to www.example.com.

The health checks are failing - Simple Records do not have health checks, so this option is incorrect.

References:

https://aws.amazon.com/route53/

https://docs.aws.amazon.com/Route53/latest/DeveloperGuide/resource-record-sets-values-basic.html

https://docs.aws.amazon.com/Route53/latest/DeveloperGuide/resource-record-sets-choosing-alias-non-alias.html

Domain

Design Resilient Architectures

Question 43Skipped

A photo hosting service publishes a collection of beautiful mountain images, every month, that aggregate over 50 gigabytes in size and downloaded all around the world. The content is currently hosted on Amazon EFS and distributed by Elastic Load Balancing (ELB) and Amazon EC2 instances. The website is experiencing high load each month and very high network costs.

As a Solutions Architect, what can you recommend that won't force an application refactor and reduce network costs and Amazon EC2 load drastically?

Upgrade the Amazon EC2 instances

Host the master pack onto Amazon S3 for faster access

Correct answer

Create an Amazon CloudFront distribution

Enable Elastic Load Balancing (ELB) caching

Overall explanation

Correct option:

Create an Amazon CloudFront distribution

Amazon CloudFront is a fast content delivery network (CDN) service that securely delivers data, videos, applications, and APIs to customers globally with low latency, high transfer speeds, all within a developer-friendly environment.

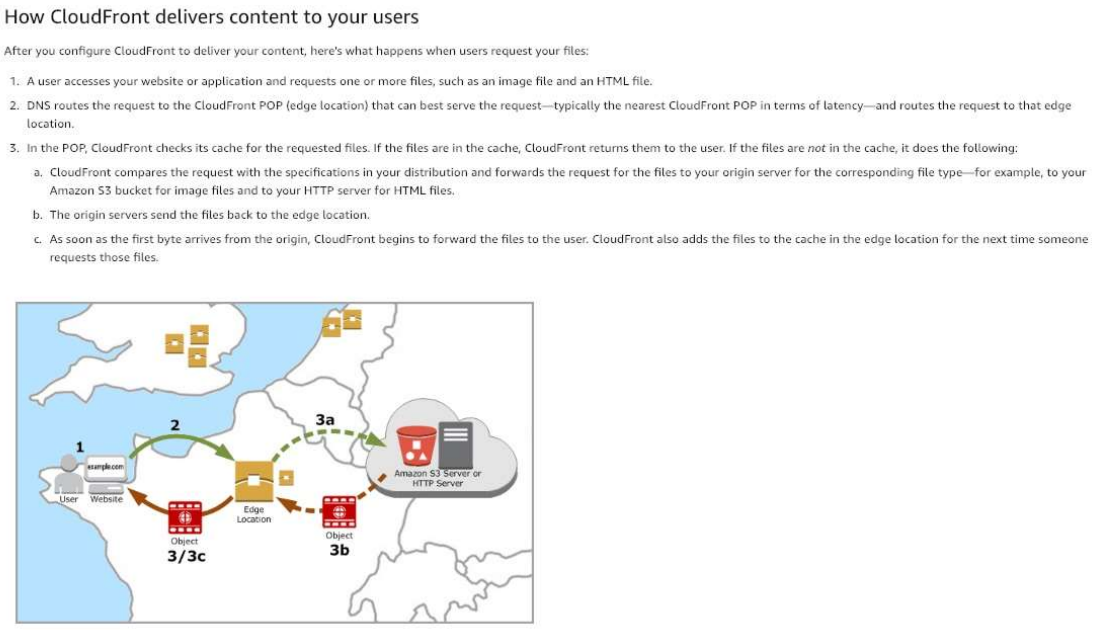
Amazon CloudFront points of presence (POPs) (edge locations) make sure that popular content can be served quickly to your viewers. Amazon CloudFront also has regional edge caches that bring more of your content closer to your viewers, even when the content is not popular enough to stay at a POP, to help improve performance for that content.

Regional edge caches help with all types of content, particularly content that tends to become less popular over time. Examples include user-generated content, such as video, photos, or

artwork; e-commerce assets such as product photos and videos; and news and event-related content that might suddenly find new popularity.

For the given use case, you need to create an Amazon CloudFront distribution to add a caching layer in front of your ELB. That caching layer will be very effective as the image pack is a static file, and therefore it would save the network costs significantly without requiring an application refactor.

How Amazon CloudFront delivers content to your users:



via
- https://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/HowCloudFrontWorks.html

Incorrect options:

Host the master pack onto Amazon S3 for faster access - Hosting the master pack into Amazon S3 will result in application code refactoring. So, this option is not correct.

Upgrade the Amazon EC2 instances - Upgrading the Amazon EC2 instances won't help save network cost. Hence, this option is incorrect for the given scenario.

Enable Elastic Load Balancing (ELB) caching - ELB does not have any caching capability. This option is just added as a distractor.

Reference:

https://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/HowCloudFrontWorks.html

Domain

Design Cost-Optimized Architectures

Question 44Skipped

A company has built a serverless application using Amazon API Gateway and AWS Lambda. The backend is leveraging an Amazon Aurora MySQL database. The web application was initially launched in the Americas and the company would now like to expand it to Europe, where a read-only version will be available to improve latency. You plan on deploying the Amazon API Gateway and AWS Lambda using AWS CloudFormation, but would like to have a read-only copy of your data in Europe as well.

As a Solutions Architect, what do you recommend?

Use Amazon Aurora Multi-AZ

Create an AWS Lambda function to periodically back up and restore the Amazon Aurora database in another region

Correct answer

Use Amazon Aurora Read Replicas

Use Amazon DynamoDB Streams

Overall explanation

Correct option:

Use Amazon Aurora Read Replicas

Amazon Aurora features a distributed, fault-tolerant, self-healing storage system that auto-scales up to 64TB per database instance. It delivers high performance and availability with up to 15 low-latency read replicas, point-in-time recovery, continuous backup to Amazon S3, and replication across three Availability Zones (AZs).

Aurora Replicas are independent endpoints in an Aurora DB cluster, best used for scaling read operations. Up to 15 Aurora Replicas can be distributed across the Availability Zones (AZs) that a DB cluster spans within an AWS Region. The DB cluster volume is made up of multiple copies of the data for the DB cluster. However, the data in the cluster volume is represented as a single, logical volume to the primary instance and Aurora Replicas in the DB cluster. You can also set up two Aurora MySQL DB clusters in different AWS Regions, by creating an Aurora Read Replica of an Amazon Aurora MySQL DB cluster in a different AWS Region. In this way, Aurora Read Replicas can be deployed globally.

Incorrect options:

Use Amazon Aurora Multi-AZ - Aurora, stores copies of the data in a DB cluster across multiple Availability Zones in a single AWS Region, regardless of whether the instances in the DB cluster span multiple Availability Zones (AZs). When data is written to the primary DB instance, Aurora synchronously replicates the data across Availability Zones (AZs) to six storage nodes associated with your cluster volume. Doing so provides data redundancy, eliminates I/O freezes, and minimizes latency spikes during system backups. Always remember that the main purpose for multi-AZ is high availability whereas the main purpose of read replicas is read scalability.

Use Amazon DynamoDB Streams - Amazon DynamoDB Streams is a powerful service that you can combine with other AWS services to solve many problems. When enabled, DynamoDB Streams captures a time-ordered sequence of item-level modifications in a DynamoDB table and durably stores the information for up to 24 hours. Applications can access a series of stream records, which contain an item change, from a DynamoDB stream in near real-time.

Create an AWS Lambda function to periodically back up and restore the Amazon Aurora database in another region - AWS Lambda can be used to create backups for Amazon RDS. But, the process is not an optimized solution, especially when Amazon Aurora already offers the read replica feature.

References:

https://docs.aws.amazon.com/AmazonRDS/latest/AuroraUserGuide/Aurora.Replication.html

https://aws.amazon.com/rds/features/multi-az/

https://aws.amazon.com/rds/features/read-replicas/

Domain

Design Resilient Architectures

Question 45Skipped

You are working as an AWS architect for a weather tracking facility. You are asked to set up a Disaster Recovery (DR) mechanism with minimum costs. In case of failure, the facility can only bear data loss of approximately 15 minutes without jeopardizing the forecasting models.

As a Solutions Architect, which DR method will you suggest?

Correct answer

Pilot Light

Multi-Site

Backup and Restore

Warm Standby

Overall explanation

Correct option:

Pilot Light

The term pilot light is often used to describe a DR scenario in which a minimal version of an environment is always running in the cloud. The idea of the pilot light is an analogy that comes from the gas heater. In a gas heater, a small flame that's always on can quickly ignite the entire furnace to heat up a house. This scenario is similar to a backup-and-restore scenario. For example, with AWS you can maintain a pilot light by configuring and running the most critical core elements of your system in AWS. For the given use-case, a small part of the backup infrastructure is always running simultaneously syncing mutable data (such as databases or documents) so that there is no loss of critical data. When the time comes for recovery, you can

rapidly provision a full-scale production environment around the critical core. For Pilot light, RPO/RTO is in 10s of minutes, so this is the correct solution.

Four Disaster Recovery scenarios:

- **Backup and restore** (RPO in hours, RTO in 24 hours or less): Back up your data and applications using point-in-time backups into the DR Region. Restore this data when necessary to recover from a disaster.
- **Pilot light** (RPO in minutes, RTO in hours): Replicate your data from one region to another and provision a copy of your core workload infrastructure. Resources required to support data replication and backup such as databases and object storage are always on. Other elements such as application servers are loaded with application code and configurations, but are switched off and are only used during testing or when Disaster Recovery failover is invoked.
- **Warm standby** (RPO in seconds, RTO in minutes): Maintain a scaled-down but fully functional version of your workload always running in the DR Region. Business-critical systems are fully duplicated and are always on, but with a scaled down fleet. When the time comes for recovery, the system is scaled up quickly to handle the production load. The more scaled-up the Warm Standby is, the lower RTO and control plane reliance will be. When scaled up to full scale this is known as a **Hot Standby**.
- **Multi-region (multi-site) active-active** (RPO near zero, RTO potentially zero): Your workload is deployed to, and actively serving traffic from, multiple AWS Regions. This strategy requires you to synchronize data across Regions. Possible conflicts caused by writes to the same record in two different regional replicas must be avoided or handled. Data replication is useful for data synchronization and will protect you against some types of disaster, but it will not protect you against data corruption or destruction unless your solution also includes options for point-in-time recovery. Use services like Amazon Route 53 or AWS Global Accelerator to route your user traffic to where your workload is healthy. For more details on AWS services you can use for active-active architectures see the **AWS Regions** section of Use Fault Isolation to Protect Your Workload.

via - https://docs.aws.amazon.com/wellarchitected/latest/reliability-pillar/plan-for-disaster-recovery-dr.html

Incorrect options:

Backup and Restore - In most traditional environments, data is backed up to tape and sent off-site regularly. If you use this method, it can take a long time to restore your system in the event of a disruption or disaster. Amazon S3 is an ideal destination for backup data that might be needed quickly to perform a restore. Transferring data to and from Amazon S3 is typically done through the network and is therefore accessible from any location. There are many commercial and open-source backup solutions that integrate with Amazon S3. You can use AWS Import/Export to transfer very large data sets by shipping storage devices directly to AWS. For longer-term data storage where retrieval times of several hours are adequate, there is Amazon Glacier, which has the same durability model as Amazon S3. Amazon Glacier is a low-cost alternative starting from $0.01/GB per month. Amazon Glacier and Amazon S3 can be used in conjunction to produce a tiered backup solution. Even though Backup and Restore method is cheaper, it has an RPO in hours, so this option is not the right fit.

Warm Standby - The term warm standby is used to describe a DR scenario in which a scaled-down version of a fully functional environment is always running in the cloud. A warm standby solution extends the pilot light elements and preparation. It further decreases the recovery time because some services are always running. By identifying your business-critical systems, you can fully duplicate these systems on AWS and have them always on. This option is costlier compared to Pilot Light.

Multi-Site - A multi-site solution runs on AWS as well as on your existing on-site infrastructure in an active-active configuration. The data replication method that you employ will be determined by the recovery point that you choose, either Recovery Time Objective (the

maximum allowable downtime before degraded operations are restored) or Recovery Point Objective (the maximum allowable time window whereby you will accept the loss of transactions during the DR process). This option is more costly compared to Pilot Light.

References:

https://aws.amazon.com/blogs/architecture/disaster-recovery-dr-architecture-on-aws-part-iii-pilot-light-and-warm-standby/

https://aws.amazon.com/blogs/publicsector/rapidly-recover-mission-critical-systems-in-a-disaster/

https://docs.aws.amazon.com/wellarchitected/latest/reliability-pillar/plan-for-disaster-recovery-dr.html

https://docs.aws.amazon.com/wellarchitected/latest/reliability-pillar/disaster-recovery-dr-objectives.html

Domain

Design Resilient Architectures

Question 46Skipped

As part of the on-premises data center migration to AWS Cloud, a company is looking at using multiple AWS Snow Family devices to move their on-premises data.

Which AWS Snow Family service offers the feature of storage clustering?

AWS Snowmobile Storage Compute

AWS Snowcone

Correct answer

AWS Snowball Edge Compute Optimized

AWS Snowmobile

Overall explanation

Correct option:

AWS Snowball Edge Compute Optimized

AWS Snowball is a data migration and edge computing device that comes in two device options: Compute Optimized and Storage Optimized. AWS Snowball Edge Storage Optimized devices provide 40 vCPUs of compute capacity coupled with 80 terabytes of usable block or Amazon S3-compatible object storage. It is well-suited for local storage and large-scale data transfer. AWS Snowball Edge Compute Optimized devices provide 52 vCPUs, 42 terabytes of usable block or object storage, and an optional GPU for use cases such as advanced machine learning and full-motion video analysis in disconnected environments.

Customers can use these two options for data collection, machine learning and processing, and storage in environments with intermittent connectivity (such as manufacturing, industrial, and transportation) or in extremely remote locations (such as military or maritime operations)

before shipping it back to AWS. These devices may also be rack mounted and clustered together to build larger, temporary installations.

Therefore, both AWS Snowball Edge Storage Optimized and AWS Snowball Edge Compute Optimized offer the storage clustering feature.

Feature comparison

	AWS Snowcone	AWS Snowball Edge Storage Optimized	AWS Snowball Edge Compute Optimized	AWS Snowmobile
Usage Scenario	Edge computing, Data transfer, Edge storage	Data transfer, Edge storage	Edge computing, Data transfer	Data transfer
Usable HDD Storage	8 TB	80 TB	42 TB	100 PB
Usable SSD Storage	No	1 TB	7.68 TB	No
Usable vCPUs	2 vCPUs	40 vCPUs	52 vCPUs	N/A
Usable Memory	4 GB	80 GB	208 GB	N/A
GPU	No	No	nVidia V100 (optional)	No
Onboard Computing Options	AWS IoT Greengrass Amazon EC2 AMIs	AWS IoT Greengrass Amazon EC2 AMIs	AWS IoT Greengrass Amazon EC2 AMIs	N/A
DataSync	Yes	No	No	No
Transfers via NFS	Yes	Yes	Yes	Yes
Transfers via S3 API	No	Yes	Yes	No
Network Interfaces	2x 1/10 Gbit - RJ45	2x 10 Gbit – RJ45 1x 25 Gbit – SFP+ 1x 100 Gbit – QSFP28	2x 10 Gbit – RJ45 1x 25 Gbit – SFP+ 1x 100 Gbit – QSFP28	6x 40 Gbit
Device Size	9 inches long, 6 inches wide, and 3 inches tall (227 mm x 148.6 mm x 82.65 mm)	28.3 inches long, 10.6 inches wide, and 15.5 inches tall (548 mm x 320 mm x 501 mm)	28.3 inches long, 10.6 inches wide, and 15.5 inches tall (548 mm x 320 mm x 501 mm)	N/A
Device Weight	4.5 lbs. (2.1 kg)	49.7 lbs. (22.3 kg)	49.7 lbs. (22.3 kg)	N/A
Encryption	Yes, 256-bit	Yes, 256-bit	Yes, 256-bit	Yes, 256-bit
Portability	Battery-based Operation	No	No	No
Wireless	Wi-Fi	No	No	No
Storage Clustering	No	Yes, 5-10 nodes	Yes, 5-10 nodes	N/A
HIPAA Compliant	Yes, eligible	Yes, eligible	Yes, eligible	No
Typical Job Lifetime	Offline or Online Data Transfer: Days-Weeks Edge Compute: Weeks-Years	Offline Data Transfer: Days-Weeks	Edge Compute: Weeks-Years	Data Migration: Months

via - https://aws.amazon.com/snow/#Feature_comparison

Incorrect options:

AWS Snowcone - AWS Snowcone is the smallest member of the AWS Snow Family of edge computing and data transfer devices. Snowcone is portable, rugged, and secure. You can use Snowcone to collect, process, and move data to AWS, either offline by shipping the device or online with AWS DataSync. Snowcone does not offer a storage Clustering option.

AWS Snowmobile - AWS Snowmobile moves up to 100 PB of data in a 45-foot long ruggedized shipping container and is ideal for multi-petabyte or Exabyte-scale digital media migrations and data centers shutdowns. AWS Snowmobile arrives at the customer site and appears as a network-attached data store for more secure, high-speed data transfer. After data is transferred to Snowmobile, it is driven back to an AWS Region where the data is loaded into Amazon S3. AWS Snowmobile does not offer a storage Clustering option.

AWS Snowmobile Storage Compute - This is a made-up option, given only as a distractor.

References:

https://aws.amazon.com/snow/#Feature_comparison

https://aws.amazon.com/snow/

Domain

Design Resilient Architectures

Question 47Skipped

Your company is deploying a website running on AWS Elastic Beanstalk. The website takes over 45 minutes for the installation and contains both static as well as dynamic files that must be generated during the installation process.

As a Solutions Architect, you would like to bring the time to create a new instance in your AWS Elastic Beanstalk deployment to be less than 2 minutes. Which of the following options should be combined to build a solution for this requirement? (Select two)

Correct selection

Create a Golden Amazon Machine Image (AMI) with the static installation components already setup

Use Amazon EC2 user data to install the application at boot time

Store the installation files in Amazon S3 so they can be quickly retrieved

Correct selection

Use Amazon EC2 user data to customize the dynamic installation parts at boot time

Use AWS Elastic Beanstalk deployment caching feature

Overall explanation

Correct options:

AWS Elastic Beanstalk is an easy-to-use service for deploying and scaling web applications and services developed with Java, .NET, PHP, Node.js, Python, Ruby, Go, and Docker on familiar servers such as Apache, Nginx, Passenger, and IIS.

You can simply upload your code and Elastic Beanstalk automatically handles the deployment, from capacity provisioning, load balancing, auto-scaling to application health monitoring. At the same time, you retain full control over the AWS resources powering your application and can access the underlying resources at any time.

When you create an AWS Elastic Beanstalk environment, you can specify an Amazon Machine Image (AMI) to use instead of the standard Elastic Beanstalk AMI included in your platform version. A custom AMI can improve provisioning times when instances are launched in your environment if you need to install a lot of software that isn't included in the standard AMIs.

Create a Golden Amazon Machine Image (AMI) with the static installation components already setup

A Golden AMI is an AMI that you standardize through configuration, consistent security patching, and hardening. It also contains agents you approve for logging, security, performance monitoring, etc. For the given use-case, you can have the static installation components already setup via the golden AMI.

Use Amazon EC2 user data to customize the dynamic installation parts at boot time

Amazon EC2 instance user data is the data that you specified in the form of a configuration script while launching your instance. You can use Amazon EC2 user data to customize the dynamic installation parts at boot time, rather than installing the application itself at boot time.

Incorrect options:

Store the installation files in Amazon S3 so they can be quickly retrieved - Amazon S3 bucket can be used as a storage location for your source code, logs, and other artifacts that are created when you use AWS Elastic Beanstalk. It cannot be used to run or generate dynamic files since Amazon S3 is not an environment but a storage service.

Use Amazon EC2 user data to install the application at boot time - User data of an instance can be used to perform common automated configuration tasks or run scripts after the instance starts. User data, cannot, however, be used to install the application since it takes over 45 minutes for the installation which contains static as well as dynamic files that must be generated during the installation process.

Use AWS Elastic Beanstalk deployment caching feature - AWS Elastic Beanstalk deployment caching is a made-up option. It is just added as a distractor.

References:

https://aws.amazon.com/elasticbeanstalk/

https://aws.amazon.com/blogs/awsmarketplace/announcing-the-golden-ami-pipeline/

https://docs.aws.amazon.com/elasticbeanstalk/latest/dg/AWSHowTo.S3.html

https://docs.aws.amazon.com/elasticbeanstalk/latest/dg/using-features.customenv.html

https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/instancedata-add-user-data.html

Domain

Design Resilient Architectures

Question 48Skipped

A startup's cloud infrastructure consists of a few Amazon EC2 instances, Amazon RDS instances and Amazon S3 storage. A year into their business operations, the startup is incurring costs that seem too high for their business requirements.

Which of the following options represents a valid cost-optimization solution?

Use Amazon S3 Storage class analysis to get recommendations for transitions of objects to Amazon S3 Glacier storage classes to reduce storage costs. You can also automate moving these objects into lower-cost storage tier using Lifecycle Policies

Use AWS Trusted Advisor checks on Amazon EC2 Reserved Instances to automatically renew reserved instances (RI). AWS Trusted advisor also suggests Amazon RDS idle database instances

Correct answer

Use AWS Cost Explorer Resource Optimization to get a report of Amazon EC2 instances that are either idle or have low utilization and use AWS Compute Optimizer to look at instance type recommendations

Use AWS Compute Optimizer recommendations to help you choose the optimal Amazon EC2 purchasing options and help reserve your instance capacities at reduced costs

Overall explanation

Correct option:

Use AWS Cost Explorer Resource Optimization to get a report of Amazon EC2 instances that are either idle or have low utilization and use AWS Compute Optimizer to look at instance type recommendations

AWS Cost Explorer helps you identify under-utilized Amazon EC2 instances that may be downsized on an instance by instance basis within the same instance family, and also understand the potential impact on your AWS bill by taking into account your Reserved Instances and Savings Plans.

AWS Compute Optimizer recommends optimal AWS Compute resources for your workloads to reduce costs and improve performance by using machine learning to analyze historical utilization metrics. Compute Optimizer helps you choose the optimal Amazon EC2 instance types, including those that are part of an Amazon EC2 Auto Scaling group, based on your utilization data.

Incorrect options:

Use Amazon S3 Storage class analysis to get recommendations for transitions of objects to Amazon S3 Glacier storage classes to reduce storage costs. You can also automate moving these objects into lower-cost storage tier using Lifecycle Policies - By using Amazon S3 Analytics Storage Class analysis you can analyze storage access patterns to help you decide when to transition the right data to the right storage class. This new Amazon S3 analytics feature observes data access patterns to help you determine when to transition less frequently accessed STANDARD storage to the STANDARD_IA (IA, for infrequent access) storage class. Storage class analysis does not give recommendations for transitions to the ONEZONE_IA or S3 Glacier storage classes.

Use AWS Trusted Advisor checks on Amazon EC2 Reserved Instances to automatically renew reserved instances (RI). AWS Trusted advisor also suggests Amazon RDS idle database instances - AWS Trusted Advisor checks for Amazon EC2 Reserved Instances that are scheduled to expire within the next 30 days or have expired in the preceding 30 days. Reserved Instances do not renew automatically; you can continue using an Amazon EC2 instance covered by the reservation without interruption, but you will be charged On-Demand rates. AWS Trusted advisor does not have a feature to auto-renew Reserved Instances.

Use AWS Compute Optimizer recommendations to help you choose the optimal Amazon EC2 purchasing options and help reserve your instance capacities at reduced costs - AWS Compute Optimizer recommends optimal AWS Compute resources for your workloads to reduce costs and improve performance by using machine learning to analyze historical utilization metrics. Over-provisioning compute can lead to unnecessary infrastructure cost and under-provisioning compute can lead to poor application performance. Compute Optimizer helps you choose the optimal Amazon EC2 instance types, including those that are part of an Amazon EC2 Auto Scaling group, based on your utilization data. It does not recommend instance purchase options.

References:

https://aws.amazon.com/compute-optimizer/

https://aws.amazon.com/premiumsupport/technology/trusted-advisor/best-practice-checklist/

https://docs.aws.amazon.com/AmazonS3/latest/dev/analytics-storage-class.html

Domain

Design Cost-Optimized Architectures

Question 49Skipped

You have developed a new REST API leveraging the Amazon API Gateway, AWS Lambda and Amazon Aurora database services. Most of the workload on the website is read-heavy. The data rarely changes and it is acceptable to serve users outdated data for about 24 hours. Recently, the website has been experiencing high load and the costs incurred on the Aurora database have been very high.

How can you easily reduce the costs while improving performance, with minimal changes?

Correct answer

Enable Amazon API Gateway Caching

Enable AWS Lambda In Memory Caching

Switch to using an Application Load Balancer

Add Amazon Aurora Read Replicas

Overall explanation

Correct option:

Enable Amazon API Gateway Caching

Amazon API Gateway is a fully managed service that makes it easy for developers to create, publish, maintain, monitor, and secure APIs at any scale. APIs act as the "front door" for applications to access data, business logic, or functionality from your backend services. Using API Gateway, you can create RESTful APIs and WebSocket APIs that enable real-time two-way communication applications. API Gateway supports containerized and serverless workloads, as well as web applications.

You can enable Amazon API caching in Amazon API Gateway to cache your endpoint's responses. With caching, you can reduce the number of calls made to your endpoint and also improve the latency of requests to your API. When you enable caching for a stage, API Gateway caches responses from your endpoint for a specified time-to-live (TTL) period, in seconds. Amazon API Gateway then responds to the request by looking up the endpoint response from the cache instead of requesting your endpoint. The default TTL value for API caching is 300 seconds. The maximum TTL value is 3600 seconds. TTL=0 means caching is disabled. Using API Gateway Caching feature is the answer for the use case, as we can accept stale data for about 24 hours.

Incorrect options:

Add Amazon Aurora Read Replicas - Amazon Aurora features a distributed, fault-tolerant, self-healing storage system that auto-scales up to 64TB per database instance. It delivers high performance and availability with up to 15 low-latency read replicas, point-in-time recovery, continuous backup to Amazon S3, and replication across three Availability Zones (AZs).

Amazon Aurora Read Replicas are independent endpoints in an Aurora DB cluster, best used for scaling read operations and increasing availability. Up to 15 Aurora Replicas can be distributed across the Availability Zones that a DB cluster spans within an AWS Region. The DB cluster volume is made up of multiple copies of the data for the DB cluster. However, the data in the cluster volume is represented as a single, logical volume to the primary instance and to Aurora Replicas in the DB cluster. Adding Aurora Read Replicas would greatly increase the cost, therefore this option is ruled out.

Switch to using an Application Load Balancer - An Application Load Balancer functions at the application layer, the seventh layer of the Open Systems Interconnection (OSI) model. After the load balancer receives a request, it evaluates the listener rules in priority order to determine which rule to apply, and then selects a target from the target group for the rule action. You can configure listener rules to route requests to different target groups based on the content of the application traffic. Switching to a Load Balancer would not improve the current status as we need a caching mechanism.

Enable AWS Lambda In Memory Caching - AWS Lambda has no native in-memory caching capability. AWS Lambda is a serverless compute capacity. This option is incorrect and has been added as a distractor.

Reference:

https://docs.aws.amazon.com/apigateway/latest/developerguide/api-gateway-caching.html

Domain

Design Cost-Optimized Architectures

Question 50Skipped

A social media company wants the capability to dynamically alter the size of a geographic area from which traffic is routed to a specific server resource.

Which feature of Amazon Route 53 can help achieve this functionality?

Weighted routing

Correct answer

Geoproximity routing

Latency-based routing

Geolocation routing

Overall explanation

Correct option:

Geoproximity routing

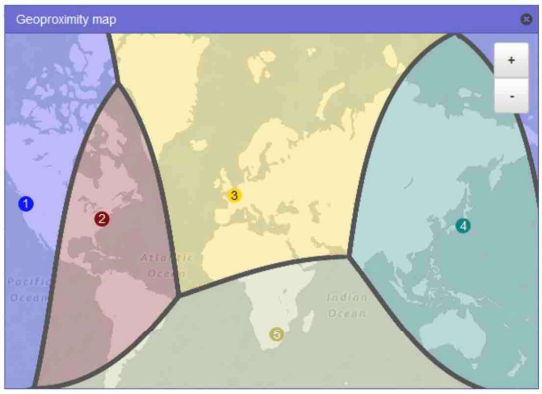
Geoproximity routing lets Amazon Route 53 route traffic to your resources based on the geographic location of your users and your resources. You can also optionally choose to route more traffic or less to a given resource by specifying a value, known as a bias. A bias expands or shrinks the size of the geographic region from which traffic is routed to a resource.

To optionally change the size of the geographic region from which Amazon Route 53 routes traffic to a resource, specify the applicable value for the bias: 1. To expand the size of the geographic region from which Amazon Route 53 routes traffic to a resource, specify a positive integer from 1 to 99 for the bias. Amazon Route 53 shrinks the size of adjacent regions.

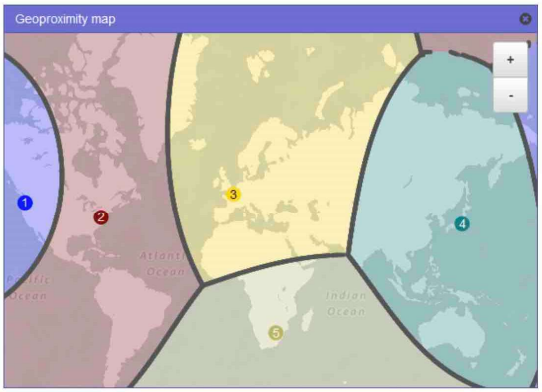
1. To shrink the size of the geographic region from which Amazon Route 53 routes traffic to a resource, specify a negative bias of -1 to -99. Amazon Route 53 expands the size of adjacent regions.

More on how bias works in Geoproximity routing:

The following map shows four AWS Regions (numbered 1 through 4) and a location in Johannesburg, South Africa that is specified by latitude and longitude (5).



The following map shows what happens if you add a bias of +25 for the US East (Northern Virginia) Region (number **2** on the map). Traffic is routed to the resource in that Region from a larger portion of North America than previously, and from all of South America.



via - https://docs.aws.amazon.com/Route53/latest/DeveloperGuide/routing-policy.html

Incorrect options:

Geolocation routing - Geolocation routing lets you choose the resources that serve your traffic based on the geographic location of your users, meaning the location that DNS queries originate from. For example, you might want all queries from Europe to be routed to an Elastic Load Balancing (ELB) load balancer in the Frankfurt region.

When you use geolocation routing, you can localize your content and present some or all of your website in the language of your users. You can also use geolocation routing to restrict the distribution of content to only the locations in which you have distribution rights. Another possible use is for balancing load across endpoints in a predictable, easy-to-manage way so that each user location is consistently routed to the same endpoint.

Latency-based routing - If your application is hosted in multiple AWS Regions, you can improve performance for your users by serving their requests from the AWS Region that provides the lowest latency.

To use latency-based routing, you create latency records for your resources in multiple AWS Regions. When Amazon Route 53 receives a DNS query for your domain or subdomain (example.com or acme.example.com), it determines which AWS Regions you've created latency

records for, determines which region gives the user the lowest latency, and then selects a latency record for that region. Amazon Route 53 responds with the value from the selected record, such as the IP address for a web server.

Weighted routing - Weighted routing lets you associate multiple resources with a single domain name (example.com) or subdomain name (acme.example.com) and choose how much traffic is routed to each resource. This can be useful for a variety of purposes, including load balancing and testing new versions of software.

To configure weighted routing, you create records that have the same name and type for each of your resources. You assign each record a relative weight that corresponds with how much traffic you want to send to each resource. Amazon Route 53 sends traffic to a resource based on the weight that you assign to the record as a proportion of the total weight for all records in the group.

Reference:

https://docs.aws.amazon.com/Route53/latest/DeveloperGuide/routing-policy.html

Domain

Design High-Performing Architectures

Question 51Skipped

A company's business logic is built on several microservices that are running in the on-premises data center. They currently communicate using a message broker that supports the MQTT protocol. The company is looking at migrating these applications and the message broker to AWS Cloud without changing the application logic.

Which technology allows you to get a managed message broker that supports the MQTT protocol?

Amazon Simple Notification Service (Amazon SNS)

Correct answer

Amazon MQ

Amazon Simple Queue Service (Amazon SQS)

Amazon Kinesis Data Streams

Overall explanation

Correct option:

Amazon MQ

Amazon MQ is a managed message broker service for Apache ActiveMQ that makes it easy to set up and operate message brokers in the cloud. Message brokers allow different software systems–often using different programming languages, and on different platforms–to communicate and exchange information. If an organization is using messaging with existing applications and wants to move the messaging service to the cloud quickly and easily, AWS recommends Amazon MQ for such a use case. Connecting your current applications to Amazon

MQ is easy because it uses industry-standard APIs and protocols for messaging, including JMS, NMS, AMQP, STOMP, MQTT, and WebSocket.

Therefore, the only possible answer is Amazon MQ.

Incorrect option:

Amazon Simple Queue Service (Amazon SQS) - Amazon Simple Queue Service (Amazon SQS) is a fully managed message queuing service that enables you to decouple and scale microservices, distributed systems, and serverless applications. SQS eliminates the complexity and overhead associated with managing and operating message-oriented middleware and empowers developers to focus on differentiating work. Using Amazon SQS, you can send, store, and receive messages between software components at any volume, without losing messages or requiring other services to be available.

Amazon Kinesis Data Streams - Amazon Kinesis Data Streams (KDS) is a massively scalable and durable real-time data streaming service. KDS can continuously capture gigabytes of data per second from hundreds of thousands of sources such as website clickstreams, database event streams, financial transactions, social media feeds, IT logs, and location-tracking events.

Amazon Simple Notification Service (Amazon SNS) - Amazon Simple Notification Service (Amazon SNS) is a highly available, durable, secure, fully managed pub/sub messaging service that enables you to decouple microservices, distributed systems, and serverless applications. Amazon SNS provides topics for high-throughput, push-based, many-to-many messaging.

Amazon SNS, Amazon SQS, and Amazon Kinesis are AWS's proprietary technologies and do not come with MQTT compatibility.

References:

https://aws.amazon.com/amazon-mq/

Domain

Design High-Performing Architectures

Question 52Skipped

A junior developer has downloaded a sample Amazon S3 bucket policy to make changes to it based on new company-wide access policies. He has requested your help in understanding this bucket policy.

As a Solutions Architect, which of the following would you identify as the correct description for the given policy?

```
{
 "Version": "2012-10-17",
 "Id": "S3PolicyId1",
 "Statement": [
  {
   "Sid": "IPAllow",
```

```
    "Effect": "Allow",
    "Principal": "*",
    "Action": "s3:*",
    "Resource": "arn:aws:s3:::examplebucket/*",
    "Condition": {
      "IpAddress": {"aws:SourceIp": "54.240.143.0/24"},
      "NotIpAddress": {"aws:SourceIp": "54.240.143.188/32"}
    }
  }
 ]
}
```

It ensures Amazon EC2 instances that have inherited a security group can access the bucket

It authorizes an IP address and a Classless Inter-Domain Routing (CIDR) to access the S3 bucket

It ensures the Amazon S3 bucket is exposing an external IP within the Classless Inter-Domain Routing (CIDR) range specified, except one IP

Correct answer

It authorizes an entire Classless Inter-Domain Routing (CIDR) except one IP address to access the Amazon S3 bucket

Overall explanation

Correct option:

It authorizes an entire Classless Inter-Domain Routing (CIDR) except one IP address to access the Amazon S3 bucket

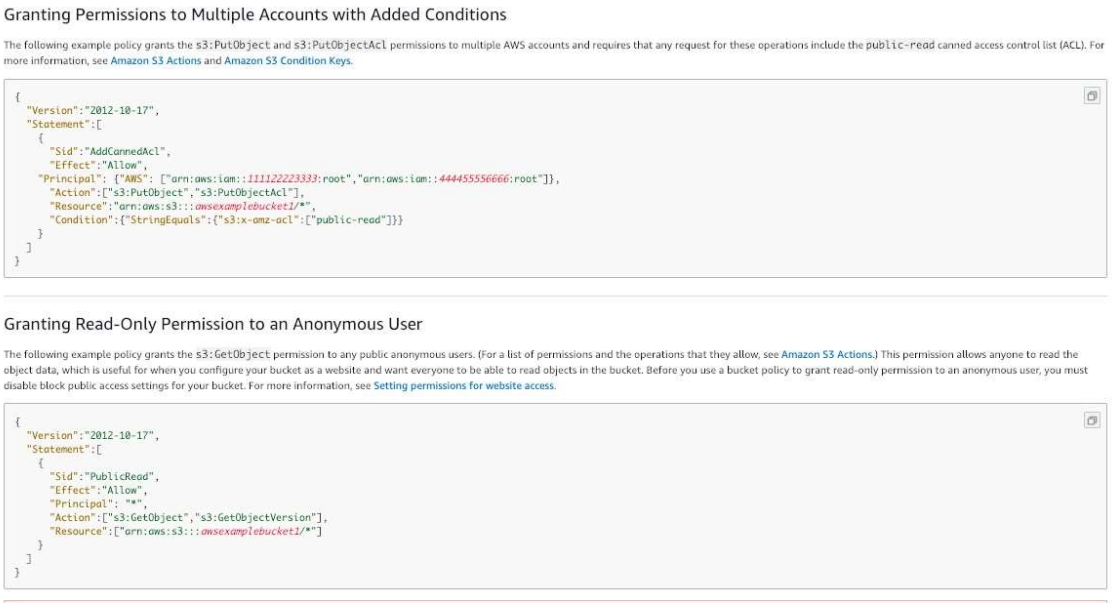
You manage access in AWS by creating policies and attaching them to IAM identities (users, groups of users, or roles) or AWS resources. A policy is an object in AWS that, when associated with an identity or resource, defines their permissions. AWS evaluates these policies when an IAM principal (user or role) makes a request. Permissions in the policies determine whether the request is allowed or denied. Most policies are stored in AWS as JSON documents. AWS supports six types of policies: identity-based policies, resource-based policies, permissions boundaries, AWS Organizations SCPs, ACLs, and session policies.

Let's analyze the bucket policy one step at a time:

The snippet "Effect": "Allow" implies an allow effect. The snippet "Principal": "*" implies any Principal. The snippet "Action": "s3:*" implies any Amazon S3 API. The snippet "Resource": "arn:aws:s3:::examplebucket/*" implies that the resource can be the bucket examplebucket and its contents. Consider the last snippet of the given bucket

policy: "Condition": { "IpAddress": {"aws:SourceIp": "54.240.143.0/24"}, "NotIpAddress": {"aws:SourceIp": "54.240.143.188/32"} } This snippet implies that if the source IP is in the CIDR block "54.240.143.0/24" (== 54.240.143.0 - 54.240.143.255), then it is allowed to access the examplebucket and its contents. However, the source IP cannot be in the CIDR "54.240.143.188/32" (== 1 IP, 54.240.143.188/32), which means one IP address is explicitly blocked from accessing the examplebucket and its contents.

Example Bucket policies:



via - https://docs.aws.amazon.com/AmazonS3/latest/dev/example-bucket-policies.html

Incorrect options:

It ensures the Amazon S3 bucket is exposing an external IP within the Classless Inter-Domain Routing (CIDR) range specified, except one IP

It ensures Amazon EC2 instances that have inherited a security group can access the bucket

It authorizes an IP address and a Classless Inter-Domain Routing (CIDR) to access the S3 bucket

These three options contradict the explanation provided above, so these options are incorrect.

Reference:

https://docs.aws.amazon.com/AmazonS3/latest/dev/example-bucket-policies.html

Domain

Design Secure Architectures

Question 53Skipped

The engineering team at a company is running batch workloads on AWS Cloud. The team has embedded Amazon RDS database connection strings within each web server hosting the

flagship application. After failing a security audit, the team is looking at a different approach to store the database secrets securely and automatically rotate the database credentials.

Which of the following solutions would you recommend to meet this requirement?

AWS Key Management Service (KMS)

AWS Systems Manager Parameter Store

AWS Systems Manager

Correct answer

AWS Secrets Manager

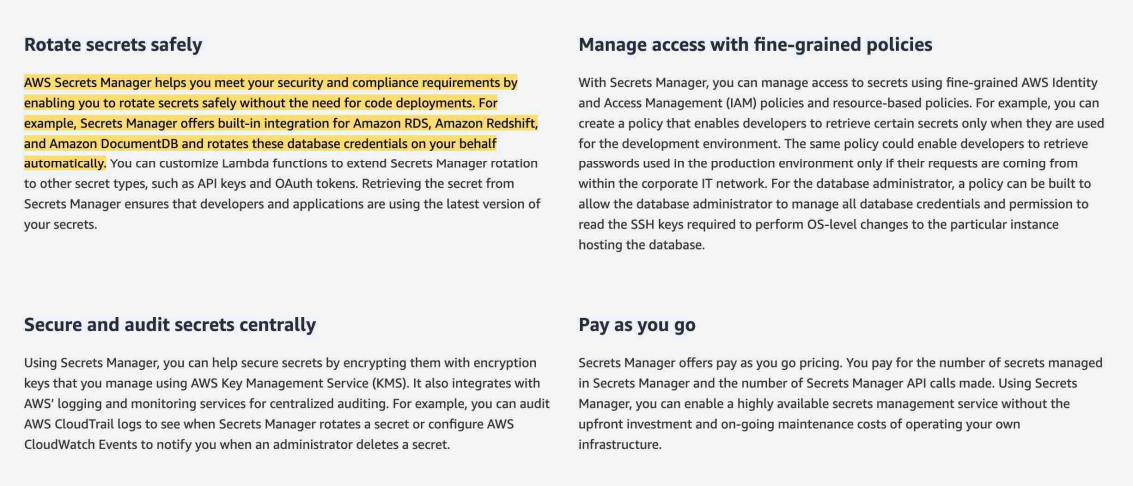
Overall explanation

Correct option:

AWS Secrets Manager

AWS Secrets Manager enables you to easily rotate, manage, and retrieve database credentials, API keys, and other secrets throughout their lifecycle. Users and applications retrieve secrets with a call to Secrets Manager APIs, eliminating the need to hardcode sensitive information in plain text. Secrets Manager offers secret rotation with built-in integration for Amazon RDS, Amazon Redshift, and Amazon DocumentDB.

Benefits of AWS Secrets Manager:



Rotate secrets safely

AWS Secrets Manager helps you meet your security and compliance requirements by enabling you to rotate secrets safely without the need for code deployments. For example, Secrets Manager offers built-in integration for Amazon RDS, Amazon Redshift, and Amazon DocumentDB and rotates these database credentials on your behalf automatically. You can customize Lambda functions to extend Secrets Manager rotation to other secret types, such as API keys and OAuth tokens. Retrieving the secret from Secrets Manager ensures that developers and applications are using the latest version of your secrets.

Manage access with fine-grained policies

With Secrets Manager, you can manage access to secrets using fine-grained AWS Identity and Access Management (IAM) policies and resource-based policies. For example, you can create a policy that enables developers to retrieve certain secrets only when they are used for the development environment. The same policy could enable developers to retrieve passwords used in the production environment only if their requests are coming from within the corporate IT network. For the database administrator, a policy can be built to allow the database administrator to manage all database credentials and permission to read the SSH keys required to perform OS-level changes to the particular instance hosting the database.

Secure and audit secrets centrally

Using Secrets Manager, you can help secure secrets by encrypting them with encryption keys that you manage using AWS Key Management Service (KMS). It also integrates with AWS' logging and monitoring services for centralized auditing. For example, you can audit AWS CloudTrail logs to see when Secrets Manager rotates a secret or configure AWS CloudWatch Events to notify you when an administrator deletes a secret.

Pay as you go

Secrets Manager offers pay as you go pricing. You pay for the number of secrets managed in Secrets Manager and the number of Secrets Manager API calls made. Using Secrets Manager, you can enable a highly available secrets management service without the upfront investment and on-going maintenance costs of operating your own infrastructure.

via - https://aws.amazon.com/secrets-manager/

Incorrect options:

AWS Systems Manager Parameter Store - AWS Systems Manager Parameter Store provides secure, hierarchical storage for configuration data management and secrets management. You can store data such as passwords, database strings, and license codes as parameter values. AWS SSM Parameter Store cannot be used to automatically rotate the database credentials.

AWS Systems Manager - AWS Systems Manager gives you visibility and control of your infrastructure on AWS. Systems Manager provides a unified user interface so you can view

operational data from multiple AWS services and allows you to automate operational tasks across your AWS resources. AWS Systems Manager cannot be used to store your secrets securely and automatically rotate the database credentials.

AWS Key Management Service (KMS) - AWS Key Management Service (KMS) makes it easy for you to create and manage cryptographic keys and control their use across a wide range of AWS services and in your applications. KMS cannot be used to store your secrets securely and automatically rotate the database credentials.

References:

https://aws.amazon.com/secrets-manager/

https://docs.aws.amazon.com/systems-manager/latest/userguide/systems-manager-paramstore.html

Domain

Design Secure Architectures

Question 54Skipped

An Elastic Load Balancer has marked all the Amazon EC2 instances in the target group as unhealthy. Surprisingly, when a developer enters the IP address of the Amazon EC2 instances in the web browser, he can access the website.

What could be the reason the instances are being marked as unhealthy? (Select two)

Correct selection

The route for the health check is misconfigured

You need to attach elastic IP address (EIP) to the Amazon EC2 instances

Your web-app has a runtime that is not supported by the Application Load Balancer

Correct selection

The security group of the Amazon EC2 instance does not allow for traffic from the security group of the Application Load Balancer

The Amazon Elastic Block Store (Amazon EBS) volumes have been improperly mounted

Overall explanation

Correct options:

The security group of the Amazon EC2 instance does not allow for traffic from the security group of the Application Load Balancer

The route for the health check is misconfigured

An Application Load Balancer periodically sends requests to its registered targets to test their status. These tests are called health checks.

Each load balancer node routes requests only to the healthy targets in the enabled Availability Zones (AZs) for the load balancer. Each load balancer node checks the health of each target, using the health check settings for the target groups with which the target is registered. If a

target group contains only unhealthy registered targets, the load balancer nodes route requests across its unhealthy targets.

You must ensure that your load balancer can communicate with registered targets on both the listener port and the health check port. Whenever you add a listener to your load balancer or update the health check port for a target group used by the load balancer to route requests, you must verify that the security groups associated with the load balancer allow traffic on the new port in both directions.

Application Load Balancer Configuration for Security Groups and Health Check Routes:

Security groups for your Application Load Balancer

PDF | Kindle | RSS

You must ensure that your load balancer can communicate with registered targets on both the listener port and the health check port. Whenever you add a listener to your load balancer or update the health check port for a target group used by the load balancer to route requests, you must verify that the security groups associated with the load balancer allow traffic on the new port in both directions. If they do not, you can edit the rules for the currently associated security groups or associate different security groups with the load balancer.

Recommended rules

The following are the recommended rules for an Internet-facing load balancer.

Inbound		
Source	**Port Range**	**Comment**
0.0.0.0/0	*listener*	Allow all inbound traffic on the load balancer listener port
Outbound		
Destination	**Port Range**	**Comment**
instance security group	*instance listener*	Allow outbound traffic to instances on the instance listener port
instance security group	*health check*	Allow outbound traffic to instances on the health check port

The following are the recommended rules for an internal load balancer.

Inbound		
Source	**Port Range**	**Comment**
VPC CIDR	*listener*	Allow inbound traffic from the VPC CIDR on the load balancer listener port
Outbound		
Destination	**Port Range**	**Comment**
instance security group	*instance listener*	Allow outbound traffic to instances on the instance listener port
instance security group	*health check*	Allow outbound traffic to instances on the health check port

via - https://docs.aws.amazon.com/elasticloadbalancing/latest/application/load-balancer-update-security-groups.html

Incorrect options:

The Amazon Elastic Block Store (Amazon EBS) volumes have been improperly mounted - You can access the website using the IP address which means there is no issue with the Amazon EBS volumes. So this option is not correct.

Your web-app has a runtime that is not supported by the Application Load Balancer - There is no connection between a web app runtime and the application load balancer. This option has been added as a distractor.

You need to attach elastic IP address (EIP) to the Amazon EC2 instances - This option is a distractor as Elastic IPs do not need to be assigned to Amazon EC2 instances while using an Application Load Balancer.

References:

https://docs.aws.amazon.com/elasticloadbalancing/latest/application/load-balancer-update-security-groups.html

https://docs.aws.amazon.com/elasticloadbalancing/latest/application/target-group-health-checks.html

Domain

Design Secure Architectures

Question 55Skipped

A small rental company had 5 employees, all working under the same AWS cloud account. These employees deployed their applications built for various functions- including billing, operations, finance, etc. Each of these employees has been operating in their own VPC. Now, there is a need to connect these VPCs so that the applications can communicate with each other.

Which of the following is the MOST cost-effective solution for this use-case?

Use an Internet Gateway

Use an AWS Direct Connect connection

Use a Network Address Translation gateway (NAT gateway)

Correct answer

Use a VPC peering connection

Overall explanation

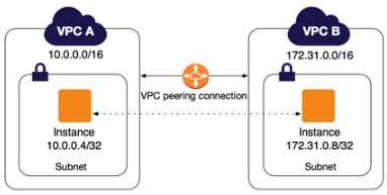
Correct option:

Use a VPC peering connection

A VPC peering connection is a networking connection between two VPCs that enables you to route traffic between them using private IPv4 addresses or IPv6 addresses. Instances in either VPC can communicate with each other as if they are within the same network. You can create a VPC peering connection between your own VPCs, or with a VPC in another AWS account. The VPCs can be in different regions (also known as an inter-region VPC peering connection). VPC Peering helps connect two VPCs and is not transitive. To connect VPCs together, the best available option is to use VPC peering.

More on VPC Peering:

A VPC peering connection is a networking connection between two VPCs that enables you to route traffic between them using private IPv4 addresses or IPv6 addresses. Instances in either VPC can communicate with each other as if they are within the same network. You can create a VPC peering connection between your own VPCs, or with a VPC in another AWS account. The VPCs can be in different regions (also known as an inter-region VPC peering connection).



AWS uses the existing infrastructure of a VPC to create a VPC peering connection; it is neither a gateway nor a VPN connection, and does not rely on a separate piece of physical hardware. There is no single point of failure for communication or a bandwidth bottleneck.

A VPC peering connection helps you to facilitate the transfer of data. For example, if you have more than one AWS account, you can peer the VPCs across those accounts to create a file sharing network. You can also use a VPC peering connection to allow other VPCs to access resources you have in one of your VPCs.

You can establish peering relationships between VPCs across different AWS Regions (also called Inter-Region VPC Peering). This allows VPC resources including EC2 instances, Amazon RDS databases and Lambda functions that run in different AWS Regions to communicate with each other using private IP addresses, without requiring gateways, VPN connections, or separate network appliances. The traffic remains in the private IP space. All inter-region traffic is encrypted with no single point of failure, or bandwidth bottleneck. Traffic always stays on the global AWS backbone, and never traverses the public internet, which reduces threats, such as common exploits, and DDoS attacks. Inter-Region VPC Peering provides a simple and cost-effective way to share resources between regions or replicate data for geographic redundancy.

via - https://docs.aws.amazon.com/vpc/latest/peering/what-is-vpc-peering.html

Incorrect options:

Use an Internet Gateway - An internet gateway is a horizontally scaled, redundant, and highly available VPC component that allows communication between instances in your VPC and the internet. It, therefore, imposes no availability risks or bandwidth constraints on your network traffic. Internet Gateway is not meant for connecting between VPCs.

Use an AWS Direct Connect connection - AWS Direct Connect is a cloud service solution that makes it easy to establish a dedicated network connection from your premises to AWS. AWS Direct Connect lets you establish a dedicated network connection between your network and one of the AWS Direct Connect locations. Using industry-standard 802.1q VLANs, this dedicated connection can be partitioned into multiple virtual interfaces. For the given use-case, direct connect gateway is overkill and is not as cost-optimal as using VPC peering.

Use a Network Address Translation gateway (NAT gateway) - You can use a network address translation (NAT) gateway to enable instances in a private subnet to connect to the internet or other AWS services, but prevent the internet from initiating a connection with those instances. You are charged for creating and using a NAT gateway in your account. NAT gateway hourly usage and data processing rates apply. NAT Gateway is not used for connection between VPCs.

Reference:

https://docs.aws.amazon.com/vpc/latest/peering/what-is-vpc-peering.html

Domain

Design Resilient Architectures

Question 56Skipped

The engineering team at an e-commerce company has been tasked with migrating to a serverless architecture. The team wants to focus on the key points of consideration when using AWS Lambda as a backbone for this architecture.

As a Solutions Architect, which of the following options would you identify as correct for the given requirement? (Select three)

The bigger your deployment package, the slower your AWS Lambda function will cold-start. Hence, AWS suggests packaging dependencies as a separate package from the actual AWS Lambda package

Correct selection

By default, AWS Lambda functions always operate from an AWS-owned VPC and hence have access to any public internet address or public AWS APIs. Once an AWS Lambda function is VPC-enabled, it will need a route through a Network Address Translation gateway (NAT gateway) in a public subnet to access public resources

Serverless architecture and containers complement each other but you cannot package and deploy AWS Lambda functions as container images

Correct selection

Since AWS Lambda functions can scale extremely quickly, it's a good idea to deploy a Amazon CloudWatch Alarm that notifies your team when function metrics such as ConcurrentExecutions or Invocations exceeds the expected threshold

Correct selection

If you intend to reuse code in more than one AWS Lambda function, you should consider creating an AWS Lambda Layer for the reusable code

AWS Lambda allocates compute power in proportion to the memory you allocate to your function. AWS, thus recommends to over provision your function time out settings for the proper performance of AWS Lambda functions

Overall explanation

Correct options:

By default, AWS Lambda functions always operate from an AWS-owned VPC and hence have access to any public internet address or public AWS APIs. Once an AWS Lambda function is VPC-enabled, it will need a route through a Network Address Translation gateway (NAT gateway) in a public subnet to access public resources

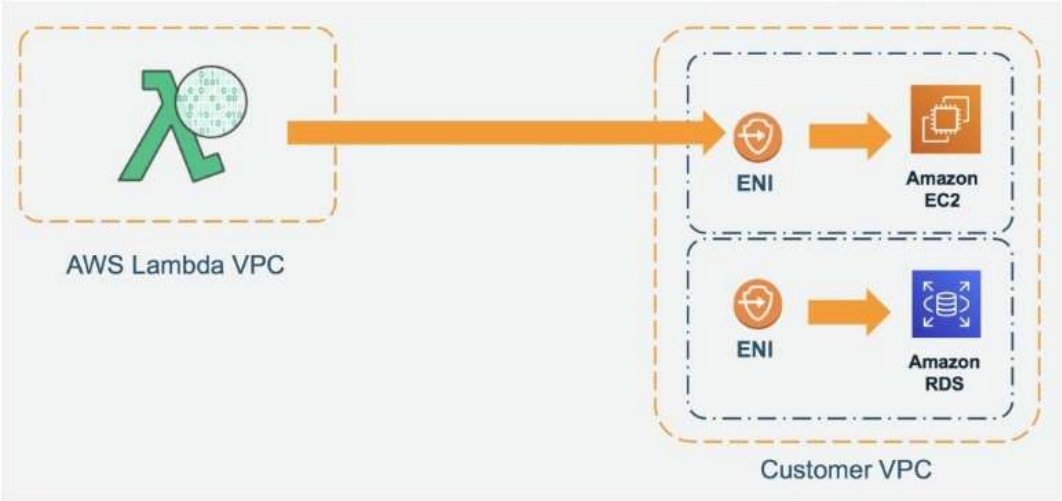
AWS Lambda functions always operate from an AWS-owned VPC. By default, your function has the full ability to make network requests to any public internet address — this includes access to any of the public AWS APIs. For example, your function can interact with AWS DynamoDB APIs to PutItem or Query for records. You should only enable your functions for VPC access when you need to interact with a private resource located in a private subnet. An Amazon RDS instance is a good example.

Once your function is VPC-enabled, all network traffic from your function is subject to the routing rules of your VPC/Subnet. If your function needs to interact with a public resource, you will need a route through a NAT gateway in a public subnet.

When to VPC-Enable an AWS Lambda Function:

Tip #1: When to VPC-Enable a Lambda Function

Lambda functions always operate from an AWS-owned VPC. By default, your function has full ability to make network requests to any public internet address — this includes access to any of the public AWS APIs. For example, your function can interact with AWS DynamoDB APIs to PutItem or Query for records. You should only enable your functions for VPC access when you need to interact with a private resource located in a private subnet. An RDS instance is a good example.



Once your function is VPC-enabled, all network traffic from your function is subject to the routing rules of your VPC/Subnet. If your function needs to interact with a public resource, you will need a route through a NAT gateway in a public subnet.

via - https://aws.amazon.com/blogs/architecture/best-practices-for-developing-on-aws-lambda/

Since AWS Lambda functions can scale extremely quickly, it's a good idea to deploy a Amazon CloudWatch Alarm that notifies your team when function metrics such as ConcurrentExecutions or Invocations exceeds the expected threshold

Since AWS Lambda functions can scale extremely quickly, this means you should have controls in place to notify you when you have a spike in concurrency. A good idea is to deploy an Amazon CloudWatch Alarm that notifies your team when function metrics such as ConcurrentExecutions or Invocations exceeds your threshold. You should create an AWS Budget so you can monitor costs on a daily basis.

If you intend to reuse code in more than one AWS Lambda function, you should consider creating an AWS Lambda Layer for the reusable code

You can configure your AWS Lambda function to pull in additional code and content in the form of layers. A layer is a ZIP archive that contains libraries, a custom runtime, or other dependencies. With layers, you can use libraries in your function without needing to include

them in your deployment package. Layers let you keep your deployment package small, which makes development easier. A function can use up to 5 layers at a time.

You can create layers, or use layers published by AWS and other AWS customers. Layers support resource-based policies for granting layer usage permissions to specific AWS accounts, AWS Organizations, or all accounts. The total unzipped size of the function and all layers can't exceed the unzipped deployment package size limit of 250 megabytes.

Incorrect options:

AWS Lambda allocates compute power in proportion to the memory you allocate to your function. AWS, thus recommends to over provision your function time out settings for the proper performance of AWS Lambda functions - AWS Lambda allocates compute power in proportion to the memory you allocate to your function. This means you can over-provision memory to run your functions faster and potentially reduce your costs. However, AWS recommends that you should not over provision your function time out settings. Always understand your code performance and set a function time out accordingly. Overprovisioning function timeout often results in Lambda functions running longer than expected and unexpected costs.

The bigger your deployment package, the slower your AWS Lambda function will cold-start. Hence, AWS suggests packaging dependencies as a separate package from the actual AWS Lambda package - This statement is incorrect and acts as a distractor. All the dependencies are also packaged into the single Lambda deployment package.

Serverless architecture and containers complement each other but you cannot package and deploy AWS Lambda functions as container images - This statement is incorrect. You can now package and deploy AWS Lambda functions as container images.

References:

https://aws.amazon.com/blogs/architecture/best-practices-for-developing-on-aws-lambda/

https://docs.aws.amazon.com/lambda/latest/dg/configuration-layers.html

https://aws.amazon.com/blogs/aws/new-for-aws-lambda-container-image-support/

Domain

Design High-Performing Architectures

Question 57Skipped

An e-commerce company wants to migrate its on-premises application to AWS. The application consists of application servers and a Microsoft SQL Server database. The solution should result in the maximum possible availability for the database layer while minimizing operational and management overhead.

As a solutions architect, which of the following would you recommend to meet the given requirements?

Migrate the data to Amazon EC2 instance hosted SQL Server database. Deploy the Amazon EC2 instances in a Multi-AZ configuration

Correct answer

Migrate the data to Amazon RDS for SQL Server database in a Multi-AZ deployment

Migrate the data to Amazon RDS for SQL Server database in a cross-region read-replica configuration

Migrate the data to Amazon RDS for SQL Server database in a cross-region Multi-AZ deployment

Overall explanation

Correct option:

Migrate the data to Amazon RDS for SQL Server database in a Multi-AZ deployment

Amazon RDS supports Multi-AZ deployments for Microsoft SQL Server by using either SQL Server Database Mirroring (DBM) or Always On Availability Groups (AGs). Amazon RDS monitors and maintains the health of your Multi-AZ deployment. If problems occur, Amazon RDS automatically repairs unhealthy database instances, reestablishes synchronization, and initiates failovers.

Multi-AZ deployments provide increased availability, data durability, and fault tolerance for database instances. In the event of planned database maintenance or unplanned service disruption, Amazon RDS automatically fails over to the up-to-date secondary database instance. This functionality lets database operations resume quickly without manual intervention. The primary and standby instances use the same endpoint, whose physical network address transitions to the secondary replica as part of the failover process. You don't have to reconfigure your application when a failover occurs.

This option provides the maximum possible availability for the database layer while minimizing operational and management overhead.

Incorrect options:

Migrate the data to Amazon EC2 instance hosted SQL Server database. Deploy the Amazon EC2 instances in a Multi-AZ configuration - Hosting SQL Server database on Amazon EC2 instance involves significant operational and management overhead in terms of OS patching, database patching, etc. So this option is incorrect.

Migrate the data to Amazon RDS for SQL Server database in a cross-region read-replica configuration - Amazon RDS Read Replicas enable you to create one or more read-only copies of your database instance within the same AWS Region or in a different AWS Region. Read replicas are used to enhance the read scalability of a database. You cannot use read replicas to improve the availability of a database. Therefore this option is incorrect.

Migrate the data to Amazon RDS for SQL Server database in a cross-region Multi-AZ deployment - Amazon RDS Multi-AZ deployments provide enhanced availability for database instances within a single AWS Region. There is no such thing as a cross-region Multi-AZ deployment. Hence this option is incorrect.

References:

https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/USER_SQLServerMultiAZ.html

https://aws.amazon.com/about-aws/whats-new/2018/01/amazon-rds-read-replicas-now-support-multi-az-deployments/

Domain

Design Resilient Architectures

Question 58Skipped

A company runs a popular dating website on the AWS Cloud. As a Solutions Architect, you've designed the architecture of the website to follow a serverless pattern on the AWS Cloud using Amazon API Gateway and AWS Lambda. The backend uses an Amazon RDS PostgreSQL database. Currently, the application uses a username and password combination to connect the AWS Lambda function to the Amazon RDS database.

You would like to improve the security at the authentication level by leveraging short-lived credentials. What will you choose? (Select two)

Deploy AWS Lambda in a VPC

Embed a credential rotation logic in the AWS Lambda, retrieving them from SSM

Correct selection

Attach an AWS Identity and Access Management (IAM) role to AWS Lambda

Restrict the Amazon RDS database security group to the AWS Lambda's security group

Correct selection

Use IAM authentication from AWS Lambda to Amazon RDS PostgreSQL

Overall explanation

Correct options:

Use IAM authentication from AWS Lambda to Amazon RDS PostgreSQL

Attach an AWS Identity and Access Management (IAM) role to AWS Lambda

You can authenticate to your database instance using AWS Identity and Access Management (IAM) database authentication. IAM database authentication works with MySQL and PostgreSQL. With this authentication method, you don't need to use a password when you connect to a database instance. Instead, you use an authentication token.

An authentication token is a unique string of characters that Amazon RDS generates on request. Authentication tokens are generated using AWS Signature Version 4. Each token has a lifetime of 15 minutes. You don't need to store user credentials in the database, because authentication is managed externally using IAM. You can also still use standard database authentication.

IAM database authentication provides the following benefits: 1. Network traffic to and from the database is encrypted using Secure Sockets Layer (SSL). 2. You can use IAM to centrally manage access to your database resources, instead of managing access individually on each DB instance. 3. For applications running on Amazon EC2, you can use profile credentials specific to your Amazon EC2 instance to access your database instead of a password, for greater security.

Incorrect options:

AWS Systems Manager Parameter Store (aka SSM Parameter Store) provides secure, hierarchical storage for configuration data management and secrets management. You can store data such as passwords, database strings, Amazon EC2 instance IDs, Amazon Machine Image (AMI) IDs, and license codes as parameter values. You can store values as plain text or encrypted data.

Embed a credential rotation logic in the AWS Lambda, retrieving them from SSM - Retrieving credentials from SSM is overkill for the expected solution and hence this is not a correct option.

Restrict the Amazon RDS database security group to the AWS Lambda's security group

Deploy AWS Lambda in a VPC

This question is very tricky because all answers do indeed increase security. But the question is related to authentication mechanisms, and as such, deploying an AWS Lambda in a VPC or tightening security groups does not change the authentication layer. IAM authentication to Amazon RDS is supported, which must be achieved by attaching an IAM role the AWS Lambda function

References:

https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/UsingWithRDS.IAMDBAuth.html

https://docs.aws.amazon.com/systems-manager/latest/userguide/systems-manager-parameter-store.html

Domain

Design Secure Architectures

Question 59Skipped

The engineering team at a global e-commerce company is currently reviewing their disaster recovery strategy. The team has outlined that they need to be able to quickly recover their application stack with a Recovery Time Objective (RTO) of 5 minutes, in all of the AWS Regions that the application runs. The application stack currently takes over 45 minutes to install on a Linux system.

As a Solutions architect, which of the following options would you recommend as the disaster recovery strategy?

Use Amazon EC2 user data to speed up the installation process

Correct answer

Create an Amazon Machine Image (AMI) after installing the software and copy the AMI across all Regions. Use this Region-specific AMI to run the recovery process in the respective Regions

Create an Amazon Machine Image (AMI) after installing the software and use this AMI to run the recovery process in other Regions

Store the installation files in Amazon S3 for quicker retrieval

Overall explanation

Correct option:

Create an Amazon Machine Image (AMI) after installing the software and copy the AMI across all Regions. Use this Region-specific AMI to run the recovery process in the respective Regions

An Amazon Machine Image (AMI) provides the information required to launch an instance. You must specify an AMI when you launch an instance. You can launch multiple instances from a single AMI when you need multiple instances with the same configuration. You can use different AMIs to launch instances when you need instances with different configurations.

For the current use case, you need to create an AMI such that the application stack is already set up. But AMIs are bound to the Region they are created in. So, you need to copy the AMI across Regions for disaster recovery readiness.

Copying a source AMI results in an identical but distinct target AMI with its own unique identifier. In the case of an Amazon EBS-backed AMI, each of its backing snapshots is, by default, copied to an identical but distinct target snapshot. (The sole exceptions are when you choose to encrypt or re-encrypt the snapshot.) You can change or deregister the source AMI with no effect on the target AMI. The reverse is also true. There are no charges for copying an AMI. However, standard storage and data transfer rates apply. If you copy an Amazon EBS-backed AMI, you will incur charges for the storage of any additional Amazon EBS snapshots.

AWS does not copy launch permissions, user-defined tags, or Amazon S3 bucket permissions from the source AMI to the new AMI. After the copy operation is complete, you can apply launch permissions, user-defined tags, and Amazon S3 bucket permissions to the new AMI.

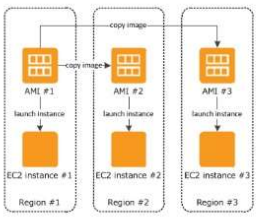
AMIs Cross-Region copying:

Cross-Region copying

Copying an AMI across geographically diverse Regions provides the following benefits:

- Consistent global deployment: Copying an AMI from one Region to another enables you to launch consistent instances in different Regions based on the same AMI.
- Scalability: You can more easily design and build global applications that meet the needs of your users, regardless of their location.
- Performance: You can increase performance by distributing your application, as well as locating critical components of your application in closer proximity to your users. You can also take advantage of Region-specific features, such as instance types or other AWS services.
- High availability: You can design and deploy applications across AWS regions, to increase availability.

The following diagram shows the relations among a source AMI and two copied AMIs in different Regions, as well as the EC2 instances launched from each. When you launch an instance from an AMI, it resides in the same Region where the AMI resides. If you make changes to the source AMI and want those changes to be reflected in the AMIs in the target Regions, you must recopy the source AMI to the target Regions.



When you first copy an instance store-backed AMI to a Region, we create an Amazon S3 bucket for the AMIs copied to that Region. All instance store-backed AMIs that you copy to that Region are stored in this bucket. The bucket names have the following format: amis-for-*account*-in-*region*-*hash*. For example: `amis-for-123456789012-in-us-east-2-yhjmxvp6`.

Prerequisite

Prior to copying an AMI, you must ensure that the contents of the source AMI are updated to support running in a different Region. For example, you should update any database connection strings or similar application configuration data to point to the appropriate resources. Otherwise, instances launched from the new AMI in the destination Region may still use the resources from the source Region, which can impact performance and cost.

via - https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/CopyingAMIs.html

Incorrect options:

Store the installation files in Amazon S3 for quicker retrieval - Amazon Simple Storage Service (Amazon S3) is an object storage service from AWS. It will not help with speeding up the installation since Amazon S3 is a storage service. You will still need an Amazon EC2 instance to have the necessary installation environment.

Use Amazon EC2 user data to speed up the installation process - User data of an Amazon EC2 instance can be used to perform common automated configuration tasks or run scripts after the instance starts. User data, cannot, however, be used to install the application. Amazon EC2 user data would not help as it would run the same installation script for the same duration of 45 minutes.

Create an Amazon Machine Image (AMI) after installing the software and use this AMI to run the recovery process in other Regions - As discussed above, AMIs are Region-specific and need to be copied to all Regions they are intended to be used in.

References:

https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/AMIs.html

https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/CopyingAMIs.html

Domain

Design Resilient Architectures

Question 60Skipped

An e-commerce company tracks user clicks on its flagship website and performs analytics to provide near-real-time product recommendations. An Amazon EC2 instance receives data from the website and sends the data to an Amazon Aurora Database instance. Another Amazon EC2 instance continuously checks the changes in the database and executes SQL queries to provide recommendations. Now, the company wants a redesign to decouple and scale the infrastructure. The solution must ensure that data can be analyzed in real-time without any data loss even when the company sees huge traffic spikes.

What would you recommend as an AWS Certified Solutions Architect - Associate?

Leverage Amazon Kinesis Data Streams to capture the data from the website and feed it into Amazon QuickSight which can query the data in real time. Lastly, the analyzed feed is output into Kinesis Data Firehose to persist the data on Amazon S3

Correct answer

Leverage Amazon Kinesis Data Streams to capture the data from the website and feed it into Amazon Kinesis Data Analytics which can query the data in real time. Lastly, the analyzed feed is output into Amazon Kinesis Data Firehose to persist the data on Amazon S3

Leverage Amazon Kinesis Data Streams to capture the data from the website and feed it into Amazon Kinesis Data Firehose to persist the data on Amazon S3. Lastly, use Amazon Athena to analyze the data in real time

Leverage Amazon SQS to capture the data from the website. Configure a fleet of Amazon EC2 instances under an Auto scaling group to process messages from the Amazon SQS queue and trigger the scaling policy based on the number of pending messages in the queue. Perform real-time analytics using a third-party library on the Amazon EC2 instances

Overall explanation

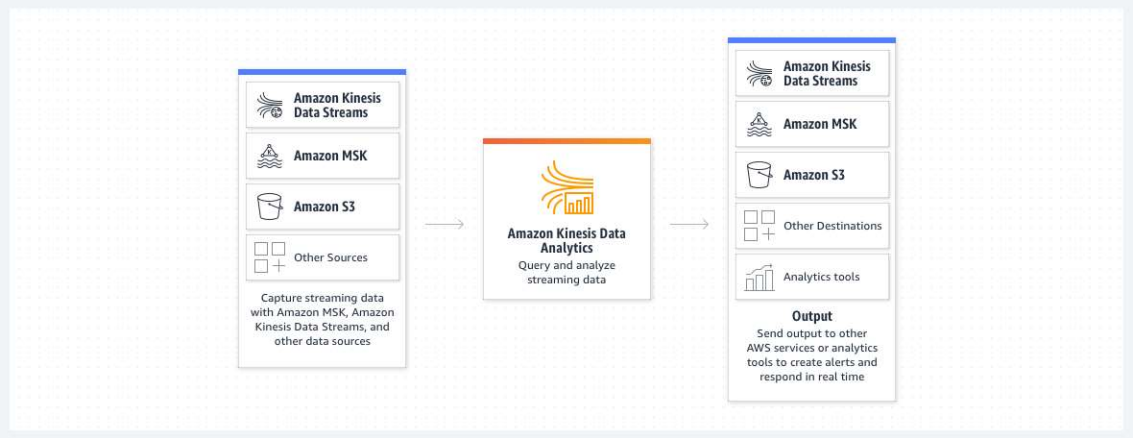
Correct option:

Leverage Amazon Kinesis Data Streams to capture the data from the website and feed it into Amazon Kinesis Data Analytics which can query the data in real time. Lastly, the analyzed feed is output into Amazon Kinesis Data Firehose to persist the data on Amazon S3

You can use Amazon Kinesis Data Streams to build custom applications that process or analyze streaming data for specialized needs. Amazon Kinesis Data Streams manages the infrastructure, storage, networking, and configuration needed to stream your data at the level of your data throughput. You don't have to worry about provisioning, deployment, or ongoing maintenance of hardware, software, or other services for your data streams.

For the given use case, you can use Amazon Kinesis Data Analytics to transform and analyze incoming streaming data from Kinesis Data Streams in real time. Kinesis Data Analytics takes care of everything required to run streaming applications continuously, and scales automatically to match the volume and throughput of your incoming data. With Amazon Kinesis Data Analytics, there are no servers to manage, no minimum fee or setup cost, and you only pay for the resources your streaming applications consume.

Amazon Kinesis Data Analytics:

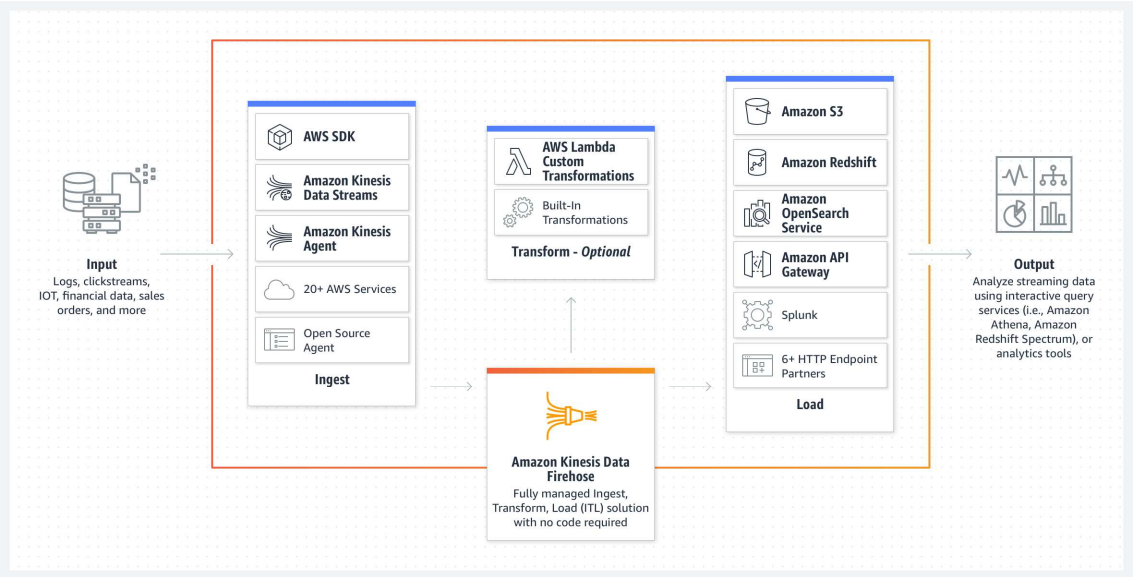


via - https://aws.amazon.com/kinesis/

Amazon Kinesis Data Firehose is an extract, transform, and load (ETL) service that reliably captures, transforms and delivers streaming data to data lakes, data stores, and analytics services.

For the given use case, post the real-time analysis, the output feed from Kinesis Data Analytics is output into Kinesis Data Firehose which dumps the data into Amazon S3 without any data loss.

Amazon Kinesis Data Firehose:



via - https://aws.amazon.com/kinesis/

Incorrect options:

Leverage Amazon Kinesis Data Streams to capture the data from the website and feed it into Amazon QuickSight which can query the data in real time. Lastly, the analyzed feed is output into Kinesis Data Firehose to persist the data on Amazon S3 - Amazon QuickSight cannot use Amazon Kinesis Data Streams as a source. In addition, Amazon QuickSight cannot be used for real-time streaming data analysis from its source. Therefore this option is incorrect.

Leverage Amazon Kinesis Data Streams to capture the data from the website and feed it into Amazon Kinesis Data Firehose to persist the data on Amazon S3. Lastly, use Amazon Athena to analyze the data in real time - Amazon Athena is an interactive query service that makes it easy to analyze data in Amazon S3 using standard SQL. Athena is serverless, so there is no infrastructure to manage, and you pay only for the queries that you run. Athena cannot be used to analyze data in real time. Therefore this option is incorrect.

Leverage Amazon SQS to capture the data from the website. Configure a fleet of Amazon EC2 instances under an Auto scaling group to process messages from the Amazon SQS queue and trigger the scaling policy based on the number of pending messages in the queue. Perform real-time analytics using a third-party library on the Amazon EC2 instances - Even though using Amazon SQS with Amazon EC2 instances can decouple the architecture, however, performing real-time analytics using a third party library on the Amazon EC2 instances is not the best fit solution for the given use case. The Amazon Kinesis family of services is the better fit for the given scenario as these services allow streaming data ingestion, real-time analysis, and reliable data delivery to the data sink.

References:

https://aws.amazon.com/kinesis/

https://aws.amazon.com/quicksight/resources/faqs/

Domain

Design High-Performing Architectures

Question 61Skipped

As a solutions architect, you have created a solution that utilizes an Application Load Balancer with stickiness and an Auto Scaling Group (ASG). The Auto Scaling Group spans across 2 Availability Zones (AZs). AZ-A has 3 Amazon EC2 instances and AZ-B has 4 Amazon EC2 instances. The Auto Scaling Group is about to go into a scale-in event due to the triggering of a Amazon CloudWatch alarm.

What will happen under the default Auto Scaling Group configuration?

Correct answer

The instance with the oldest launch template or launch configuration will be terminated in AZ-B

An instance in the AZ-A will be created

A random instance in the AZ-A will be terminated

A random instance will be terminated in AZ-B

Overall explanation

Correct option:

The instance with the oldest launch template or launch configuration will be terminated in AZ-B

Amazon EC2 Auto Scaling helps you ensure that you have the correct number of Amazon EC2 instances available to handle the load for your application. You create collections of Amazon EC2 instances, called Auto Scaling groups. You can specify the minimum number of instances in each Auto Scaling group, and Amazon EC2 Auto Scaling ensures that your group never goes below this size.

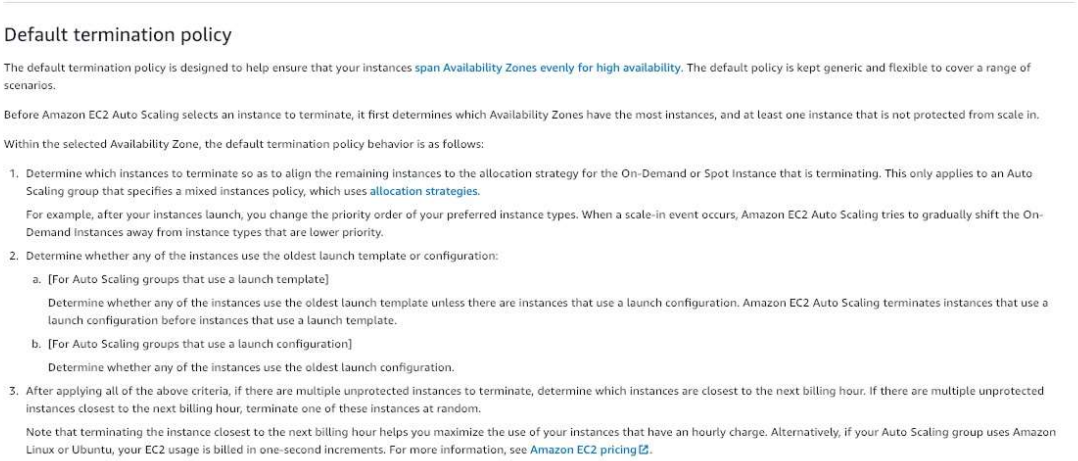
With each Auto Scaling group, you can control when it adds instances (referred to as scaling out) or removes instances (referred to as scaling in) from your network architecture.

The default termination policy is designed to help ensure that your instances span Availability Zones evenly for high availability. The default policy is kept generic and flexible to cover a range of scenarios.

The default termination policy behavior is as follows: 1. Determine which Availability Zones (Azs) have the most instances and at least one instance that is not protected from scale-in. 2. Determine which instances to terminate to align the remaining instances to the allocation strategy for the On-Demand or Spot Instance that is terminating. 3. Determine whether any of the instances use the oldest launch template or configuration: 3.a. Determine whether any of the instances use the oldest launch template unless there are instances that use a launch configuration. 3.b. Determine whether any of the instances use the oldest launch configuration. 4. After applying all of the above criteria, if there are multiple unprotected instances to terminate, determine which instances are closest to the next billing hour.

Per the given use-case, AZs will be balanced first, then the instance with the oldest launch template or launch configuration within the applicable AZ (AZ-B) will be terminated.

Default Termination policy:



Default termination policy

The default termination policy is designed to help ensure that your instances span Availability Zones evenly for high availability. The default policy is kept generic and flexible to cover a range of scenarios.

Before Amazon EC2 Auto Scaling selects an instance to terminate, it first determines which Availability Zones have the most instances, and at least one instance that is not protected from scale in.

Within the selected Availability Zone, the default termination policy behavior is as follows:

1. Determine which instances to terminate so as to align the remaining instances to the allocation strategy for the On-Demand or Spot Instance that is terminating. This only applies to an Auto Scaling group that specifies a mixed instances policy, which uses allocation strategies.

   For example, after your instances launch, you change the priority order of your preferred instance types. When a scale-in event occurs, Amazon EC2 Auto Scaling tries to gradually shift the On-Demand Instances away from instance types that are lower priority.
2. Determine whether any of the instances use the oldest launch template or configuration:
   a. [For Auto Scaling groups that use a launch template]

      Determine whether any of the instances use the oldest launch template unless there are instances that use a launch configuration. Amazon EC2 Auto Scaling terminates instances that use a launch configuration before instances that use a launch template.
   b. [For Auto Scaling groups that use a launch configuration]

      Determine whether any of the instances use the oldest launch configuration.
3. After applying all of the above criteria, if there are multiple unprotected instances to terminate, determine which instances are closest to the next billing hour. If there are multiple unprotected instances closest to the next billing hour, terminate one of these instances at random.

   Note that terminating the instance closest to the next billing hour helps you maximize the use of your instances that have an hourly charge. Alternatively, if your Auto Scaling group uses Amazon Linux or Ubuntu, your EC2 usage is billed in one-second increments. For more information, see Amazon EC2 pricing.

via - https://docs.aws.amazon.com/autoscaling/ec2/userguide/as-instance-termination.html

Incorrect options:

A random instance in the AZ-A will be terminated

An instance in the AZ-A will be created

A random instance will be terminated in AZ-B

These three options contradict the details provided in the explanation above. Hence these are incorrect.

Reference:

https://docs.aws.amazon.com/autoscaling/ec2/userguide/as-instance-termination.html

Domain

Design Resilient Architectures

Question 62Skipped

The engineering team at a leading e-commerce company is anticipating a surge in the traffic because of a flash sale planned for the weekend. You have estimated the web traffic to be 10x. The content of your website is highly dynamic and changes very often.

As a Solutions Architect, which of the following options would you recommend to make sure your infrastructure scales for that day?

Use an Amazon CloudFront distribution in front of your website

Correct answer

Use an Auto Scaling Group

Use an Amazon Route 53 Multi Value record

Deploy the website on Amazon S3

Overall explanation

Correct option:

Use an Auto Scaling Group

An Auto Scaling group (ASG) contains a collection of Amazon EC2 instances that are treated as a logical grouping for automatic scaling and management. An Auto Scaling group also enables you to use Amazon EC2 Auto Scaling features such as health check replacements and scaling policies. Both maintaining the number of instances in an Auto Scaling group and automatic scaling are the core functionality of the Amazon EC2 Auto Scaling service.

The size of an Auto Scaling group depends on the number of instances that you set as the desired capacity. You can adjust its size to meet demand, either manually or by using automatic scaling.

An Auto Scaling group starts by launching enough instances to meet its desired capacity. It maintains this number of instances by performing periodic health checks on the instances in the group. The Auto Scaling group continues to maintain a fixed number of instances even if an instance becomes unhealthy. If an instance becomes unhealthy, the group terminates the unhealthy instance and launches another instance to replace it.

Auto Scaling group is the correct answer here.

Incorrect option:

Use an Amazon CloudFront distribution in front of your website - Amazon CloudFront is a fast content delivery network (CDN) service that securely delivers data, videos, applications, and APIs to customers globally with low latency, high transfer speeds, all within a developer-friendly environment. You can use Amazon CloudFront to improve the performance of your website. Amazon CloudFront makes your website files (such as HTML, images, and video) available from data centers around the world (called edge locations). When a visitor requests a file from your website, CloudFront automatically redirects the request to a copy of the file at the nearest edge location. This results in faster download times than if the visitor had requested the content from a data center that is located farther away.

Amazon CloudFront is not a good solution here as the content is highly dynamic, and Amazon CloudFront will cache things.

Deploy the website on Amazon S3 - You can use Amazon S3 to host a static website. On a static website, individual web pages include static content. They might also contain client-side scripts. To host a static website on Amazon S3, you configure an Amazon S3 bucket for website hosting and then upload your website content to the bucket. When you configure a bucket as a static website, you enable static website hosting, set permissions, and add an index document. Depending on your website requirements, you can also configure other options, including redirects, web traffic logging, and custom error documents.

Dynamic applications cannot be deployed to Amazon S3. This option has been added as a distractor.

Use an Amazon Route 53 Multi Value record - Amazon Route 53 is a highly available and scalable cloud Domain Name System (DNS) web service. Use Multi Value answer routing policy

when you want Route 53 to respond to DNS queries with up to eight healthy records selected at random. Amazon Route 53 does not help in scaling your application. This option has been added as a distractor.

References:

https://docs.aws.amazon.com/autoscaling/ec2/userguide/AutoScalingGroup.html

https://docs.aws.amazon.com/AmazonS3/latest/dev/WebsiteHosting.html

Domain

Design High-Performing Architectures

Question 63Skipped

A healthcare company is evaluating storage options on Amazon S3 to meet regulatory guidelines. The data should be stored in such a way on Amazon S3 that it cannot be deleted until the regulatory time period has expired.

As a solutions architect, which of the following would you recommend for the given requirement?

Correct answer

Use Amazon S3 Object Lock

Use Amazon S3 cross-region replication (S3 CRR)

Activate AWS Multi-Factor Authentication (AWS MFA) delete on the Amazon S3 bucket

Use Amazon S3 Glacier Vault Lock

Overall explanation

Correct option:

Use Amazon S3 Object Lock

Amazon S3 Object Lock is an Amazon S3 feature that allows you to store objects using a write once, read many (WORM) model. You can use WORM protection for scenarios where it is imperative that data is not changed or deleted after it has been written. Whether your business has a requirement to satisfy compliance regulations in the financial or healthcare sector, or you simply want to capture a golden copy of business records for later auditing and reconciliation, Amazon S3 Object Lock is the right tool for you. Object Lock can help prevent objects from being deleted or overwritten for a fixed amount of time or indefinitely.

Using S3 Object Lock

PDF | RSS

With S3 Object Lock, you can store objects using a *write-once-read-many* (WORM) model. Object Lock can help prevent objects from being deleted or overwritten for a fixed amount of time or indefinitely. You can use Object Lock to help meet regulatory requirements that require WORM storage, or to simply add another layer of protection against object changes and deletion.

S3 Object Lock has been assessed by Cohasset Associates for use in environments that are subject to SEC 17a-4, CFTC, and FINRA regulations. For more information about how Object Lock relates to these regulations, see the Cohasset Associates Compliance Assessment.

Object Lock provides two ways to manage object retention: *retention periods* and *legal holds*.

- **Retention period** — Specifies a fixed period of time during which an object remains locked. During this period, your object is WORM-protected and can't be overwritten or deleted. For more information, see Retention periods
- **Legal hold** — Provides the same protection as a retention period, but it has no expiration date. Instead, a legal hold remains in place until you explicitly remove it. Legal holds are independent from retention periods. For more information, see Legal holds.

An object version can have both a retention period and a legal hold, one but not the other, or neither. For more information, see How S3 Object Lock works.

Object Lock works only in versioned buckets, and retention periods and legal holds apply to individual object versions. When you lock an object version, Amazon S3 stores the lock information in the metadata for that object version. Placing a retention period or legal hold on an object protects only the version specified in the request. It doesn't prevent new versions of the object from being created.

If you put an object into a bucket that has the same key name as an existing protected object, Amazon S3 creates a new version of that object, stores it in the bucket as requested, and reports the request as completed successfully. The existing protected version of the object remains locked according to its retention configuration.

To use S3 Object Lock, you follow these basic steps:

1. Create a new bucket with Object Lock enabled.
2. (Optional) Configure a default retention period for objects placed in the bucket.
3. Place the objects that you want to lock in the bucket.
4. Apply a retention period, a legal hold, or both, to the objects that you want to protect.

via - https://docs.aws.amazon.com/AmazonS3/latest/userguide/object-lock.html

Governance mode or Compliance mode?

When setting a retention period for your objects or buckets, you can choose the retention mode you wish to apply to your objects. You can choose either the *Governance* mode or the *Compliance* mode for your objects.

You should use the *Governance* mode if you want to protect objects from being deleted by most users during a pre-defined retention period, but at the same time want some users with special permissions to have the flexibility to alter the retention settings or delete the objects. Users with the s3:BypassGovernanceRetention permission can override or remove governance-mode retention settings. Most customers will use the *Governance* mode since they don't have compliant storage requirements.

You should use the *Compliance* mode if you have a requirement to store compliant data. You should only use the *Compliance* mode if you never want any user, including the root user in your AWS account, to be able to delete the objects during a pre-defined retention period.

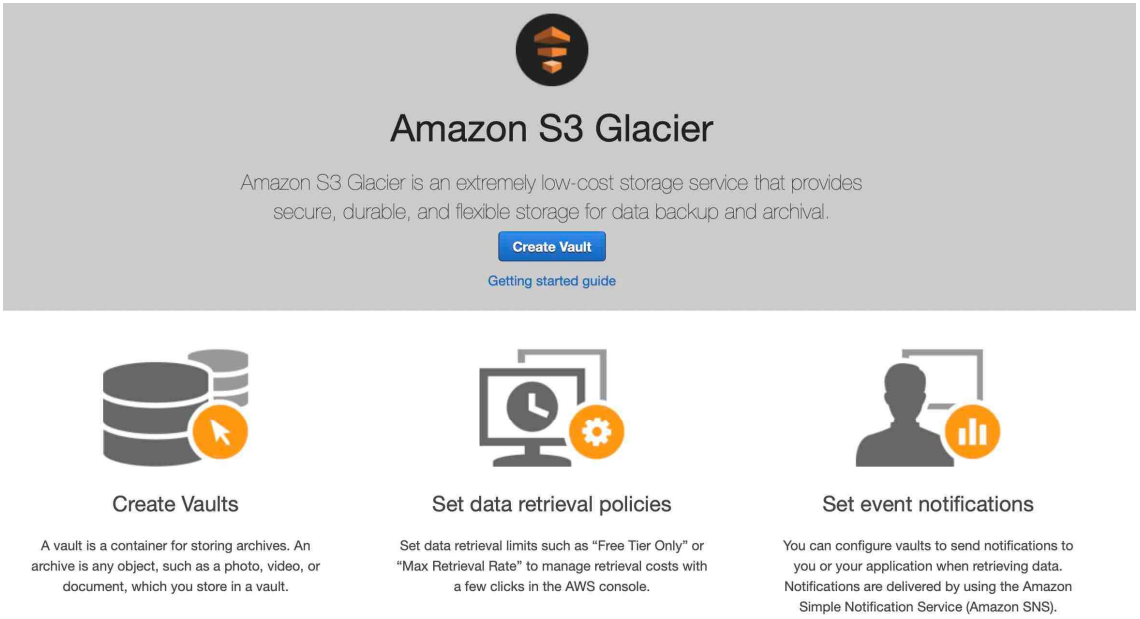
> Note: The only way to delete an object under the *Compliance* mode before its retention date expires is to delete the associated AWS account.

via - https://aws.amazon.com/blogs/storage/protecting-data-with-amazon-s3-object-lock/

Incorrect options:

Use Amazon S3 Glacier Vault Lock

A vault is a container for storing archives on Glacier. When you create a vault, you specify a vault name and the AWS Region in which you want to create the vault. Since Vault Lock is only for Glacier and not for Amazon S3, so it cannot be used for the given use-case.



Use Amazon S3 cross-region replication (S3 CRR) - Replication enables automatic, asynchronous copying of objects across Amazon S3 buckets. The object may be replicated to a single destination bucket or multiple destination buckets. Both source and destination buckets must have versioning enabled. By default, when Amazon S3 Replication is enabled and an object is deleted in the source bucket, Amazon S3 adds a delete marker in the source bucket only. This action protects data from malicious deletions. If you have delete marker replication enabled, these markers are copied to the destination buckets, and Amazon S3 behaves as if the object was deleted in both source and destination buckets. However, someone with administrative access to Amazon S3 can disable cross-Region replication and then delete all versions from both source as well as the destination, so this option will not be able to safeguard your data compared to Amazon S3 Object Lock.

Activate AWS Multi-Factor Authentication (AWS MFA) delete on the Amazon S3 bucket - When working with Amazon S3 Versioning in Amazon S3 buckets, you can optionally add another layer of security by configuring a bucket to enable MFA (multi-factor authentication) delete. When you do this, the bucket owner must include two forms of authentication in any request to delete a version or change the versioning state of the bucket. Only the root account can enable MFA delete. MFA delete cannot be used for the given use case because it just represents an additional security layer and can be disabled by anyone having access to the root account credentials.

References:

https://docs.aws.amazon.com/AmazonS3/latest/userguide/object-lock.html

https://aws.amazon.com/blogs/storage/protecting-data-with-amazon-s3-object-lock/

Domain

Design Secure Architectures

Question 64Skipped

A company uses Application Load Balancers in multiple AWS Regions. The Application Load Balancers receive inconsistent traffic that varies throughout the year. The engineering team at the company needs to allow the IP addresses of the Application Load Balancers in the on-premises firewall to enable connectivity.

Which of the following represents the MOST scalable solution with minimal configuration changes?

Migrate all Application Load Balancers in different Regions to the Network Load Balancers. Configure the on-premises firewall's rule to allow the Elastic IP addresses of all the Network Load Balancers

Set up a Network Load Balancer in one Region. Register the private IP addresses of the Application Load Balancers in different Regions with the Network Load Balancer. Configure the on-premises firewall's rule to allow the Elastic IP address attached to the Network Load Balancer

Develop an AWS Lambda script to get the IP addresses of the Application Load Balancers in different Regions. Configure the on-premises firewall's rule to allow the IP addresses of the Application Load Balancers

Correct answer

Set up AWS Global Accelerator. Register the Application Load Balancers in different Regions to the AWS Global Accelerator. Configure the on-premises firewall's rule to allow static IP addresses associated with the AWS Global Accelerator

Overall explanation

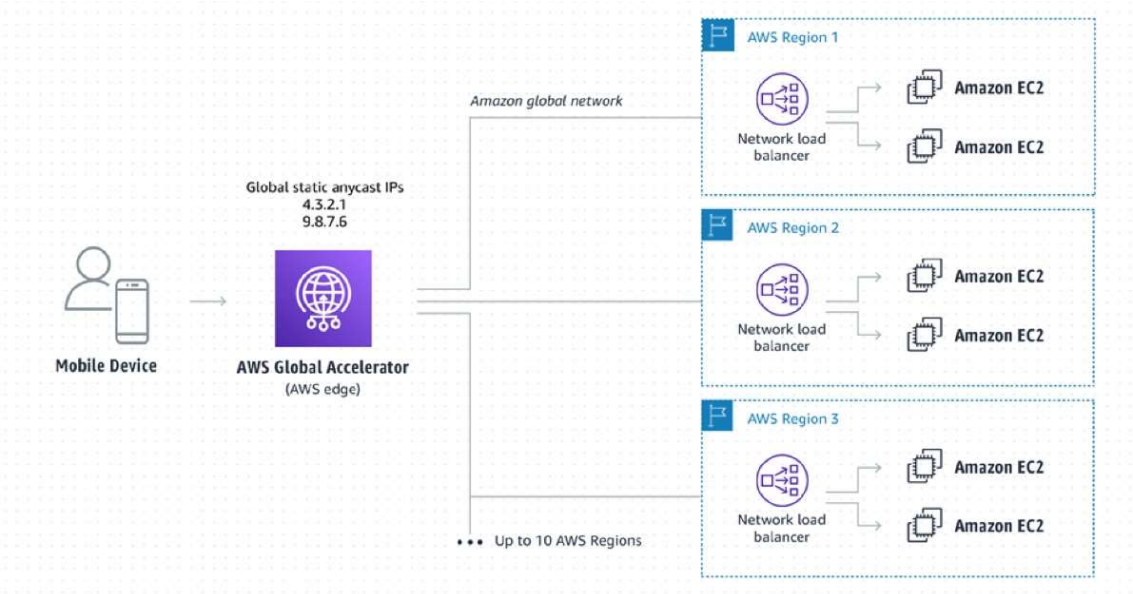
Correct option:

Set up AWS Global Accelerator. Register the Application Load Balancers in different Regions to the AWS Global Accelerator. Configure the on-premises firewall's rule to allow static IP addresses associated with the AWS Global Accelerator

AWS Global Accelerator is a networking service that helps you improve the availability and performance of the applications that you offer to your global users. AWS Global Accelerator is easy to set up, configure, and manage. It provides static IP addresses that provide a fixed entry point to your applications and eliminate the complexity of managing specific IP addresses for different AWS Regions and Availability Zones.

Associate the static IP addresses provided by AWS Global Accelerator to regional AWS resources or endpoints, such as Network Load Balancers, Application Load Balancers, Amazon EC2 Instances, and Elastic IP addresses. The IP addresses are anycast from AWS edge locations so they provide onboarding to the AWS global network close to your users.

Simplified and resilient traffic routing for multi-Region applications using AWS Global Accelerator:



via - https://aws.amazon.com/global-accelerator/

Incorrect options:

Migrate all Application Load Balancers in different Regions to the Network Load Balancers. Configure the on-premises firewall's rule to allow the Elastic IP addresses of all the Network Load Balancers - Although you could potentially migrate the Application Load Balancers to Network Load Balancers, this option requires changes to the on-premises firewall's configuration rules, hence this is not the right fit for the given use-case. It is more optimal to manage the two static IPs provided by the AWS Global Accelerator for configuring the firewall.

Set up a Network Load Balancer in one Region. Register the private IP addresses of the Application Load Balancers in different Regions with the Network Load Balancer. Configure the on-premises firewall's rule to allow the Elastic IP address attached to the Network Load Balancer - Using a single Network Load Balancer is not possible across AWS regions since an Network Load Balancer is Region bound. Multiple Network Load Balancers have to be registered for the on-premises firewall.

Develop an AWS Lambda script to get the IP addresses of the Application Load Balancers in different Regions. Configure the on-premises firewall's rule to allow the IP addresses of the Application Load Balancers - This option requires on-going changes to the on-premises firewall's configuration rules because the IP addresses of the Application Load Balancers would keep changing. Hence this is not the right fit for the given use-case. It is more optimal to configure the firewall with a one-time change for the two static IPs provided by the AWS Global Accelerator.

Reference:

https://aws.amazon.com/global-accelerator/faqs/

Domain

Design High-Performing Architectures

Question 65Skipped

A CRM company has a software as a service (SaaS) application that feeds updates to other in-house and third-party applications. The SaaS application and the in-house applications are being migrated to use AWS services for this inter-application communication.

As a Solutions Architect, which of the following would you suggest to asynchronously decouple the architecture?

Correct answer

Use Amazon EventBridge to decouple the system architecture

Use Amazon Simple Notification Service (Amazon SNS) to communicate between systems and decouple the architecture

Use Amazon Simple Queue Service (Amazon SQS) to decouple the architecture

Use Elastic Load Balancing (ELB) for effective decoupling of system architecture

Overall explanation

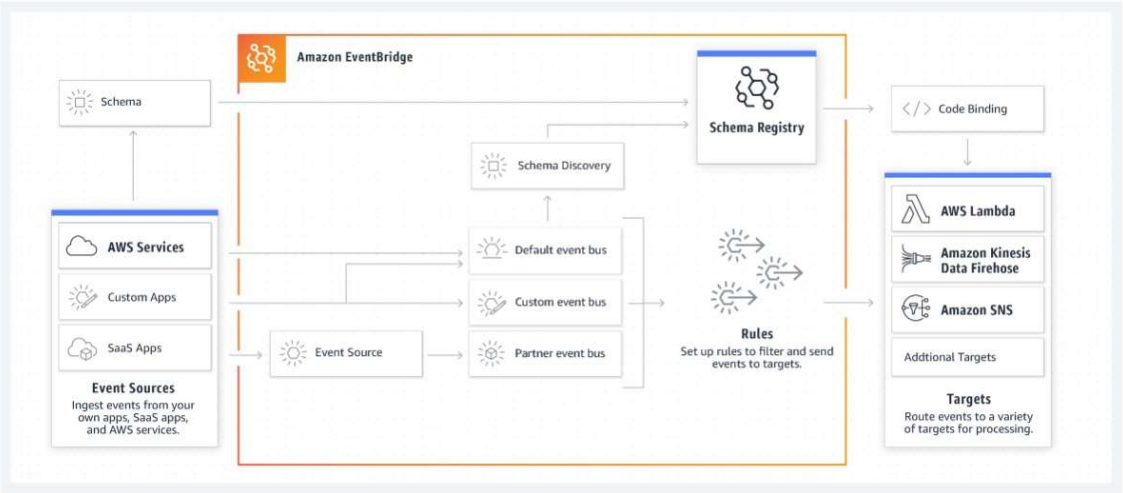
Correct option:

Use Amazon EventBridge to decouple the system architecture

Both Amazon EventBridge and Amazon SNS can be used to develop event-driven applications, but for this use case, EventBridge is the right fit.

Amazon EventBridge is recommended when you want to build an application that reacts to events from SaaS applications and/or AWS services. Amazon EventBridge is the only event-based service that integrates directly with third-party SaaS partners. Amazon EventBridge also automatically ingests events from over 90 AWS services without requiring developers to create any resources in their account. Further, Amazon EventBridge uses a defined JSON-based structure for events and allows you to create rules that are applied across the entire event body to select events to forward to a target. Amazon EventBridge currently supports over 15 AWS services as targets, including AWS Lambda, Amazon SQS, Amazon SNS, and Amazon Kinesis Streams and Firehose, among others. At launch, Amazon EventBridge is has limited throughput (see Service Limits) which can be increased upon request, and typical latency of around half a second.

How Amazon EventBridge works:



via - https://aws.amazon.com/eventbridge/

Incorrect options:

Use Amazon Simple Notification Service (Amazon SNS) to communicate between systems and decouple the architecture - As discussed above, Amazon SNS can be used for event-based services. But, our use case needs integration with third-party SaaS services, hence Amazon EventBridge is the right choice, as Amazon SNS does not support third-party services integration.

Use Amazon Simple Queue Service (Amazon SQS) to decouple the architecture - Amazon SQS is a message queuing service from amazon and works well for decoupling applications. It does not directly integrate with third-party SaaS services.

Use Elastic Load Balancing (ELB) for effective decoupling of system architecture - Elastic Load Balancing (ELB) offers a synchronous decoupling of applications, which is not the right fit for the current use case.

References:

https://aws.amazon.com/eventbridge/

https://aws.amazon.com/sns/

Domain

Design Resilient Architectures