Question 1Skipped

Which of the following AWS services provides a highly available and fault-tolerant solution to capture the clickstream events from the source and then provide a concurrent feed of the data stream to the downstream applications?

Amazon Kinesis Data Analytics

Amazon Simple Queue Service (Amazon SQS)

Amazon Kinesis Data Firehose

Correct answer

Amazon Kinesis Data Streams

Overall explanation

Correct option:

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 data collected is available in milliseconds to enable real-time analytics use cases such as real-time dashboards, real-time anomaly detection, dynamic pricing, and more.

Amazon Kinesis Data Streams enables real-time processing of streaming big data. It provides ordering of records, as well as the ability to read and/or replay records in the same order to multiple Amazon Kinesis Applications. The Amazon Kinesis Client Library (KCL) delivers all records for a given partition key to the same record processor, making it easier to build multiple applications reading from the same Amazon Kinesis data stream (for example, to perform counting, aggregation, and filtering).

Amazon Kinesis Data Streams is recommended when you need the ability for multiple applications to consume the same stream concurrently. For example, you have one application that updates a real-time dashboard and another application that archives data to Amazon Redshift. You want both applications to consume data from the same stream concurrently and independently.

KDS provides the ability for multiple applications to consume the same stream concurrently

Q: When should I use Amazon Kinesis Data Streams, and when should I use Amazon SQS?

We recommend Amazon Kinesis Data Streams for use cases with requirements that are similar to the following:

- Routing related records to the same record processor (as in streaming MapReduce). For example, counting and aggregation are simpler when all records for a given key are routed to the same record processor.
- Ordering of records. For example, you want to transfer log data from the application host to the processing/archival host while maintaining the order of log statements.
- Ability for multiple applications to consume the same stream concurrently. For example, you have one application that updates a real-time dashboard and another that archives data to Amazon Redshift. You want both applications to consume data from the same stream concurrently and independently.
- Ability to consume records in the same order a few hours later. For example, you have a billing application and an audit application that runs a few hours behind the billing application. Because Amazon Kinesis Data Streams stores data for up to 7 days, you can run the audit application up to 7 days behind the billing application.

via - https://aws.amazon.com/kinesis/data-streams/faqs/

Incorrect options:

Amazon Kinesis Data Firehose - Amazon Kinesis Data Firehose is the easiest way to load streaming data into data stores and analytics tools. It can capture, transform, and load streaming data into Amazon S3, Amazon Redshift, Amazon Elasticsearch Service, and Splunk, enabling near real-time analytics with existing business intelligence tools and dashboards you're already using today. It is a fully managed service that automatically scales to match the throughput of your data and requires no ongoing administration. It can also batch, compress, and encrypt the data before loading it, minimizing the amount of storage used at the destination and increasing security. As Kinesis Data Firehose is used to load streaming data into data stores, therefore this option is incorrect.

Amazon Kinesis Data Analytics - Amazon Kinesis Data Analytics is the easiest way to analyze streaming data in real-time. You can quickly build SQL queries and sophisticated Java applications using built-in templates and operators for common processing functions to organize, transform, aggregate, and analyze data at any scale. Kinesis Data Analytics enables you to easily and quickly build queries and sophisticated streaming applications in three simple steps: setup your streaming data sources, write your queries or streaming applications and set up your destination for processed data. As Kinesis Data Analytics is used to build SQL queries and sophisticated Java applications, therefore this option is incorrect.

Amazon Simple Queue Service (Amazon SQS) - 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. 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. For SQS, you cannot have the same message being consumed by multiple consumers at the same time, therefore this option is incorrect.

Exam alert:

Please remember that Amazon Kinesis Data Firehose is used to load streaming data into data stores (Amazon S3, Amazon Redshift, Amazon Elasticsearch Service, and Splunk) whereas Kinesis Data Streams provides support for real-time processing of streaming data. It provides ordering of records, as well as the ability to read and/or replay records in the same order to multiple downstream Amazon Kinesis Applications.

References:

https://aws.amazon.com/kinesis/data-streams/faqs/

https://aws.amazon.com/kinesis/data-firehose/faqs/

https://aws.amazon.com/kinesis/data-analytics/faqs/

Domain

Design Resilient Architectures

Question 2Skipped

The engineering team at a social media company wants to use Amazon CloudWatch alarms to automatically recover Amazon EC2 instances if they become impaired. The team has hired you as a solutions architect to provide subject matter expertise.

As a solutions architect, which of the following statements would you identify as CORRECT regarding this automatic recovery process? (Select two)

If your instance has a public IPv4 address, it does not retain the public IPv4 address after recovery

Correct selection

If your instance has a public IPv4 address, it retains the public IPv4 address after recovery

Correct selection

A recovered instance is identical to the original instance, including the instance ID, private IP addresses, Elastic IP addresses, and all instance metadata

Terminated Amazon EC2 instances can be recovered if they are configured at the launch of instance

During instance recovery, the instance is migrated during an instance reboot, and any data that is in-memory is retained

Overall explanation

Correct options:

A recovered instance is identical to the original instance, including the instance ID, private IP addresses, Elastic IP addresses, and all instance metadata

If your instance has a public IPv4 address, it retains the public IPv4 address after recovery

You can create an Amazon CloudWatch alarm to automatically recover the Amazon EC2 instance if it becomes impaired due to an underlying hardware failure or a problem that requires AWS involvement to repair. Terminated instances cannot be recovered. A recovered instance is identical to the original instance, including the instance ID, private IP addresses, Elastic IP addresses, and all instance metadata. If the impaired instance is in a placement group, the recovered instance runs in the placement group. If your instance has a public IPv4 address, it retains the public IPv4 address after recovery. During instance recovery, the instance is migrated during an instance reboot, and any data that is in-memory is lost.

Incorrect options:

Terminated Amazon EC2 instances can be recovered if they are configured at the launch of instance - This is incorrect as terminated instances cannot be recovered.

During instance recovery, the instance is migrated during an instance reboot, and any data that is in-memory is retained - As mentioned above, during instance recovery, the instance is migrated during an instance reboot, and any data that is in-memory is lost.

If your instance has a public IPv4 address, it does not retain the public IPv4 address after recovery - As mentioned above, if your instance has a public IPv4 address, it retains the public IPv4 address after recovery.

Reference:

https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ec2-instance-recover.html

Domain

Design Resilient Architectures

Question 3Skipped

A media startup is looking at hosting their web application on AWS Cloud. The application will be accessed by users from different geographic regions of the world to upload and download video files that can reach a maximum size of 10 gigabytes. The startup wants the solution to be cost-effective and scalable with the lowest possible latency for a great user experience.

As a Solutions Architect, which of the following will you suggest as an optimal solution to meet the given requirements?

Use Amazon EC2 with Amazon ElastiCache for faster distribution of content, while Amazon S3 can be used as a storage service

Use Amazon EC2 with AWS Global Accelerator for faster distribution of content, while using Amazon S3 as storage service

Correct answer

Use Amazon S3 for hosting the web application and use Amazon S3 Transfer Acceleration (Amazon S3TA) to reduce the latency that geographically dispersed users might face

Use Amazon S3 for hosting the web application and use Amazon CloudFront for faster distribution of content to geographically dispersed users

Overall explanation

Correct option:

Use Amazon S3 for hosting the web application and use Amazon S3 Transfer Acceleration (Amazon S3TA) to reduce the latency that geographically dispersed users might face

Amazon S3 Transfer Acceleration (S3TA) can speed up content transfers to and from Amazon S3 by as much as 50-500% for long-distance transfer of larger objects. Customers who have either web or mobile applications with widespread users or applications hosted far away from their S3 bucket can experience long and variable upload and download speeds over the Internet. S3 Transfer Acceleration (S3TA) reduces the variability in Internet routing, congestion, and speeds that can affect transfers, and logically shortens the distance to S3 for remote applications.

S3TA improves transfer performance by routing traffic through Amazon CloudFront's globally distributed Edge Locations and over AWS backbone networks, and by using network protocol optimizations.

For applications interacting with your Amazon S3 buckets through the S3 API from outside of your bucket's region, S3TA helps avoid the variability in Internet routing and congestion. It does this by routing your uploads and downloads over the AWS global network infrastructure, so you get the benefit of AWS network optimizations.

Incorrect options:

Use Amazon S3 for hosting the web application and use Amazon CloudFront for faster distribution of content to geographically dispersed users - Amazon S3 with Amazon CloudFront is a very powerful way of distributing static content to geographically dispersed users with low latency speeds. If you have objects that are smaller than 1GB or if the data set is less than 1GB in size, you should consider using Amazon CloudFront's PUT/POST commands for optimal performance. The given use case has data larger than 1GB and hence S3 Transfer Acceleration is a better option.

Q: How should I choose between S3 Transfer Acceleration and Amazon CloudFront's PUT/POST?

S3 Transfer Acceleration optimizes the TCP protocol and adds additional intelligence between the client and the S3 bucket, making S3 Transfer Acceleration a better choice if a higher throughput is desired. If you have objects that are smaller than 1GB or if the data set is less than 1GB in size, you should consider using Amazon CloudFront's PUT/POST commands for optimal performance.

via - https://aws.amazon.com/s3/faqs/

Use Amazon EC2 with AWS Global Accelerator for faster distribution of content, while using Amazon S3 as storage service- AWS Global Accelerator is a networking service that sends your user's traffic through Amazon Web Service's global network infrastructure, improving your internet user performance by up to 60%. With AWS Global Accelerator, you are provided two global static customer-facing IPs to simplify traffic management. On the back end, add or remove your AWS application origins, such as Network Load Balancers, Application Load Balancers, Elastic IPs, and Amazon EC2 Instances, without making user-facing changes. As discussed, AWS Global Accelerator is meant for a different use case and is not meant for increasing the speed of Amazon S3 uploads or downloads.

Use Amazon EC2 with Amazon ElastiCache for faster distribution of content, while Amazon S3 can be used as a storage service - 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. Amazon S3 Transfer Acceleration is a better performing option than opting for Amazon EC2 with Amazon ElastiCache, which is not meant to address the given use-case.

Reference:

[https://aws.amazon.com/s3/transfer-acceleration/](https://aws.amazon.com/s3/transfer-acceleration

https://aws.amazon.com/s3/faqs/

Domain

Design High-Performing Architectures

Question 4Skipped

A media company has its corporate headquarters in Los Angeles with an on-premises data center using an AWS Direct Connect connection to the AWS VPC. The branch offices in San Francisco and Miami use AWS Site-to-Site VPN connections to connect to the AWS VPC. The company is looking for a solution to have the branch offices send and receive data with each other as well as with their corporate headquarters.

As a solutions architect, which of the following AWS services would you recommend addressing this use-case?

Software VPN

Correct answer

AWS VPN CloudHub

VPC Peering connection

VPC Endpoint

Overall explanation

Correct option:

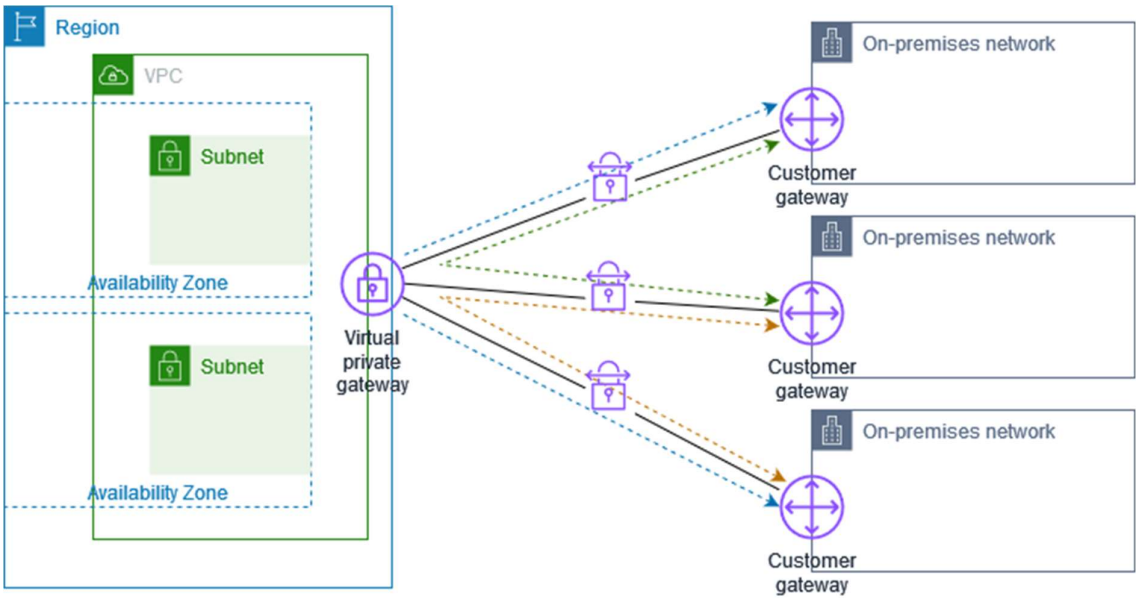
AWS VPN CloudHub

If you have multiple AWS Site-to-Site VPN connections, you can provide secure communication between sites using the AWS VPN CloudHub. This enables your remote sites to communicate with each other, and not just with the VPC. Sites that use AWS Direct Connect connections to the virtual private gateway can also be part of the AWS VPN CloudHub. The VPN CloudHub operates on a simple hub-and-spoke model that you can use with or without a VPC. This design is suitable if you have multiple branch offices and existing internet connections and would like

to implement a convenient, potentially low-cost hub-and-spoke model for primary or backup connectivity between these remote offices.

Per the given use-case, the corporate headquarters has an AWS Direct Connect connection to the VPC and the branch offices have Site-to-Site VPN connections to the VPC. Therefore using the AWS VPN CloudHub, branch offices can send and receive data with each other as well as with their corporate headquarters.

AWS VPN CloudHub:



via - https://docs.aws.amazon.com/vpn/latest/s2svpn/VPN_CloudHub.html

Incorrect options:

VPC Endpoint - 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. AWS PrivateLink simplifies the security of data shared with cloud-based applications by eliminating the exposure of data to the public Internet. When you use VPC endpoint, the traffic between your VPC and the other AWS service does not leave the Amazon network, therefore this option cannot be used to send and receive data between the remote branch offices of the company.

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. VPC peering facilitates a connection between two VPCs within the AWS network, therefore this option cannot be used to send and receive data between the remote branch offices of the company.

Software VPN - Amazon VPC offers you the flexibility to fully manage both sides of your Amazon VPC connectivity by creating a VPN connection between your remote network and a software VPN appliance running in your Amazon VPC network. Since Software VPN just handles

connectivity between the remote network and Amazon VPC, therefore it cannot be used to send and receive data between the remote branch offices of the company.

References:

https://docs.aws.amazon.com/whitepapers/latest/aws-vpc-connectivity-options/aws-vpn-cloudhub-network-to-amazon.html

https://docs.aws.amazon.com/vpn/latest/s2svpn/VPN_CloudHub.html

Domain

Design Secure Architectures

Question 5Skipped

The business analytics team at a company has been running ad-hoc queries on Oracle and PostgreSQL services on Amazon RDS to prepare daily reports for senior management. To facilitate the business analytics reporting, the engineering team now wants to continuously replicate this data and consolidate these databases into a petabyte-scale data warehouse by streaming data to Amazon Redshift.

As a solutions architect, which of the following would you recommend as the MOST resource-efficient solution that requires the LEAST amount of development time without the need to manage the underlying infrastructure?

Correct answer

Use AWS Database Migration Service (AWS DMS) to replicate the data from the databases into Amazon Redshift

Use AWS Glue to replicate the data from the databases into Amazon Redshift

Use AWS EMR to replicate the data from the databases into Amazon Redshift

Use Amazon Kinesis Data Streams to replicate the data from the databases into Amazon Redshift

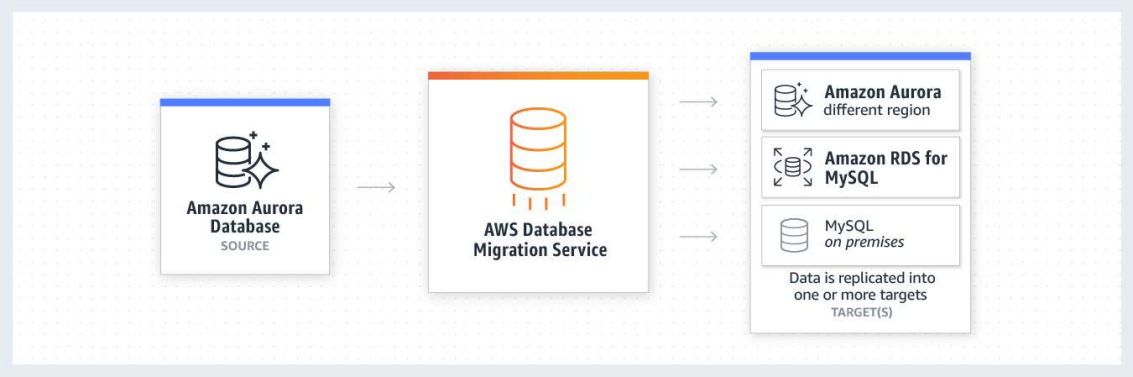
Overall explanation

Correct option:

Use AWS Database Migration Service (AWS DMS) to replicate the data from the databases into Amazon Redshift

AWS Database Migration Service helps you migrate databases to AWS quickly and securely. The source database remains fully operational during the migration, minimizing downtime to applications that rely on the database. With AWS Database Migration Service, you can continuously replicate your data with high availability and consolidate databases into a petabyte-scale data warehouse by streaming data to Amazon Redshift and Amazon S3.

Continuous Data Replication



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

You can migrate data to Amazon Redshift databases using AWS Database Migration Service. Amazon Redshift is a fully managed, petabyte-scale data warehouse service in the cloud. With an Amazon Redshift database as a target, you can migrate data from all of the other supported source databases.

The Amazon Redshift cluster must be in the same AWS account and the same AWS Region as the replication instance. During a database migration to Amazon Redshift, AWS DMS first moves data to an Amazon S3 bucket. When the files reside in an Amazon S3 bucket, AWS DMS then transfers them to the proper tables in the Amazon Redshift data warehouse. AWS DMS creates the S3 bucket in the same AWS Region as the Amazon Redshift database. The AWS DMS replication instance must be located in that same region.

Incorrect options:

Use AWS Glue to replicate the data from the databases into Amazon Redshift - 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.

Using AWS Glue involves significant development efforts to write custom migration scripts to copy the database data into Redshift.

Use AWS EMR to replicate the data from the databases into Amazon Redshift - Amazon EMR is the industry-leading cloud big data platform for processing vast amounts of data using open source tools such as Apache Spark, Apache Hive, Apache HBase, Apache Flink, Apache Hudi, and Presto. With EMR you can run Petabyte-scale analysis at less than half of the cost of traditional on-premises solutions and over 3x faster than standard Apache Spark. For short-running jobs, you can spin up and spin down clusters and pay per second for the instances used. For long-running workloads, you can create highly available clusters that automatically scale to meet demand. Amazon EMR uses Hadoop, an open-source framework, to distribute your data and processing across a resizable cluster of Amazon EC2 instances.

Using EMR involves significant infrastructure management efforts to set up and maintain the EMR cluster. Additionally this option involves a major development effort to write custom migration jobs to copy the database data into Redshift.

Use Amazon Kinesis Data Streams to replicate the data from the databases into Amazon Redshift - 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.

However, the user is expected to manually provision an appropriate number of shards to process the expected volume of the incoming data stream. The throughput of an Amazon Kinesis data stream is designed to scale without limits via increasing the number of shards within a data stream. Therefore Kinesis Data Streams is not the right fit for this use-case.

References:

https://docs.aws.amazon.com/dms/latest/userguide/CHAP_Target.Redshift.html

https://aws.amazon.com/dms/

Domain

Design Resilient Architectures

Question 6Skipped

A small business has been running its IT systems on the on-premises infrastructure but the business now plans to migrate to AWS Cloud for operational efficiencies.

As a Solutions Architect, can you suggest a cost-effective serverless solution for its flagship application that has both static and dynamic content?

Host the static content on Amazon S3 and use Amazon EC2 with Amazon RDS for generating the dynamic content. Amazon CloudFront can be configured in front of Amazon EC2 instance, to make global distribution easy

Host both the static and dynamic content of the web application on Amazon EC2 with Amazon RDS as database. Amazon CloudFront should be configured to distribute the content across geographically disperse regions

Correct answer

Host the static content on Amazon S3 and use AWS Lambda with Amazon DynamoDB for the serverless web application that handles dynamic content. Amazon CloudFront will sit in front of AWS Lambda for distribution across diverse regions

Host both the static and dynamic content of the web application on Amazon S3 and use Amazon CloudFront for distribution across diverse regions/countries

Overall explanation

Correct option:

Host the static content on Amazon S3 and use AWS Lambda with Amazon DynamoDB for the serverless web application that handles dynamic content. Amazon CloudFront will sit in front of AWS Lambda for distribution across diverse regions

AWS Lambda with Amazon DynamoDB is the right answer for a serverless solution. Amazon CloudFront will help in enhancing user experience by delivering content, across different geographic locations with low latency. Amazon S3 is a cost-effective and faster way of distributing static content for web applications.

Incorrect options:

Host both the static and dynamic content of the web application on Amazon S3 and use Amazon CloudFront for distribution across diverse regions/countries - Amazon S3 is not the right fit for hosting Dynamic content, so this option is incorrect.

Host the static content on Amazon S3 and use Amazon EC2 with Amazon RDS for generating the dynamic content. Amazon CloudFront can be configured in front of Amazon EC2 instance, to make global distribution easy - The company is looking for a serverless solution, and Amazon EC2 is not a serverless service as the Amazon EC2 instances have to be managed by AWS customers.

Host both the static and dynamic content of the web application on Amazon EC2 with Amazon RDS as database. Amazon CloudFront should be configured to distribute the content across geographically disperse regions - This is a possible solution, but not a cost-effective or optimal one. Since static content can be cost-effectively managed on Amazon S3 and can be accessed and distributed faster when compared to fetching the content from the Amazon EC2 server.

Reference:

https://aws.amazon.com/blogs/networking-and-content-delivery/deliver-your-apps-dynamic-content-using-amazon-cloudfront-getting-started-template/

Domain

Design High-Performing Architectures

Question 7Skipped

The engineering team at a company is moving the static content from the company's logistics website hosted on Amazon EC2 instances to an Amazon S3 bucket. The team wants to use an Amazon CloudFront distribution to deliver the static content. The security group used by the Amazon EC2 instances allows the website to be accessed by a limited set of IP ranges from the company's suppliers. Post-migration to Amazon CloudFront, access to the static content should only be allowed from the aforementioned IP addresses.

Which options would you combine to build a solution to meet these requirements? (Select two)

Correct selection

Create an AWS WAF ACL and use an IP match condition to allow traffic only from those IPs that are allowed in the Amazon EC2 security group. Associate this new AWS WAF ACL with the Amazon CloudFront distribution

Correct selection

Configure an origin access identity (OAI) and associate it with the Amazon CloudFront distribution. Set up the permissions in the Amazon S3 bucket policy so that only the OAI can read the objects

Create a new NACL that allows traffic from the same IPs as specified in the current Amazon EC2 security group. Associate this new NACL with the Amazon CloudFront distribution

Create an AWS Web Application Firewall (AWS WAF) ACL and use an IP match condition to allow traffic only from those IPs that are allowed in the Amazon EC2 security group. Associate this new AWS WAF ACL with the Amazon S3 bucket policy

Create a new security group that allows traffic from the same IPs as specified in the current Amazon EC2 security group. Associate this new security group with the Amazon CloudFront distribution

Overall explanation

Correct options:

Configure an origin access identity (OAI) and associate it with the Amazon CloudFront distribution. Set up the permissions in the Amazon S3 bucket policy so that only the OAI can read the objects

When you use Amazon CloudFront with an Amazon S3 bucket as the origin, you can configure Amazon CloudFront and Amazon S3 in a way that provides the following benefits:

Restricts access to the Amazon S3 bucket so that it's not publicly accessible

Makes sure that viewers (users) can access the content in the bucket only through the specified Amazon CloudFront distribution—that is, prevents them from accessing the content directly from the bucket, or through an unintended CloudFront distribution.

To do this, configure Amazon CloudFront to send authenticated requests to Amazon S3, and configure Amazon S3 to only allow access to authenticated requests from Amazon CloudFront. Amazon CloudFront provides two ways to send authenticated requests to an Amazon S3 origin: origin access control (OAC) and origin access identity (OAI).

Exam Alert:

Please note that AWS recommends using OAC because it supports:

All Amazon S3 buckets in all AWS Regions, including opt-in Regions launched after December 2022

Amazon S3 server-side encryption with AWS KMS (SSE-KMS)

Dynamic requests (POST, PUT, etc.) to Amazon S3

OAI doesn't work for the scenarios in the preceding list, or it requires extra workarounds in those scenarios. However, you will continue to see answers enlisting OAI as the preferred option in the actual exam as it takes about 6 months/1 year for a new feature to appear in the exam.

Create an AWS WAF ACL and use an IP match condition to allow traffic only from those IPs that are allowed in the Amazon EC2 security group. Associate this new AWS WAF ACL with the Amazon CloudFront distribution

AWS WAF is a web application firewall that lets you monitor the HTTP and HTTPS requests that are forwarded to your protected web application resources. You can protect the following resource types:

Amazon CloudFront distribution

Amazon API Gateway REST API

Application Load Balancer

AWS AppSync GraphQL API

Amazon Cognito user pool

AWS WAF also lets you control access to your content. Based on conditions that you specify, such as the IP addresses that requests originate from or the values of query strings, your protected resource responds to requests either with the requested content, with an HTTP 403 status code (Forbidden), or with a custom response.

If you want to allow or block web requests based on the IP addresses that the requests originate from, create one or more IP match conditions via your AWS WAF. An IP match condition lists up to 10,000 IP addresses or IP address ranges that your requests originate from.

For the given use case, you should add those IP addresses that are allowed in the Amazon EC2 security group into the IP match condition.

Incorrect options:

Create an AWS Web Application Firewall (AWS WAF) ACL and use an IP match condition to allow traffic only from those IPs that are allowed in the Amazon EC2 security group. Associate this new AWS WAF ACL with the Amazon S3 bucket policy - You cannot associate an AWS WAF ACL with an Amazon S3 bucket policy.

Create a new NACL that allows traffic from the same IPs as specified in the current Amazon EC2 security group. Associate this new NACL with the Amazon CloudFront distribution - NACL is associated with a subnet within a VPC. Amazon CloudFront delivers your content through a worldwide network of data centers called edge locations. So a NACL cannot be associated with a Amazon CloudFront distribution.

Create a new security group that allows traffic from the same IPs as specified in the current Amazon EC2 security group. Associate this new security group with the Amazon CloudFront distribution - A security group acts as a virtual firewall for your Amazon EC2 instances to control incoming and outgoing traffic. Inbound rules control the incoming traffic to your instance, and outbound rules control the outgoing traffic from your instance. Amazon CloudFront delivers your content through a worldwide network of data centers called edge locations. So a security group cannot be associated with Amazon CloudFront distribution.

References:

https://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/private-content-restricting-access-to-s3.html

https://docs.aws.amazon.com/waf/latest/developerguide/what-is-aws-waf.html

https://docs.aws.amazon.com/waf/latest/developerguide/classic-web-acl-ip-conditions.html

Domain

Design Secure Architectures

Question 8Skipped

A company wants to improve its gaming application by adding a leaderboard that uses a complex proprietary algorithm based on the participating user's performance metrics to identify the top users on a real-time basis. The technical requirements mandate high elasticity, low latency, and real-time processing to deliver customizable user data for the community of users. The leaderboard would be accessed by millions of users simultaneously.

Which of the following options support the case for using Amazon ElastiCache to meet the given requirements? (Select two)

Use Amazon ElastiCache to improve the performance of Extract-Transform-Load (ETL) workloads

Correct selection

Use Amazon ElastiCache to improve latency and throughput for read-heavy application workloads

Correct selection

Use Amazon ElastiCache to improve the performance of compute-intensive workloads

Use Amazon ElastiCache to improve latency and throughput for write-heavy application workloads

Use Amazon ElastiCache to run highly complex JOIN queries

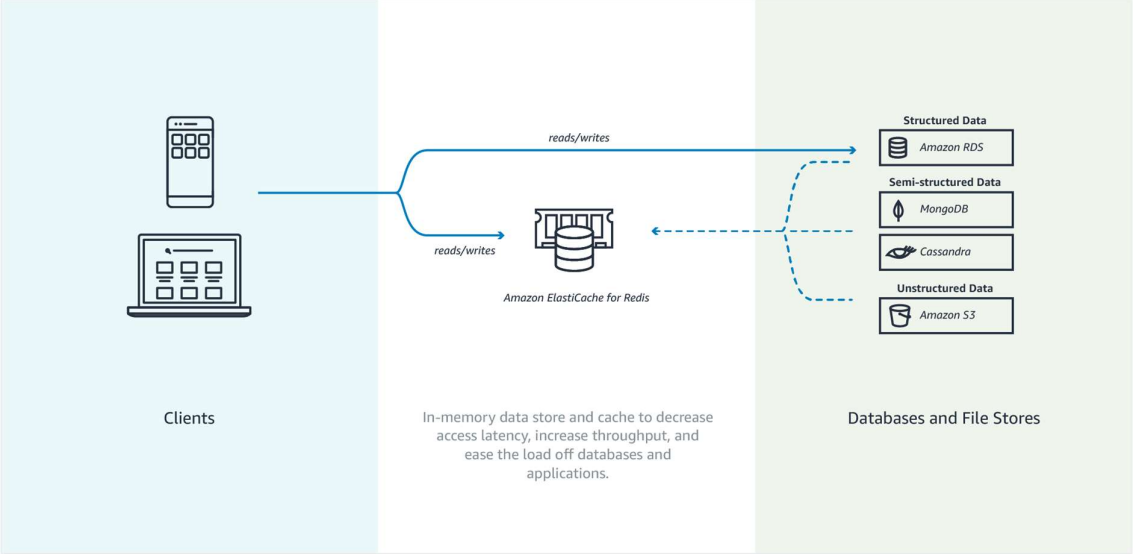
Overall explanation

Correct option:

Use Amazon ElastiCache to improve latency and throughput for read-heavy application workloads

Use Amazon ElastiCache to improve the performance of compute-intensive workloads

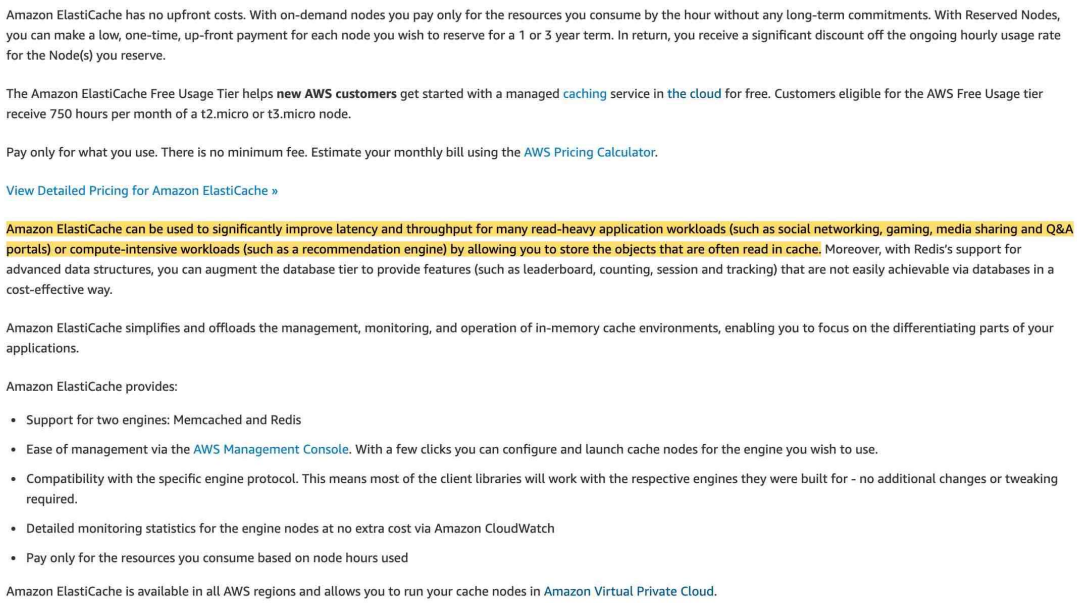
Amazon ElastiCache allows you to run in-memory data stores in the AWS cloud. Amazon ElastiCache is a popular choice for real-time use cases like Caching, Session Stores, Gaming, Geospatial Services, Real-Time Analytics, and Queuing.



via - https://docs.aws.amazon.com/AmazonElastiCache/latest/mem-ug/elasticache-use-cases.html

Amazon ElastiCache can be used to significantly improve latency and throughput for many read-heavy application workloads (such as social networking, gaming, media sharing, leaderboard, and Q&A portals) or compute-intensive workloads (such as a recommendation engine) by allowing you to store the objects that are often read in the cache.

Overview of Amazon ElastiCache features:



Amazon ElastiCache has no upfront costs. With on-demand nodes you pay only for the resources you consume by the hour without any long-term commitments. With Reserved Nodes, you can make a low, one-time, up-front payment for each node you wish to reserve for a 1 or 3 year term. In return, you receive a significant discount off the ongoing hourly usage rate for the Node(s) you reserve.

The Amazon ElastiCache Free Usage Tier helps **new AWS customers** get started with a managed caching service in the cloud for free. Customers eligible for the AWS Free Usage tier receive 750 hours per month of a t2.micro or t3.micro node.

Pay only for what you use. There is no minimum fee. Estimate your monthly bill using the AWS Pricing Calculator.

View Detailed Pricing for Amazon ElastiCache »

Amazon ElastiCache can be used to significantly improve latency and throughput for many read-heavy application workloads (such as social networking, gaming, media sharing and Q&A portals) or compute-intensive workloads (such as a recommendation engine) by allowing you to store the objects that are often read in cache. Moreover, with Redis's support for advanced data structures, you can augment the database tier to provide features (such as leaderboard, counting, session and tracking) that are not easily achievable via databases in a cost-effective way.

Amazon ElastiCache simplifies and offloads the management, monitoring, and operation of in-memory cache environments, enabling you to focus on the differentiating parts of your applications.

Amazon ElastiCache provides:

- Support for two engines: Memcached and Redis
- Ease of management via the AWS Management Console. With a few clicks you can configure and launch cache nodes for the engine you wish to use.
- Compatibility with the specific engine protocol. This means most of the client libraries will work with the respective engines they were built for - no additional changes or tweaking required.
- Detailed monitoring statistics for the engine nodes at no extra cost via Amazon CloudWatch
- Pay only for the resources you consume based on node hours used

Amazon ElastiCache is available in all AWS regions and allows you to run your cache nodes in Amazon Virtual Private Cloud.

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

Incorrect options:

Use Amazon ElastiCache to improve latency and throughput for write-heavy application workloads - As mentioned earlier in the explanation, Amazon ElastiCache can be used to significantly improve latency and throughput for many read-heavy application workloads. Caching is not a good fit for write-heavy applications as the cache goes stale at a very fast rate.

Use Amazon ElastiCache to improve the performance of Extract-Transform-Load (ETL) workloads - ETL workloads involve reading and transforming high-volume data which is not a good fit for caching. You should use AWS Glue or Amazon EMR to facilitate ETL workloads.

Use Amazon ElastiCache to run highly complex JOIN queries - Complex JSON queries can be run on relational databases such as Amazon RDS or Amazon Aurora. Amazon ElastiCache is not a good fit for this use case.

References:

https://docs.aws.amazon.com/AmazonElastiCache/latest/mem-ug/elasticache-use-cases.html

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

Domain

Design High-Performing Architectures

Question 9Skipped

A startup has recently moved their monolithic web application to AWS Cloud. The application runs on a single Amazon EC2 instance. Currently, the user base is small and the startup does not want to spend effort on elaborate disaster recovery strategies or Auto Scaling Group. The application can afford a maximum downtime of 10 minutes.

In case of a failure, which of these options would you suggest as a cost-effective and automatic recovery procedure for the instance?

Configure AWS Trusted Advisor to monitor the health check of Amazon EC2 instance and provide a remedial action in case an unhealthy flag is detected

Configure Amazon EventBridge events that can trigger the recovery of the Amazon EC2 instance, in case the instance or the application fails

Configure an Amazon CloudWatch alarm that triggers the recovery of the Amazon EC2 instance, in case the instance fails. The instance can be configured with Amazon Elastic Block Store (Amazon EBS) or with instance store volumes

Correct answer

Configure an Amazon CloudWatch alarm that triggers the recovery of the Amazon EC2 instance, in case the instance fails. The instance, however, should only be configured with an Amazon EBS volume

Overall explanation

Correct option:

Configure an Amazon CloudWatch alarm that triggers the recovery of the Amazon EC2 instance, in case the instance fails. The instance, however, should only be configured with an Amazon EBS volume

If your instance fails a system status check, you can use Amazon CloudWatch alarm actions to automatically recover it. The recover option is available for over 90% of deployed customer Amazon EC2 instances. The Amazon CloudWatch recovery option works only for system check failures, not for instance status check failures. Also, if you terminate your instance, then it can't be recovered.

You can create an Amazon CloudWatch alarm that monitors an Amazon EC2 instance and automatically recovers the instance if it becomes impaired due to an underlying hardware failure or a problem that requires AWS involvement to repair. Terminated instances cannot be recovered. A recovered instance is identical to the original instance, including the instance ID, private IP addresses, Elastic IP addresses, and all instance metadata. If the impaired instance is in a placement group, the recovered instance runs in the placement group.

The automatic recovery process attempts to recover your instance for up to three separate failures per day. Your instance may subsequently be retired if automatic recovery fails and a hardware degradation is determined to be the root cause for the original system status check failure.

Incorrect options:

Configure Amazon EventBridge events that can trigger the recovery of the Amazon EC2 instance, in case the instance or the application fails - You cannot use Amazon EventBridge events to directly trigger the recovery of the Amazon EC2 instance.

Configure an Amazon CloudWatch alarm that triggers the recovery of the Amazon EC2 instance, in case the instance fails. The instance can be configured with Amazon Elastic Block Store (Amazon EBS) or with instance store volumes - The recover action is supported only on instances that have Amazon EBS volumes configured on them, instance store volumes are not supported for automatic recovery by Amazon CloudWatch alarms.

Configure AWS Trusted Advisor to monitor the health check of Amazon EC2 instance and provide a remedial action in case an unhealthy flag is detected - You can use Amazon EventBridge events to detect and react to changes in the status of AWS Trusted Advisor checks. This support is only available with AWS Business Support and AWS Enterprise Support. AWS Trusted Advisor by itself does not support health checks of Amazon EC2 instances or their recovery.

Reference:

https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ec2-instance-recover.html

Domain

Design Resilient Architectures

Question 10Skipped

An online gaming application has a large chunk of its traffic coming from users who download static assets such as historic leaderboard reports and the game tactics for various games. The current infrastructure and design are unable to cope up with the traffic and application freezes on most of the pages.

Which of the following is a cost-optimal solution that does not need provisioning of infrastructure?

Use Amazon CloudFront with Amazon DynamoDB for greater speed and low latency access to static assets

Configure AWS Lambda with an Amazon RDS database to provide a serverless architecture

Correct answer

Use Amazon CloudFront with Amazon S3 as the storage solution for the static assets

Use AWS Lambda with Amazon ElastiCache and Amazon RDS for serving static assets at high speed and low latency

Overall explanation

Correct option:

Use Amazon CloudFront with Amazon S3 as the storage solution for the static assets

When you put your content in an Amazon S3 bucket in the cloud, a lot of things become much easier. First, you don't need to plan for and allocate a specific amount of storage space because Amazon S3 buckets scale automatically. As Amazon S3 is a serverless service, you don't need to manage or patch servers that store files yourself; you just put and get your content. Finally, even if you require a server for your application (for example, because you have a dynamic application), the server can be smaller because it doesn't have to handle requests for static content.

Amazon CloudFront is a content delivery network (CDN) service that delivers static and dynamic web content, video streams, and APIs around the world, securely and at scale. By design, delivering data out of Amazon CloudFront can be more cost-effective than delivering it from Amazon S3 directly to your users. Amazon CloudFront serves content through a worldwide network of data centers called Edge Locations. Using edge servers to cache and serve content improves performance by providing content closer to where viewers are located.

When a user requests content that you serve with Amazon CloudFront, their request is routed to a nearby Edge Location. If Amazon CloudFront has a cached copy of the requested file, CloudFront delivers it to the user, providing a fast (low-latency) response. If the file they've requested isn't yet cached, CloudFront retrieves it from your origin – for example, the Amazon S3 bucket where you've stored your content. Then, for the next local request for the same content, it's already cached nearby and can be served immediately.

By caching your content in Edge Locations, Amazon CloudFront reduces the load on your Amazon S3 bucket and helps ensure a faster response for your users when they request content. Also, data transfer out for content by using Amazon CloudFront is often more cost-effective than serving files directly from Amazon S3, and there is no data transfer fee from Amazon S3 to Amazon CloudFront. You only pay for what is delivered to the internet from Amazon CloudFront, plus request fees.

Incorrect options:

Configure AWS Lambda with an Amazon RDS database to provide a serverless architecture - Amazon RDS is not the right choice for the current scenario because of the overhead of a database management system, as the given use-case can be addressed by using Amazon S3 storage solution.

Use Amazon CloudFront with Amazon DynamoDB for greater speed and low latency access to static assets - Amazon DynamoDB is a key-value and document database that delivers single-digit millisecond performance at any scale. But, Amazon DynamoDB is overkill for the given use-case and will prove to be a very costly solution.

Use AWS Lambda with Amazon ElastiCache and Amazon RDS for serving static assets at high speed and low latency - As discussed above, Amazon RDS is not needed for this use case where web application needs to display static pages and facilitate downloads of historic data. Amazon S3 is much better suited for this requirement.

Reference:

https://aws.amazon.com/blogs/networking-and-content-delivery/amazon-s3-amazon-cloudfront-a-match-made-in-the-cloud/

Domain

Design High-Performing Architectures

Question 11Skipped

The DevOps team at an IT company has recently migrated to AWS and they are configuring security groups for their two-tier application with public web servers and private database servers. The team wants to understand the allowed configuration options for an inbound rule for a security group.

As a solutions architect, which of the following would you identify as an INVALID option for setting up such a configuration?

You can use a security group as the custom source for the inbound rule

Correct answer

You can use an Internet Gateway ID as the custom source for the inbound rule

You can use a range of IP addresses in CIDR block notation as the custom source for the inbound rule

You can use an IP address as the custom source for the inbound rule

Overall explanation

Correct option:

You can use an Internet Gateway ID as the custom source for the inbound rule

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, you can 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.

Please see this list of allowed source or destination for security group rules:

Source or destination: The source (inbound rules) or destination (outbound rules) for the traffic. Specify one of these options:

- An individual IPv4 address. You must use the `/32` prefix length; for example, `203.0.113.1/32`.
- An individual IPv6 address. You must use the `/128` prefix length; for example, `2001:db8:1234:1a00::123/128`.
- A range of IPv4 addresses, in CIDR block notation; for example, `203.0.113.0/24`.
- A range of IPv6 addresses, in CIDR block notation; for example, `2001:db8:1234:1a00::/64`.
- The prefix list ID for the AWS service; for example, `pl-1a2b3c4d`. For more information, see Gateway VPC Endpoints in the *Amazon VPC User Guide*.
- Another security group. This allows instances that are associated with the specified security group to access instances associated with this security group. Choosing this option does not add rules from the source security group to this security group. You can specify one of the following security groups:
  - The current security group
  - A different security group for the same VPC
  - A different security group for a peer VPC in a VPC peering connection

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

Therefore, you cannot use an Internet Gateway ID as the custom source for the inbound rule.

Incorrect options:

You can use a security group as the custom source for the inbound rule

You can use a range of IP addresses in CIDR block notation as the custom source for the inbound rule

You can use an IP address as the custom source for the inbound rule

As described in the list of allowed sources or destinations for security group rules, the above options are supported.

Reference:

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

Domain

Design Secure Architectures

Question 12Skipped

The DevOps team at an IT company has created a custom VPC (V1) and attached an Internet Gateway (I1) to the VPC. The team has also created a subnet (S1) in this custom VPC and added a route to this subnet's route table (R1) that directs internet-bound traffic to the Internet Gateway. Now the team launches an Amazon EC2 instance (E1) in the subnet S1 and assigns a public IPv4 address to this instance. Next the team also launches a Network Address Translation (NAT) instance (N1) in the subnet S1.

Under the given infrastructure setup, which of the following entities is doing the Network Address Translation for the Amazon EC2 instance E1?

Correct answer

Internet Gateway (I1)

Subnet (S1)

Network Address Translation (NAT) instance (N1)

Route Table (R1)

Overall explanation

Correct option:

Internet Gateway (I1)

An Internet Gateway is a horizontally scaled, redundant, and highly available VPC component that allows communication between 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. Therefore, for instance E1, the Network Address Translation is done by Internet Gateway I1.

Additionally, an Internet Gateway supports IPv4 and IPv6 traffic. It does not cause availability risks or bandwidth constraints on your network traffic.

To enable access to or from the internet for instances in a subnet in a VPC, you must do the following:

Attach an Internet gateway to your VPC.

Add a route to your subnet's route table that directs internet-bound traffic to the internet gateway. If a subnet is associated with a route table that has a route to an internet gateway, it's known as a public subnet. If a subnet is associated with a route table that does not have a route to an internet gateway, it's known as a private subnet.

Ensure that instances in your subnet have a globally unique IP address (public IPv4 address, Elastic IP address, or IPv6 address).

Ensure that your network access control lists and security group rules allow the relevant traffic to flow to and from your instance.

Internet Gateway Overview:

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

Incorrect options:

Network Address Translation (NAT) instance (N1) - You can use a network address translation (NAT) instance in a public subnet in your VPC to enable instances in the private subnet to initiate outbound IPv4 traffic to the Internet or other AWS services, but prevent the instances from receiving inbound traffic initiated by someone on the Internet. As the instance E1 is in a public subnet, therefore this option is not correct.

Subnet (S1)

Route Table (R1)

A virtual private cloud (VPC) is a virtual network dedicated to your AWS account. A subnet is a range of IP addresses in your VPC. A route table contains a set of rules, called routes, that are used to determine where network traffic is directed. Therefore neither Subnet nor Route Table can be used for Network Address Translation.

References:

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

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

Domain

Design Resilient Architectures

Question 13Skipped

An e-commerce company uses Microsoft Active Directory to provide users and groups with access to resources on the on-premises infrastructure. The company has extended its IT infrastructure to AWS in the form of a hybrid cloud. The engineering team at the company wants to run directory-aware workloads on AWS for a SQL Server-based application. The team also wants to configure a trust relationship to enable single sign-on (SSO) for its users to access resources in either domain.

As a solutions architect, which of the following AWS services would you recommend for this use-case?

Active Directory Connector

Simple Active Directory (Simple AD)

Correct answer

AWS Directory Service for Microsoft Active Directory (AWS Managed Microsoft AD)

Amazon Cloud Directory

Overall explanation

Correct option:

AWS Directory Service for Microsoft Active Directory (AWS Managed Microsoft AD)

AWS Directory Service provides multiple ways to use Amazon Cloud Directory and Microsoft Active Directory (AD) with other AWS services.

AWS Directory Service for Microsoft Active Directory (aka AWS Managed Microsoft AD) is powered by an actual Microsoft Windows Server Active Directory (AD), managed by AWS. With AWS Managed Microsoft AD, you can run directory-aware workloads in the AWS Cloud such as SQL Server-based applications. You can also configure a trust relationship between AWS Managed Microsoft AD in the AWS Cloud and your existing on-premises Microsoft Active Directory, providing users and groups with access to resources in either domain, using single sign-on (SSO).

Incorrect options:

Active Directory Connector - Use AD Connector if you only need to allow your on-premises users to log in to AWS applications and services with their Active Directory credentials. AD Connector simply connects your existing on-premises Active Directory to AWS. You cannot use it to run directory-aware workloads on AWS, hence this option is not correct.

Simple Active Directory (Simple AD) - Simple AD provides a subset of the features offered by AWS Managed Microsoft AD. Simple AD is a standalone managed directory that is powered by a Samba 4 Active Directory Compatible Server. Simple AD does not support features such as trust relationships with other domains. Therefore, this option is not correct.

Amazon Cloud Directory - Amazon Cloud Directory is a cloud-native directory that can store hundreds of millions of application-specific objects with multiple relationships and schemas. Use Amazon Cloud Directory if you need a highly scalable directory store for your application's hierarchical data. You cannot use it to establish trust relationships with other domains on the on-premises infrastructure. Therefore, this option is not correct.

Exam Alert:

You may see questions on choosing "AWS Managed Microsoft AD" vs "AD Connector" vs "Simple AD" on the exam. Just remember that you should use AD Connector if you only need to allow your on-premises users to log in to AWS applications with their Active Directory credentials. AWS Managed Microsoft AD would also allow you to run directory-aware workloads in the AWS Cloud. AWS Managed Microsoft AD is your best choice if you have more than 5,000 users and need a trust relationship set up between an AWS hosted directory and your on-premises directories. Simple AD is the least expensive option and your best choice if you have 5,000 or fewer users and don't need the more advanced Microsoft Active Directory features such as trust relationships with other domains.

Reference:

https://docs.aws.amazon.com/directoryservice/latest/admin-guide/what_is.html

Domain

Design High-Performing Architectures

Question 14Skipped

A global manufacturing company with facilities in the US, Europe, and Asia is designing a new distributed application to optimize its procurement workflow. The orders booked in one AWS Region should be visible to all AWS Regions in a second or less. The database should be able to facilitate failover with a short Recovery Time Objective (RTO). The uptime of the application is critical to ensure that the manufacturing processes are not impacted.

As a solutions architect, which of the following will you recommend as the MOST cost-effective solution?

Provision Amazon RDS for MySQL with a cross-Region read replica

Provision Amazon DynamoDB global tables

Correct answer

Provision Amazon Aurora Global Database

Provision Amazon RDS for PostgreSQL with a cross-Region read replica

Overall explanation

Correct option:

Provision Amazon Aurora Global Database

An Aurora global database provides more comprehensive failover capabilities than the failover provided by a default Aurora DB cluster. By using an Aurora global database, you can plan for and recover from disaster fairly quickly. Recovery from disaster is typically measured using values for RTO and RPO.

Recovery time objective (RTO) – The time it takes a system to return to a working state after a disaster. In other words, RTO measures downtime. For an Aurora global database, RTO can be in the order of minutes.

Recovery point objective (RPO) – The amount of data that can be lost (measured in time). For an Aurora global database, RPO is typically measured in seconds.

With an Aurora global database, you can choose from two different approaches to failover:

1. Managed planned failover – This feature is intended for controlled environments, such as disaster recovery (DR) testing scenarios, operational maintenance, and other planned operational procedures. Managed planned failover allows you to relocate the primary DB cluster of your Aurora global database to one of the secondary Regions. Because this feature synchronizes secondary DB clusters with the primary before making any other changes, RPO is 0 (no data loss).

2. Unplanned failover ("detach and promote") – To recover from an unplanned outage, you can perform a cross-Region failover to one of the secondaries in your Aurora global database. The RTO for this manual process depends on how quickly you can perform the tasks listed in Recovering an Amazon Aurora global database from an unplanned outage. The RPO is typically measured in seconds, but this depends on the Aurora storage replication lag across the network at the time of the failure.

Disaster Recovery in Aurora Global Databases:

Disaster recovery and Amazon Aurora global databases

PDF | Kindle | RSS

An Aurora global database provides more comprehensive failover capabilities than the failover provided by a default Aurora DB cluster. By using an Aurora global database, you can plan for and recover from disaster fairly quickly. Recovery from disaster is typically measured using values for RTO and RPO.

- **Recovery time objective (RTO)** – The time it takes a system to return to a working state after a disaster. In other words, RTO measures downtime. For an Aurora global database, RTO can be in the order of minutes.
- **Recovery point objective (RPO)** – The amount of data that can be lost (measured in time). For an Aurora global database, RPO is typically measured in seconds. With an Aurora PostgreSQL–based global database, you can use the `rds.global_db_rpo` parameter to set and track the upper bound on RPO, but doing so might affect transaction processing on the primary cluster's writer node. For more information, see Managing RPOs for Aurora PostgreSQL–based global databases.

With an Aurora global database, you can choose from two different approaches to failover.

- **Managed planned failover** – This feature is intended for controlled environments, such as disaster recovery (DR) testing scenarios, operational maintenance, and other planned operational procedures. Managed planned failover allows you to relocate the primary DB cluster of your Aurora global database to one of the secondary Regions. Because this feature synchronizes secondary DB clusters with the primary before making any other changes, RPO is 0 (no data loss). RTO for this automated process is typically less than that of the "detach and promote" failover process because the demotion, promotion, and all synchronization are handled for you. To learn more, see Managed planned failover for Amazon Aurora global databases.
- **Unplanned failover** ("detach and promote") – To recover from an unplanned outage, you can perform a cross-Region failover to one of the secondaries in your Aurora global database. The RTO for this manual process depends on how quickly you can perform the tasks listed in Recovering an Amazon Aurora global database from an unplanned outage. The RPO is typically measured in seconds, but this depends on the Aurora storage replication lag across the network at the time of the failure.

via - https://docs.aws.amazon.com/AmazonRDS/latest/AuroraUserGuide/aurora-global-database-disaster-recovery.html

Incorrect options:

Provision Amazon RDS for MySQL with a cross-Region read replica

Provision Amazon RDS for PostgreSQL with a cross-Region read replica

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. For a failover, read replicas have to be manually promoted to a standalone database instance since the process is not automatic. Hence, the RTO will be quite high, so both these options are not correct for this use case.

Provision Amazon DynamoDB global tables - Aurora Global Database is good for applications that need to support cross-Region reads with low latency updates and the ability to quickly failover between regions. DynamoDB global tables provide cross-region active-active capabilities with high performance, but you lose some of the data access flexibility that comes with SQL-based databases. Due to the active-active configuration of DynamoDB global tables, there is no concept of failover because the application writes to the table in its region, and then the data is replicated to keep the other regions' table in sync. DynamoDB global tables is a much costlier solution than Aurora Global Database for the given requirement.

References:

https://docs.aws.amazon.com/AmazonRDS/latest/AuroraUserGuide/aurora-global-database-disaster-recovery.html#aurora-global-database-failover

https://aws.amazon.com/blogs/database/how-to-use-amazon-dynamodb-global-tables-to-power-multiregion-architectures/

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

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

https://docs.aws.amazon.com/AmazonRDS/latest/AuroraUserGuide/aurora-global-database.html

Domain

Design Resilient Architectures

Question 15Skipped

A developer has configured inbound traffic for the relevant ports in both the Security Group of the Amazon EC2 instance as well as the Network Access Control List (Network ACL) of the subnet for the Amazon EC2 instance. The developer is, however, unable to connect to the service running on the Amazon EC2 instance.

As a solutions architect, how will you fix this issue?

Network ACLs are stateful, so allowing inbound traffic to the necessary ports enables the connection. Security Groups are stateless, so you must allow both inbound and outbound traffic

Correct answer

Security Groups are stateful, so allowing inbound traffic to the necessary ports enables the connection. Network ACLs are stateless, so you must allow both inbound and outbound traffic

Rules associated with Network ACLs should never be modified from command line. An attempt to modify rules from command line blocks the rule and results in an erratic behavior

IAM Role defined in the Security Group is different from the IAM Role that is given access in the Network ACLs

Overall explanation

Correct option:

Security Groups are stateful, so allowing inbound traffic to the necessary ports enables the connection. Network ACLs are stateless, so you must allow both inbound and outbound traffic

Security groups are stateful, so allowing inbound traffic to the necessary ports enables the connection. Network ACLs are stateless, so you must allow both inbound and outbound traffic.

To enable the connection to a service running on an instance, the associated network ACL must allow both inbound traffic on the port that the service is listening on as well as allow outbound traffic from ephemeral ports. When a client connects to a server, a random port from the ephemeral port range (1024-65535) becomes the client's source port.

The designated ephemeral port then becomes the destination port for return traffic from the service, so outbound traffic from the ephemeral port must be allowed in the network ACL.

By default, network ACLs allow all inbound and outbound traffic. If your network ACL is more restrictive, then you need to explicitly allow traffic from the ephemeral port range.

If you accept traffic from the internet, then you also must establish a route through an internet gateway. If you accept traffic over VPN or AWS Direct Connect, then you must establish a route through a virtual private gateway (VGW).

Incorrect options:

Network ACLs are stateful, so allowing inbound traffic to the necessary ports enables the connection. Security Groups are stateless, so you must allow both inbound and outbound traffic - This is incorrect as already discussed.

IAM Role defined in the Security Group is different from the IAM Role that is given access in the Network ACLs - This is a made-up option and just added as a distractor.

Rules associated with Network ACLs should never be modified from command line. An attempt to modify rules from command line blocks the rule and results in an erratic behavior - This option is a distractor. AWS does not support modifying rules of Network ACLs from the command line tool.

Reference:

https://aws.amazon.com/premiumsupport/knowledge-center/resolve-connection-sg-acl-inbound/

Domain

Design Secure Architectures

Question 16Skipped

A global pharmaceutical company wants to move most of the on-premises data into Amazon S3, Amazon Elastic File System (Amazon EFS), and Amazon FSx for Windows File Server easily, quickly, and cost-effectively.

As a solutions architect, which of the following solutions would you recommend as the BEST fit to automate and accelerate online data transfers to these AWS storage services?

Use AWS Snowball Edge Storage Optimized device to automate and accelerate online data transfers to the given AWS storage services

Correct answer

Use AWS DataSync to automate and accelerate online data transfers to the given AWS storage services

Use File Gateway to automate and accelerate online data transfers to the given AWS storage services

Use AWS Transfer Family to automate and accelerate online data transfers to the given AWS storage services

Overall explanation

Correct option:

Use AWS DataSync to automate and accelerate online data transfers to the given AWS storage services

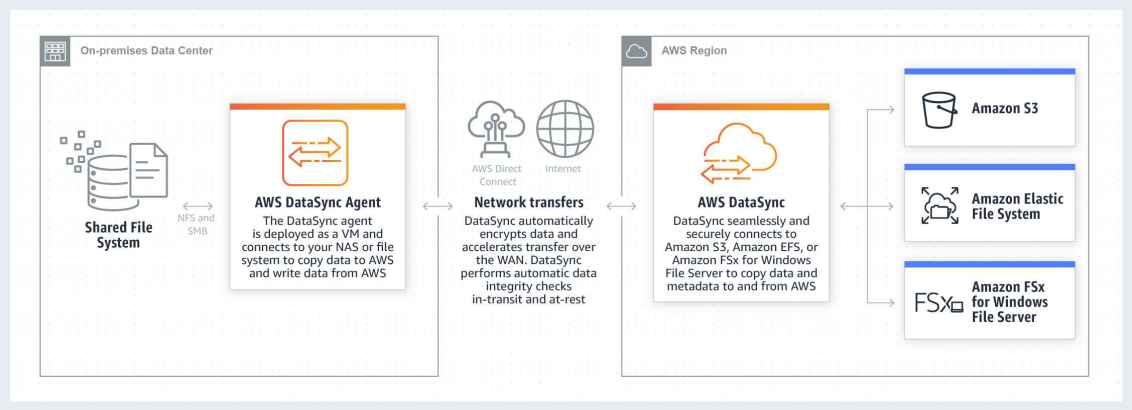
AWS DataSync is an online data transfer service that simplifies, automates, and accelerates copying large amounts of data to and from AWS storage services over the internet or AWS Direct Connect.

AWS DataSync fully automates and accelerates moving large active datasets to AWS, up to 10 times faster than command-line tools. It is natively integrated with Amazon S3, Amazon EFS, Amazon FSx for Windows File Server, Amazon CloudWatch, and AWS CloudTrail, which provides seamless and secure access to your storage services, as well as detailed monitoring of the transfer.

AWS DataSync uses a purpose-built network protocol and scale out architecture to transfer data. A single DataSync agent is capable of saturating a 10 Gbps network link.

AWS DataSync fully automates the data transfer. It comes with retry and network resiliency mechanisms, network optimizations, built-in task scheduling, monitoring via the DataSync API and Console, and Amazon CloudWatch metrics, events, and logs that provide granular visibility into the transfer process. AWS DataSync performs data integrity verification both during the transfer and at the end of the transfer.

How AWS DataSync Works:



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

Incorrect options:

Use AWS Snowball Edge Storage Optimized device to automate and accelerate online data transfers to the given AWS storage services - 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 TB of usable HDD storage, 40 vCPUs, 1 TB of SATA SSD storage, and up to 40 Gb network connectivity to address large scale data transfer and pre-processing use cases. As each Snowball Edge Storage Optimized device can handle 80TB of data, you can order 10 such devices to take care of the data transfer for all applications. The original Snowball devices were transitioned out of service and 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 Snowball device had 80TB of storage space.

AWS Snowball Edge is suitable for offline data transfers, for customers who are bandwidth constrained or transferring data from remote, disconnected, or austere environments. Therefore, it cannot support automated and accelerated online data transfers.

Use AWS Transfer Family to automate and accelerate online data transfers to the given AWS storage services - The AWS Transfer Family provides fully managed support for file transfers directly into and out of Amazon S3 and Amazon EFS. Therefore, it cannot support migration into the other AWS storage services mentioned in the given use-case (Amazon FSx for Windows File Server).

Use File Gateway to automate and accelerate online data transfers to the given AWS storage services - AWS Storage Gateway's file interface, or file gateway, offers you a seamless way to connect to the cloud to store application data files and backup images as durable objects on Amazon S3 cloud storage. File gateway offers SMB or NFS-based access to data in Amazon S3 with local caching. It can be used for on-premises applications, and for Amazon EC2-based applications that need file protocol access to S3 object storage. Therefore, it cannot support migration into the other AWS storage services mentioned in the given use-case (such as EFS and Amazon FSx for Windows File Server).

References:

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

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

https://aws.amazon.com/aws-transfer-family/

Domain

Design High-Performing Architectures

Question 17Skipped

An IT company is looking to move its on-premises infrastructure to AWS Cloud. The company has a portfolio of applications with a few of them using server bound licenses that are valid for the next year. To utilize the licenses, the CTO wants to use dedicated hosts for a one year term and then migrate the given instances to default tenancy thereafter.

As a solutions architect, which of the following options would you identify as CORRECT for changing the tenancy of an instance after you have launched it? (Select two)

You can change the tenancy of an instance from dedicated to default

Correct selection

You can change the tenancy of an instance from host to dedicated

You can change the tenancy of an instance from default to dedicated

Correct selection

You can change the tenancy of an instance from dedicated to host

You can change the tenancy of an instance from default to host

Overall explanation

Correct options:

You can change the tenancy of an instance from dedicated to host

You can change the tenancy of an instance from host to dedicated

Each Amazon EC2 instance that you launch into a VPC has a tenancy attribute. This attribute has the following values.

Tenancy Value	Description
`default`	Your instance runs on shared hardware.
`dedicated`	Your instance runs on single-tenant hardware.
`host`	Your instance runs on a Dedicated Host, which is an isolated server with configurations that you can control.

via - https://docs.aws.amazon.com/autoscaling/ec2/userguide/as-instance-monitoring.html

By default, Amazon EC2 instances run on a shared-tenancy basis.

Dedicated Instances are Amazon EC2 instances that run in a virtual private cloud (VPC) on hardware that's dedicated to a single customer. Dedicated Instances that belong to different AWS accounts are physically isolated at the hardware level. However, Dedicated Instances may share hardware with other instances from the same AWS account that is not Dedicated Instances.

A Dedicated Host is also a physical server that's dedicated to your use. With a Dedicated Host, you have visibility and control over how instances are placed on the server.

Incorrect options:

You can change the tenancy of an instance from default to dedicated - You can only change the tenancy of an instance from dedicated to host, or from host to dedicated after you've launched it. Therefore, this option is incorrect.

You can change the tenancy of an instance from dedicated to default - You can only change the tenancy of an instance from dedicated to host, or from host to dedicated after you've launched it. Therefore, this option is incorrect.

You can change the tenancy of an instance from default to host - You can only change the tenancy of an instance from dedicated to host, or from host to dedicated after you've launched it. Therefore, this option is incorrect.

Reference:

https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/dedicated-instance.html

Domain

Design Cost-Optimized Architectures

Question 18Skipped

An IT company is using Amazon Simple Queue Service (Amazon SQS) queues for decoupling the various components of application architecture. As the consuming components need additional time to process Amazon Simple Queue Service (Amazon SQS) messages, the company wants to postpone the delivery of new messages to the queue for a few seconds.

As a solutions architect, which of the following solutions would you suggest to the company?

Use visibility timeout to postpone the delivery of new messages to the queue for a few seconds

Use Amazon SQS FIFO queues to postpone the delivery of new messages to the queue for a few seconds

Correct answer

Use delay queues to postpone the delivery of new messages to the queue for a few seconds

Use dead-letter queues to postpone the delivery of new messages to the queue for a few seconds

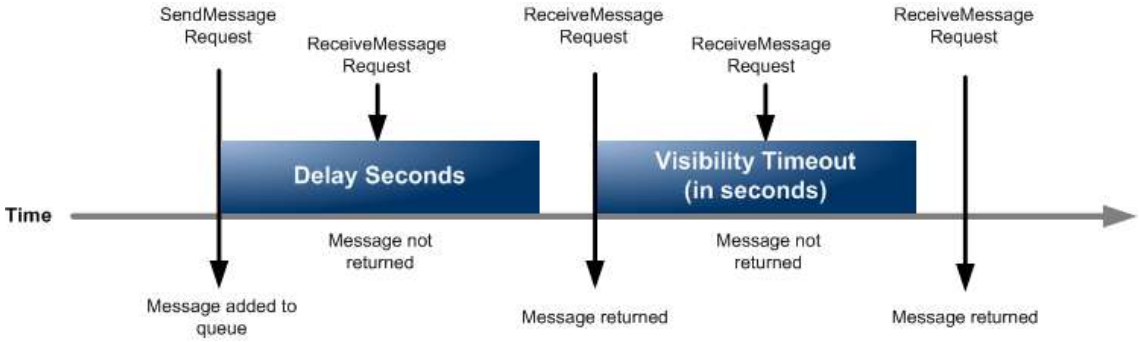
Overall explanation

Correct option:

Use delay queues to postpone the delivery of new messages to the queue for a few seconds

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.

Delay queues let you postpone the delivery of new messages to a queue for several seconds, for example, when your consumer application needs additional time to process messages. If you create a delay queue, any messages that you send to the queue remain invisible to consumers for the duration of the delay period. The default (minimum) delay for a queue is 0 seconds. The maximum is 15 minutes.



via - https://docs.aws.amazon.com/AWSSimpleQueueService/latest/SQSDeveloperGuide/sqs-delay-queues.html

Incorrect options:

Use Amazon SQS FIFO queues to postpone the delivery of new messages to the queue for a few seconds - 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. You cannot use Amazon SQS FIFO queues to postpone the delivery of new messages to the queue for a few seconds.

Use dead-letter queues to postpone the delivery of new messages to the queue for a few seconds - Dead-letter queues can be used by other queues (source queues) as a target for messages that can't be processed (consumed) successfully. Dead-letter queues are useful for debugging your application or messaging system because they let you isolate problematic messages to determine why their processing doesn't succeed. You cannot use dead-letter queues to postpone the delivery of new messages to the queue for a few seconds.

Use visibility timeout to postpone the delivery of new messages to the queue for a few seconds - Visibility timeout is a period during which Amazon SQS prevents other consumers from receiving and processing a given message. The default visibility timeout for a message is 30 seconds. The minimum is 0 seconds. The maximum is 12 hours. You cannot use visibility timeout to postpone the delivery of new messages to the queue for a few seconds.

Reference:

https://docs.aws.amazon.com/AWSSimpleQueueService/latest/SQSDeveloperGuide/sqs-delay-queues.html

Domain

Design Resilient Architectures

Question 19Skipped

A video conferencing application is hosted on a fleet of EC2 instances which are part of an Auto Scaling group. The Auto Scaling group uses a Launch Template (LT1) with "dedicated" instance tenancy but the VPC (V1) used by the Launch Template LT1 has the instance tenancy set to default. Later the DevOps team creates a new Launch Template (LT2) with shared (default) instance tenancy but the VPC (V2) used by the Launch Template LT2 has the instance tenancy set to dedicated.

Which of the following is correct regarding the instances launched via Launch Template LT1 and Launch Template LT2?

The instances launched by both Launch Template LT1 and Launch Template LT2 will have default instance tenancy

The instances launched by Launch Template LT1 will have default instance tenancy while the instances launched by the Launch Template LT2 will have dedicated instance tenancy

The instances launched by Launch Template LT1 will have dedicated instance tenancy while the instances launched by the Launch Template LT2 will have shared (default) instance tenancy

Correct answer

The instances launched by both Launch Template LT1 and Launch Template LT2 will have dedicated instance tenancy

Overall explanation

Correct option:

The instances launched by both Launch Template LT1 and Launch Template LT2 will have dedicated instance tenancy

A launch template specifies instance configuration information. It includes the ID of the Amazon Machine Image (AMI), the instance type, a key pair, security groups, and other parameters used to launch EC2 instances. If you've launched an EC2 instance before, you specified the same information to launch the instance.

When you create a Launch Template, the default value for the instance tenancy is shared and the instance tenancy is controlled by the tenancy attribute of the VPC. If you set the Launch Template Tenancy to shared (default) and the VPC Tenancy is set to dedicated, then the instances have dedicated tenancy. If you set the Launch Template Tenancy to dedicated and the VPC Tenancy is set to default, then again the instances have dedicated tenancy.

Amazon EC2 provides three options for the tenancy of your EC2 instances:

Shared (Shared) – Multiple AWS accounts may share the same physical hardware. This is the default tenancy option when launching an instance.

Dedicated instances (Dedicated) – Your instance runs on single-tenant hardware. No other AWS customer shares the same physical server.

Dedicated Hosts (Dedicated host) – The instance runs on a physical server that is dedicated to your use. Using Dedicated Hosts makes it easier to bring your own licenses (BYOL) that have dedicated hardware requirements to EC2 and meet compliance use cases. If you choose this option, you must provide a host resource group for Tenancy host resource group.

Incorrect options:

The instances launched by Launch Template LT1 will have dedicated instance tenancy while the instances launched by the Launch Template LT2 will have shared (default) instance tenancy - If either Launch Template Tenancy or VPC Tenancy is set to dedicated, then the instance tenancy is also dedicated. Therefore, this option is incorrect.

The instances launched by Launch Template LT1 will have default instance tenancy while the instances launched by the Launch Template LT2 will have dedicated instance tenancy - If either Launch Template Tenancy or VPC Tenancy is set to dedicated, then the instance tenancy is also dedicated. Therefore, this option is incorrect.

The instances launched by both Launch Template LT1 and Launch Template LT2 will have default instance tenancy - If either Launch Template Tenancy or VPC Tenancy is set to dedicated, then the instance tenancy is also dedicated. Therefore, this option is incorrect.

Reference:

https://docs.aws.amazon.com/autoscaling/ec2/userguide/advanced-settings-for-your-launch-template.html

Domain

Design High-Performing Architectures

Question 20Skipped

A healthcare company has deployed its web application on Amazon Elastic Container Service (Amazon ECS) container instances running behind an Application Load Balancer. The website slows down when the traffic spikes and the website availability is also reduced. The development team has configured Amazon CloudWatch alarms to receive notifications whenever there is an availability constraint so the team can scale out resources. The company wants an automated solution to respond to such events.

Which of the following addresses the given use case?

Correct answer

Configure AWS Auto Scaling to scale out the Amazon ECS cluster when the ECS service's CPU utilization rises above a threshold

Configure AWS Auto Scaling to scale out the Amazon ECS cluster when the Application Load Balancer's target group's CPU utilization rises above a threshold

Configure AWS Auto Scaling to scale out the Amazon ECS cluster when the CloudWatch alarm's CPU utilization rises above a threshold

Configure AWS Auto Scaling to scale out the Amazon ECS cluster when the Application Load Balancer's CPU utilization rises above a threshold

Overall explanation

Correct option:

Configure AWS Auto Scaling to scale out the Amazon ECS cluster when the ECS service's CPU utilization rises above a threshold

You use the Amazon ECS first-run wizard to create a cluster and a service that runs behind an Elastic Load Balancing load balancer. Then you can configure a target tracking scaling policy that scales your service automatically based on the current application load as measured by the service's CPU utilization (from the ECS, ClusterName, and ServiceName category in CloudWatch).

When the average CPU utilization of your service rises above 75% (meaning that more than 75% of the CPU that is reserved for the service is being used), a scale out alarm triggers Service Auto Scaling to add another task to your service to help out with the increased load. Conversely, when the average CPU utilization of your service drops below the target utilization for a sustained period, a scale-in alarm triggers a decrease in the service's desired count to free up those cluster resources for other tasks and services.

To configure target tracking scaling policies for your service

1. For **Scaling policy type**, choose **Target tracking**.
2. For **Policy name**, enter a descriptive name for your policy.
3. For **ECS service metric**, choose the metric to track. The following metrics are available:
   - **ECSServiceAverageCPUUtilization**—Average CPU utilization of the service.
   - **ECSServiceAverageMemoryUtilization**—Average memory utilization of the service.
   - **ALBRequestCountPerTarget**—Number of requests completed per target in an Application Load Balancer target group.
4. For **Target value**, enter the metric value that the policy should maintain. For example, use a target value of `1000` for `ALBRequestCountPerTarget`, or a target value of `75` (%) for `ECSServiceAverageCPUUtilization`.
5. For **Scale-out cooldown period**, enter the amount of time, in seconds, after a scale-out activity completes before another scale-out activity can start. While the scale-out cooldown period is in effect, the capacity that has been added by the previous scale-out activity that initiated the cooldown is calculated as part of the desired capacity for the next scale out. The intention is to continuously (but not excessively) scale out.
6. For **Scale-in cooldown period**, enter the amount of time, in seconds, after a scale-in activity completes before another scale-in activity can start. The scale-in cooldown period is used to block subsequent scale-in requests until it has expired. The intention is to scale in conservatively to protect your application's availability. However, if another alarm triggers a scale out activity during the cooldown period after a scale-in, Service Auto Scaling scales out your scalable target immediately.
7. (Optional) To turn off the scale-in actions for this policy, choose **Disable scale-in**. This allows you to create a separate scaling policy for scale-in later.
8. Choose **Next step**.

via - https://docs.aws.amazon.com/AmazonECS/latest/developerguide/service-configure-auto-scaling.html

Incorrect options:

Configure AWS Auto Scaling to scale out the Amazon ECS cluster when the Application Load Balancer's target group's CPU utilization rises above a threshold

Configure AWS Auto Scaling to scale out the Amazon ECS cluster when the Application Load Balancer's CPU utilization rises above a threshold

Configure AWS Auto Scaling to scale out the Amazon ECS cluster when the CloudWatch alarm's CPU utilization rises above a threshold

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

References:

https://docs.aws.amazon.com/AmazonECS/latest/developerguide/service-autoscaling-targettracking.html

https://docs.aws.amazon.com/AmazonECS/latest/developerguide/service-configure-auto-scaling.html

Domain

Design High-Performing Architectures

Question 21Skipped

A media company wants a low-latency way to distribute live sports results which are delivered via a proprietary application using UDP protocol.

As a solutions architect, which of the following solutions would you recommend such that it offers the BEST performance for this use case?

Use Auto Scaling group to provide a low latency way to distribute live sports results

Correct answer

Use AWS Global Accelerator to provide a low latency way to distribute live sports results

Use Elastic Load Balancing (ELB) to provide a low latency way to distribute live sports results

Use Amazon CloudFront to provide a low latency way to distribute live sports results

Overall explanation

Correct option:

Use AWS Global Accelerator to provide a low latency way to distribute live sports results

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 (AZs). AWS Global Accelerator always routes user traffic to the optimal endpoint based on performance, reacting instantly to changes in application health, your user's location, and policies that you configure. Global Accelerator is a good fit for non-HTTP use cases, such as gaming (UDP), IoT (MQTT), or Voice over IP. Therefore, this option is correct.

How AWS Global Accelerator Works:

via - https://aws.amazon.com/global-accelerator/

Incorrect options:

Use Amazon CloudFront to provide a low latency way to distribute live sports results - 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. CloudFront supports HTTP/RTMP protocol based requests, therefore this option is incorrect.

Use Elastic Load Balancing (ELB) to provide a low latency way to distribute live sports results - Elastic Load Balancing (ELB) automatically distributes incoming application traffic across multiple targets, such as Amazon EC2 instances, containers, IP addresses, and AWS Lambda functions. It can handle the varying load of your application traffic in a single Availability Zone or across multiple Availability Zones. Elastic Load Balancer cannot help with decreasing latency of incoming traffic from the source.

Use Auto Scaling group to provide a low latency way to distribute live sports results - 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. Auto Scaling group cannot help with decreasing latency of incoming traffic from the source.

Exam Alert:

Please note the differences between the capabilities of AWS Global Accelerator and Amazon CloudFront -

AWS Global Accelerator and Amazon CloudFront are separate services that use the AWS global network and its edge locations around the world. Amazon CloudFront improves performance for both cacheable content (such as images and videos) and dynamic content (such as API acceleration and dynamic site delivery). AWS Global Accelerator improves performance for a wide range of applications over TCP or UDP by proxying packets at the edge to applications running in one or more AWS Regions.

AWS Global Accelerator is a good fit for non-HTTP use cases, such as gaming (UDP), IoT (MQTT), or Voice over IP, as well as for HTTP use cases that specifically require static IP addresses or deterministic, fast regional failover. Both services integrate with AWS Shield for DDoS protection.

References:

https://aws.amazon.com/global-accelerator/

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

Domain

Design High-Performing Architectures

Question 22Skipped

The application maintenance team at a company has noticed that the production application is very slow when the business reports are run on the Amazon RDS database. These reports fetch

a large amount of data and have complex queries with multiple joins, spanning across multiple business-critical core tables. CPU, memory, and storage metrics are around 50% of the total capacity.

Can you recommend an improved and cost-effective way of generating the business reports while keeping the production application unaffected?

Correct answer

Create a read replica and connect the report generation tool/application to it

Migrate from General Purpose SSD to magnetic storage to enhance IOPS

Increase the size of Amazon RDS instance

Configure the Amazon RDS instance to be Multi-AZ DB instance, and connect the report generation tool to the DB instance in a different AZ

Overall explanation

Correct option:

Create a read replica and connect the report generation tool/application to it

Amazon RDS Read Replicas provide enhanced performance and durability for Amazon 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.

There are a variety of scenarios where deploying one or more read replicas for a given source DB instance may make sense. Common reasons for deploying a read replica include:

1. Scaling beyond the compute or I/O capacity of a single DB instance for read-heavy database workloads. This excess read traffic can be directed to one or more read replicas.
2. Serving read traffic while the source DB instance is unavailable. If your source DB Instance cannot take I/O requests (e.g. due to I/O suspension for backups or scheduled maintenance), you can direct read traffic to your read replica(s). For this use case, keep in mind that the data on the read replica may be “stale” since the source DB Instance is unavailable.
3. Business reporting or data warehousing scenarios; you may want business reporting queries to run against a read replica, rather than your primary, production DB Instance.
4. You may use a read replica for disaster recovery of the source DB instance, either in the same AWS Region or in another Region.

Comparing Read Replicas with Multi-AZ and Multi-Region Amazon RDS deployments:

Multi-AZ deployments	Multi-Region deployments	Read replicas
Main purpose is high availability	Main purpose is disaster recovery and local performance	Main purpose is scalability
Non-Aurora: synchronous replication; Aurora: asynchronous replication	Asynchronous replication	Asynchronous replication
Non-Aurora: only the primary instance is active; Aurora: all instances are active	All regions are accessible and can be used for reads	All read replicas are accessible and can be used for readscaling
Non-Aurora: automated backups are taken from standby; Aurora: automated backups are taken from shared storage layer	Automated backups can be taken in each region	No backups configured by default
Always span at least two Availability Zones within a single region	Each region can have a Multi-AZ deployment	Can be within an Availability Zone, Cross-AZ, or Cross-Region
Non-Aurora: database engine version upgrades happen on primary; Aurora: all instances are updated together	Non-Aurora: database engine version upgrade is independent in each region; Aurora: all instances are updated together	Non-Aurora: database engine version upgrade is independent from source instance; Aurora: all instances are updated together
Automatic failover to standby (non-Aurora) or read replica (Aurora) when a problem is detected	Aurora allows promotion of a secondary region to be the master	Can be manually promoted to a standalone database instance (non-Aurora) or to be the primary instance (Aurora)

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

Incorrect options:

Increase the size of Amazon RDS instance - This will not help as it's mentioned that the CPU, memory, and storage are running at only 50% of the total capacity.

Migrate from General Purpose SSD to magnetic storage to enhance IOPS - This is incorrect. Amazon RDS supports magnetic storage for backward compatibility only. AWS recommends that you use General Purpose SSD or Provisioned IOPS for any storage needs.

Configure the Amazon RDS instance to be Multi-AZ DB instance, and connect the report generation tool to the DB instance in a different AZ - 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. However, you cannot read from the standby database, making multi-AZ, an incorrect option for the given scenario.

Reference:

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

Domain

Design High-Performing Architectures

Question 23Skipped

A retail company has its flagship application running on a fleet of Amazon EC2 instances behind Elastic Load Balancing (ELB). The engineering team has been seeing recurrent issues wherein

the in-flight requests from the ELB to the Amazon EC2 instances are getting dropped when an instance becomes unhealthy.

Which of the following features can be used to address this issue?

Idle Timeout

Cross Zone load balancing

Correct answer

Connection Draining

Sticky Sessions

Overall explanation

Correct option:

Connection Draining

To ensure that Elastic Load Balancing stops sending requests to instances that are de-registering or unhealthy while keeping the existing connections open, use connection draining. This enables the load balancer to complete in-flight requests made to instances that are de-registering or unhealthy. The maximum timeout value can be set between 1 and 3,600 seconds (the default is 300 seconds). When the maximum time limit is reached, the load balancer forcibly closes connections to the de-registering instance.

Incorrect options:

Cross Zone load balancing - 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). Cross Zone load balancing cannot be used to complete in-flight requests made to instances that are de-registering or unhealthy.

Sticky Sessions - You can use the sticky session feature (also known as session affinity) to enable the load balancer to bind a user's session to a specific instance. This ensures that all requests from the user during the session are sent to the same instance. Sticky sessions cannot be used to complete in-flight requests made to instances that are de-registering or unhealthy.

Idle Timeout - For each request that a client makes through Elastic Load Balancing, the load balancer maintains two connections. The front-end connection is between the client and the load balancer. The back-end connection is between the load balancer and a registered Amazon EC2 instance. The load balancer has a configured "idle timeout" period that applies to its connections. If no data has been sent or received by the time that the "idle timeout" period elapses, the load balancer closes the connection. "Idle timeout" can not be used to complete in-flight requests made to instances that are de-registering or unhealthy.

Reference:

https://docs.aws.amazon.com/elasticloadbalancing/latest/classic/config-conn-drain.html

Domain

Design Secure Architectures

Question 24Skipped

A retail company has connected its on-premises data center to the AWS Cloud via AWS Direct Connect. The company wants to be able to resolve Domain Name System (DNS) queries for any resources in the on-premises network from the AWS VPC and also resolve any DNS queries for resources in the AWS VPC from the on-premises network.

As a solutions architect, which of the following solutions can be combined to address the given use case? (Select two)

Create a universal endpoint on Amazon Route 53 Resolver and then Amazon Route 53 Resolver can receive and forward queries to resolvers on the on-premises network via this endpoint

Create an outbound endpoint on Amazon Route 53 Resolver and then DNS resolvers on the on-premises network can forward DNS queries to Amazon Route 53 Resolver via this endpoint

Create an inbound endpoint on Amazon Route 53 Resolver and then Amazon Route 53 Resolver can conditionally forward queries to resolvers on the on-premises network via this endpoint

Correct selection

Create an outbound endpoint on Amazon Route 53 Resolver and then Amazon Route 53 Resolver can conditionally forward queries to resolvers on the on-premises network via this endpoint

Correct selection

Create an inbound endpoint on Amazon Route 53 Resolver and then DNS resolvers on the on-premises network can forward DNS queries to Amazon Route 53 Resolver via this endpoint

Overall explanation

Correct options:

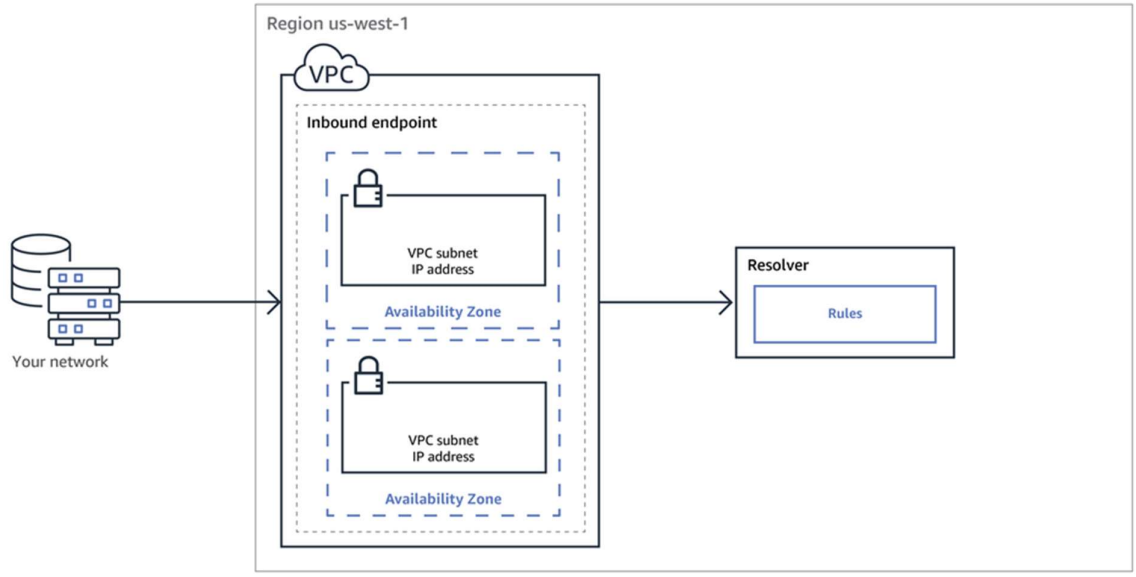
Create an inbound endpoint on Amazon Route 53 Resolver and then DNS resolvers on the on-premises network can forward DNS queries to Amazon Route 53 Resolver via this endpoint

Create an outbound endpoint on Amazon Route 53 Resolver and then Amazon Route 53 Resolver can conditionally forward queries to resolvers on the on-premises network via this endpoint

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 – and can also be used to route users to infrastructure outside of AWS. By default, Amazon Route 53 Resolver automatically answers DNS queries for local VPC domain names for Amazon EC2 instances. You can integrate DNS resolution between Resolver and DNS resolvers on your on-premises network by configuring forwarding rules.

To resolve any DNS queries for resources in the AWS VPC from the on-premises network, you can create an inbound endpoint on Amazon Route 53 Resolver and then DNS resolvers on the on-premises network can forward DNS queries to Amazon Route 53 Resolver via this endpoint.

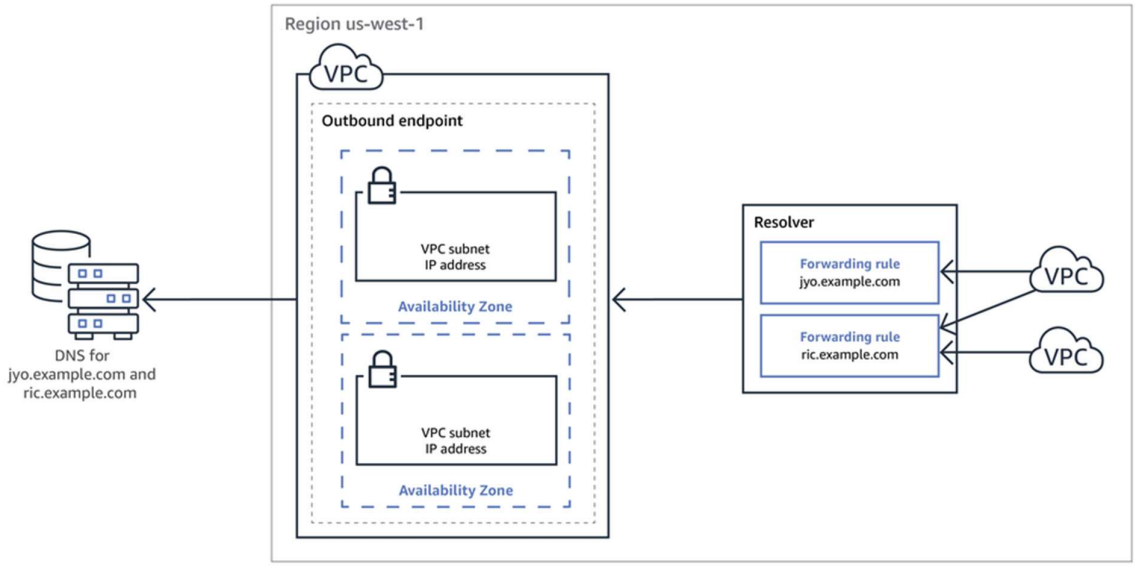
Resolver Inbound Endpoint:



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

To resolve DNS queries for any resources in the on-premises network from the AWS VPC, you can create an outbound endpoint on Amazon Route 53 Resolver and then Amazon Route 53 Resolver can conditionally forward queries to resolvers on the on-premises network via this endpoint. To conditionally forward queries, you need to create Resolver rules that specify the domain names for the DNS queries that you want to forward (such as example.com) and the IP addresses of the DNS resolvers on the on-premises network that you want to forward the queries to.

Resolver Outbound Endpoint:



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

Incorrect options:

Create an outbound endpoint on Amazon Route 53 Resolver and then DNS resolvers on the on-premises network can forward DNS queries to Amazon Route 53 Resolver via this endpoint - DNS resolvers on the on-premises network can forward DNS queries to Amazon Route 53 Resolver via an inbound endpoint. Hence, this option is incorrect.

Create an inbound endpoint on Amazon Route 53 Resolver and then Amazon Route 53 Resolver can conditionally forward queries to resolvers on the on-premises network via this endpoint - Amazon Route 53 Resolver can conditionally forward queries to resolvers on the on-premises network via an outbound endpoint. Hence, this option is incorrect.

Create a universal endpoint on Amazon Route 53 Resolver and then Amazon Route 53 Resolver can receive and forward queries to resolvers on the on-premises network via this endpoint - There is no such thing as a universal endpoint on Amazon Route 53 Resolver. This option has been added as a distractor.

References:

https://docs.aws.amazon.com/Route53/latest/DeveloperGuide/resolver-getting-started.html

https://docs.aws.amazon.com/Route53/latest/DeveloperGuide/resolver.html

Domain

Design Secure Architectures

Question 25Skipped

An e-commerce company has deployed its application on several Amazon EC2 instances that are configured in a private subnet using IPv4. These Amazon EC2 instances read and write a huge volume of data to and from Amazon S3 in the same AWS region. The company has set up subnet routing to direct all the internet-bound traffic through a Network Address Translation gateway (NAT gateway). The company wants to build the most cost-optimal solution without impacting the application's ability to communicate with Amazon S3 or the internet.

As an AWS Certified Solutions Architect Associate, which of the following would you recommend?

Correct answer

Set up a VPC gateway endpoint for Amazon S3. Attach an endpoint policy to the endpoint. Update the route table to direct the S3-bound traffic to the VPC endpoint

Provision an internet gateway. Update the route table in the private subnet to route traffic to the internet gateway. Update the network ACL (NACL) to allow the S3-bound traffic

Set up an egress-only internet gateway in the public subnet. Update the route table in the private subnet to route traffic to the internet gateway. Update the network ACL to allow the S3-bound traffic

Set up a Gateway Load Balancer (GWLB) endpoint for Amazon S3. Update the route table in the private subnet to direct the S3-bound traffic via the Gateway Load Balancer (GWLB) endpoint

Overall explanation

Correct option:

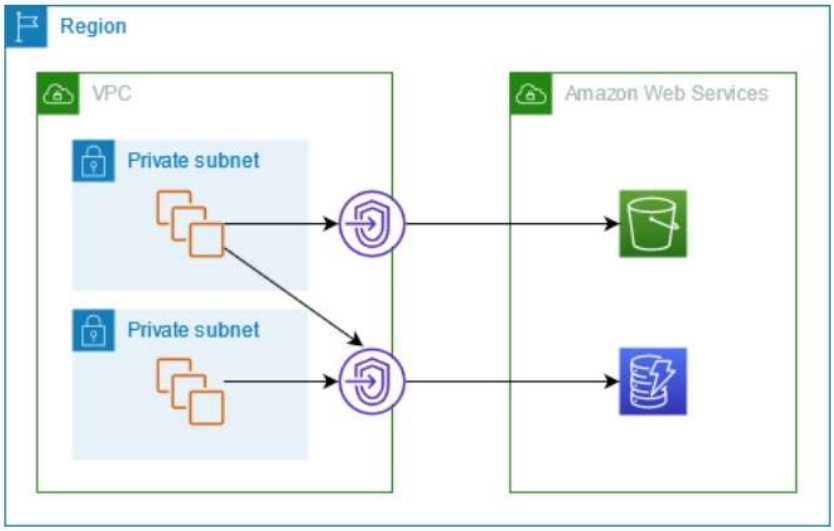
Set up a VPC gateway endpoint for Amazon S3. Attach an endpoint policy to the endpoint. Update the route table to direct the S3-bound traffic to the VPC endpoint

Gateway endpoints provide reliable connectivity to Amazon S3 without requiring an internet gateway or a NAT device for your VPC. After you create the gateway endpoint, you can add it as a target in your route table for traffic destined from your VPC to Amazon S3. There is no additional charge for using gateway endpoints.

The VPC endpoint policy for the gateway endpoint controls access to Amazon S3 from the VPC through the endpoint. The default policy allows full access.

Access through a gateway endpoint

The following diagram shows how instances access Amazon S3 and DynamoDB through a gateway endpoint. Traffic from your VPC to Amazon S3 or DynamoDB is routed to the gateway endpoint. Each subnet route table must have a route that sends traffic destined for the service to the gateway endpoint using the prefix list for the service.



Routing

When you create a gateway endpoint, you select the VPC route tables for the subnets that you enable. The following route is automatically added to each route table that you select. The destination is a prefix list for the service owned by AWS and the target is the gateway endpoint.

Destination	Target
prefix_list_id	*gateway_endpoint_id*

via - https://docs.aws.amazon.com/vpc/latest/privatelink/gateway-endpoints.html

Using the VPC gateway endpoint allows the Amazon EC2 instances to reach Amazon S3 without using the public internet. Since the data transfer remains within the same AWS region, so there is no data transfer costs for ingress as well as egress traffic. Hence this is the most cost-optimal solution.

Incorrect options:

Provision an internet gateway. Update the route table in the private subnet to route traffic to the internet gateway. Update the network ACL (NACL) to allow the S3-bound traffic - If a subnet is associated with a route table that has a route to an internet gateway, it's known as a public subnet. If a subnet is associated with a route table that does not have a route to an internet gateway, it's known as a private subnet. This option has been added as a distractor as adding a route to the internet gateway in the route table associated with the private subnet would make the subnet public. This would also make the internet-bound routing to the NAT gateway redundant. This option has been added as a distractor.

Set up an egress-only internet gateway in the public subnet. Update the route table in the private subnet to route traffic to the internet gateway. Update the network ACL to allow the S3-bound traffic - 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. Since the use case talks about only IPv4 traffic, so this option is incorrect.

Set up a Gateway Load Balancer (GWLB) endpoint for Amazon S3. Update the route table in the private subnet to direct the S3-bound traffic via the Gateway Load Balancer (GWLB) endpoint - Gateway Load Balancers use Gateway Load Balancer endpoints to securely exchange traffic across VPC boundaries. A Gateway Load Balancer endpoint is a VPC endpoint that provides private connectivity between virtual appliances in the service provider VPC and application servers in the service consumer VPC. You cannot set up a gateway load balancer endpoint to access Amazon S3. This option has been added as a distractor.

References:

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

https://aws.amazon.com/premiumsupport/knowledge-center/vpc-reduce-nat-gateway-transfer-costs/

https://docs.aws.amazon.com/vpc/latest/privatelink/vpce-gateway-load-balancer.html

https://docs.aws.amazon.com/vpc/latest/userguide/egress-only-internet-gateway.html

Domain

Design Cost-Optimized Architectures

Question 26Skipped

A retail company uses AWS Cloud to manage its IT infrastructure. The company has set up AWS Organizations to manage several departments running their AWS accounts and using resources such as Amazon EC2 instances and Amazon RDS databases. The company wants to provide shared and centrally-managed VPCs to all departments using applications that need a high degree of interconnectivity.

As a solutions architect, which of the following options would you choose to facilitate this use-case?

Use VPC sharing to share a VPC with other AWS accounts belonging to the same parent organization from AWS Organizations

Correct answer

Use VPC sharing to share one or more subnets with other AWS accounts belonging to the same parent organization from AWS Organizations

Use VPC peering to share one or more subnets with other AWS accounts belonging to the same parent organization from AWS Organizations

Use VPC peering to share a VPC with other AWS accounts belonging to the same parent organization from AWS Organizations

Overall explanation

Correct option:

Use VPC sharing to share one or more subnets with other AWS accounts belonging to the same parent organization from AWS Organizations

VPC sharing (part of Resource Access Manager) allows multiple AWS accounts to create their application resources such as Amazon EC2 instances, Amazon RDS databases, Amazon Redshift clusters, and AWS Lambda functions, into shared and centrally-managed Amazon Virtual Private Clouds (VPCs). To set this up, the account that owns the VPC (owner) shares one or more subnets with other accounts (participants) that belong to the same organization from AWS Organizations. After a subnet is shared, the participants can view, create, modify, and delete their application resources in the subnets shared with them. Participants cannot view, modify, or delete resources that belong to other participants or the VPC owner.

You can share Amazon VPCs to leverage the implicit routing within a VPC for applications that require a high degree of interconnectivity and are within the same trust boundaries. This reduces the number of VPCs that you create and manage while using separate accounts for billing and access control.

Incorrect options:

Use VPC sharing to share a VPC with other AWS accounts belonging to the same parent organization from AWS Organizations - Using VPC sharing, an account that owns the VPC (owner) shares one or more subnets with other accounts (participants) that belong to the same organization from AWS Organizations. The owner account cannot share the VPC itself. Therefore this option is incorrect.

Use VPC peering to share a VPC with other AWS accounts belonging to the same parent organization from AWS Organizations - 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. VPC peering does not facilitate centrally managed VPCs. Therefore this option is incorrect.

Use VPC peering to share one or more subnets with other AWS accounts belonging to the same parent organization from AWS Organizations - 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. VPC peering does not facilitate centrally managed

VPCs. Moreover, an AWS owner account cannot share the VPC itself with another AWS account. Therefore this option is incorrect.

References:

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

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

Domain

Design Secure Architectures

Question 27Skipped

The DevOps team at an IT company is provisioning a two-tier application in a VPC with a public subnet and a private subnet. The team wants to use either a Network Address Translation (NAT) instance or a Network Address Translation (NAT) gateway in the public subnet to enable instances in the private subnet to initiate outbound IPv4 traffic to the internet but needs some technical assistance in terms of the configuration options available for the Network Address Translation (NAT) instance and the Network Address Translation (NAT) gateway.

As a solutions architect, which of the following options would you identify as CORRECT? (Select three)

NAT gateway can be used as a bastion server

Correct selection

Security Groups can be associated with a NAT instance

Correct selection

NAT instance can be used as a bastion server

Correct selection

NAT instance supports port forwarding

NAT gateway supports port forwarding

Security Groups can be associated with a NAT gateway

Overall explanation

Correct options:

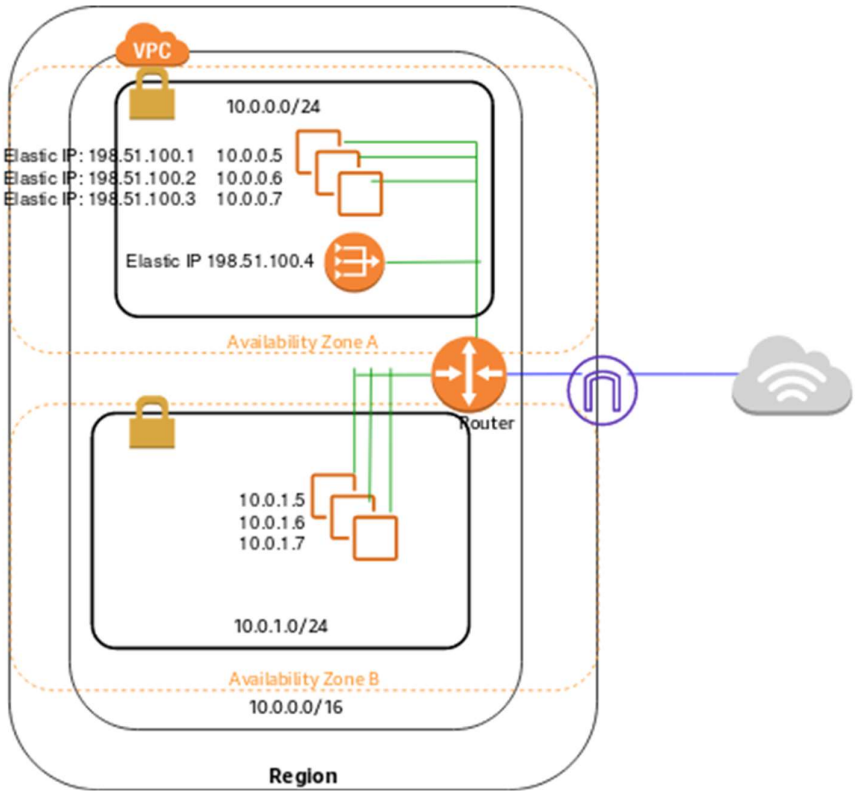
NAT instance can be used as a bastion server

Security Groups can be associated with a NAT instance

NAT instance supports port forwarding

A NAT instance or a NAT Gateway can be used in a public subnet in your VPC to enable instances in the private subnet to initiate outbound IPv4 traffic to the Internet.

How NAT Gateway works:



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

How NAT Instance works:

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

Please see this high-level summary of the differences between NAT instances and NAT gateways relevant to the options described in the question:

The following is a high-level summary of the differences between NAT instances and NAT gateways.

Attribute	NAT gateway	NAT instance
Security groups	Cannot be associated with a NAT gateway. You can associate security groups with your resources behind the NAT gateway to control inbound and outbound traffic.	Associate with your NAT instance and the resources behind your NAT instance to control inbound and outbound traffic.
Network ACLs	Use a network ACL to control the traffic to and from the subnet in which your NAT gateway resides.	Use a network ACL to control the traffic to and from the subnet in which your NAT instance resides.
Flow logs	Use flow logs to capture the traffic.	Use flow logs to capture the traffic.
Port forwarding	Not supported.	Manually customize the configuration to support port forwarding.
Bastion servers	Not supported.	Use as a bastion server.
Traffic metrics	View CloudWatch metrics for the NAT gateway.	View CloudWatch metrics for the instance.
Timeout behavior	When a connection times out, a NAT gateway returns an RST packet to any resources behind the	When a connection times out, a NAT instance sends a FIN packet to resources behind the NAT

via - https://docs.aws.amazon.com/vpc/latest/userguide/vpc-nat-comparison.html

Incorrect options:

NAT gateway supports port forwarding

Security Groups can be associated with a NAT gateway

NAT gateway can be used as a bastion server

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

Reference:

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

Domain

Design High-Performing Architectures

Question 28Skipped

A big data analytics company is working on a real-time vehicle tracking solution. The data processing workflow involves both I/O intensive and throughput intensive database workloads. The development team needs to store this real-time data in a NoSQL database hosted on an Amazon EC2 instance and needs to support up to 25,000 IOPS per volume.

As a solutions architect, which of the following Amazon Elastic Block Store (Amazon EBS) volume types would you recommend for this use-case?

Cold HDD (sc1)

General Purpose SSD (gp2)

Throughput Optimized HDD (st1)

Correct answer

Provisioned IOPS SSD (io1)

Overall explanation

Correct option:

Provisioned IOPS SSD (io1)

Provisioned IOPS SSD (io1) is backed by solid-state drives (SSDs) and is a high-performance Amazon EBS storage option designed for critical, I/O intensive database and application workloads, as well as throughput-intensive database workloads. io1 is designed to deliver a consistent baseline performance of up to 50 IOPS/GB to a maximum of 64,000 IOPS and provide up to 1,000 MB/s of throughput per volume. Therefore, the io1 volume type would be able to meet the requirement of 25,000 IOPS per volume for the given use-case.

Incorrect options:

General Purpose SSD (gp2) - gp2 is backed by solid-state drives (SSDs) and is suitable for a broad range of transactional workloads, including dev/test environments, low-latency interactive applications, and boot volumes. It supports max IOPS/Volume of 16,000.

Cold HDD (sc1) - sc1 is backed by hard disk drives (HDDs). It is ideal for less frequently accessed workloads with large, cold datasets. It supports max IOPS/Volume of 250.

Throughput Optimized HDD (st1) - st1 is backed by hard disk drives (HDDs) and is ideal for frequently accessed, throughput-intensive workloads with large datasets and large I/O sizes, such as MapReduce, Kafka, log processing, data warehouse, and ETL workloads. It supports max IOPS/Volume of 500.

Reference:

https://aws.amazon.com/ebs/volume-types/

Domain

Design High-Performing Architectures

Question 29Skipped

A company has its application servers in the public subnet that connect to the Amazon RDS instances in the private subnet. For regular maintenance, the Amazon RDS instances need patch fixes that need to be downloaded from the internet.

Considering the company uses only IPv4 addressing and is looking for a fully managed service, which of the following would you suggest as an optimal solution?

Configure a Network Address Translation instance (NAT instance) in the public subnet of the VPC

Correct answer

Configure a Network Address Translation gateway (NAT gateway) in the public subnet of the VPC

Configure the Internet Gateway of the VPC to be accessible to the private subnet resources by changing the route tables

Configure an Egress-only internet gateway for the resources in the private subnet of the VPC

Overall explanation

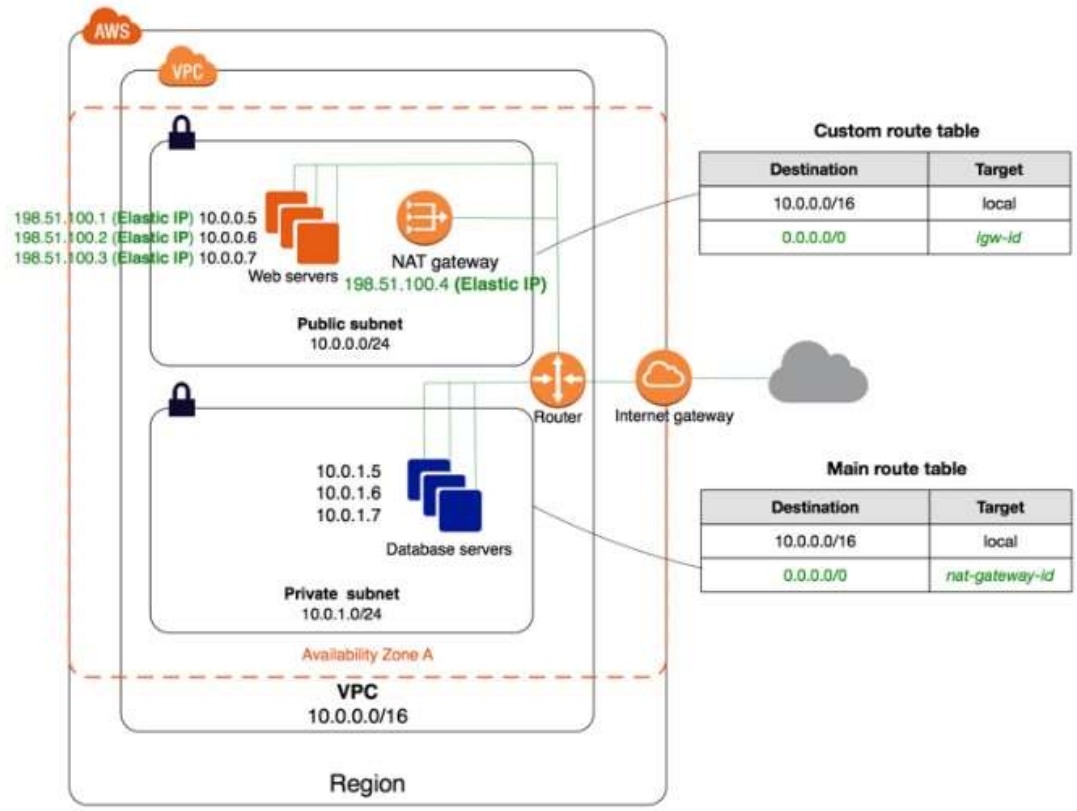
Correct option:

Configure a Network Address Translation gateway (NAT gateway) in the public subnet of the VPC

You can use a Network Address Translation gateway (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. To create a NAT gateway, you must specify the public subnet in which the NAT gateway should reside.

You must also specify an Elastic IP address to associate with the NAT gateway when you create it. The Elastic IP address cannot be changed after you associate it with the NAT Gateway. After you've created a NAT gateway, you must update the route table associated with one or more of your private subnets to point internet-bound traffic to the NAT gateway. This enables instances in your private subnets to communicate with the internet. If you no longer need a NAT gateway, you can delete it. Deleting a NAT gateway disassociates its Elastic IP address, but does not release the address from your account.

VPC architecture with NAT:



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

Incorrect options:

Configure an Egress-only internet gateway for the resources in the private subnet of the VPC - An Egress-only internet gateway is an Internet Gateway that supports IPv6 traffic, so this option is not correct for the given use-case.

Configure a Network Address Translation instance (NAT instance) in the public subnet of the VPC - You can use a network address translation (NAT) instance in a public subnet in your VPC to enable instances in the private subnet to initiate outbound IPv4 traffic to the internet or other AWS services, but prevent the instances from receiving inbound traffic initiated by someone on the internet. NAT instances are not a managed service, it has to be managed and maintained by the customer.

Configure the Internet Gateway of the VPC to be accessible to the private subnet resources by changing the route tables - Internet Gateway cannot be used directly with a private subnet. It is not possible to set up this configuration, without a NAT instance or a NAT gateway in the public subnet.

References:

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

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

https://docs.aws.amazon.com/vpc/latest/userguide/egress-only-internet-gateway.html

Domain

Design Secure Architectures

Question 30Skipped

A leading online gaming company is migrating its flagship application to AWS Cloud for delivering its online games to users across the world. The company would like to use a Network Load Balancer to handle millions of requests per second. The engineering team has provisioned multiple instances in a public subnet and specified these instance IDs as the targets for the NLB.

As a solutions architect, can you help the engineering team understand the correct routing mechanism for these target instances?

Correct answer

Traffic is routed to instances using the primary private IP address specified in the primary network interface for the instance

Traffic is routed to instances using the primary public IP address specified in the primary network interface for the instance

Traffic is routed to instances using the instance ID specified in the primary network interface for the instance

Traffic is routed to instances using the primary elastic IP address specified in the primary network interface for the instance

Overall explanation

Correct option:

Traffic is routed to instances using the primary private IP address specified in the primary network interface for the instance

A Network Load Balancer functions at the fourth layer of the Open Systems Interconnection (OSI) model. It can handle millions of requests per second. After the load balancer receives a connection request, it selects a target from the target group for the default rule. It attempts to open a TCP connection to the selected target on the port specified in the listener configuration.

Request Routing and IP Addresses -

If you specify targets using an instance ID, traffic is routed to instances using the primary private IP address specified in the primary network interface for the instance. The load balancer rewrites the destination IP address from the data packet before forwarding it to the target instance.

If you specify targets using IP addresses, you can route traffic to an instance using any private IP address from one or more network interfaces. This enables multiple applications on an instance to use the same port. Note that each network interface can have its security group. The load balancer rewrites the destination IP address before forwarding it to the target.

Incorrect options:

Traffic is routed to instances using the primary public IP address specified in the primary network interface for the instance - If you specify targets using an instance ID, traffic is routed to instances using the primary private IP address specified in the primary network interface for the instance. So public IP address cannot be used to route the traffic to the instance.

Traffic is routed to instances using the primary elastic IP address specified in the primary network interface for the instance - If you specify targets using an instance ID, traffic is routed to instances using the primary private IP address specified in the primary network interface for the instance. So elastic IP address cannot be used to route the traffic to the instance.

Traffic is routed to instances using the instance ID specified in the primary network interface for the instance - You cannot use instance ID to route traffic to the instance. This option is just added as a distractor.

References:

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

https://docs.aws.amazon.com/elasticloadbalancing/latest/network/load-balancer-target-groups.html

Domain

Design High-Performing Architectures

Question 31Skipped

A health care application processes the real-time health data of the patients into an analytics workflow. With a sharp increase in the number of users, the system has become slow and sometimes even unresponsive as it does not have a retry mechanism. The startup is looking at a scalable solution that has minimal implementation overhead.

Which of the following would you recommend as a scalable alternative to the current solution?

Use Amazon Simple Notification Service (Amazon SNS) for data ingestion and configure AWS Lambda to trigger logic for downstream processing

Use Amazon Simple Queue Service (Amazon SQS) for data ingestion and configure AWS Lambda to trigger logic for downstream processing

Correct answer

Use Amazon Kinesis Data Streams to ingest the data, process it using AWS Lambda or run analytics using Amazon Kinesis Data Analytics

Use Amazon API Gateway with the existing REST-based interface to create a high performing architecture

Overall explanation

Correct option:

Use Amazon Kinesis Data Streams to ingest the data, process it using AWS Lambda or run analytics using Amazon Kinesis Data Analytics

Amazon Kinesis Data Streams (KDS) is a massively scalable and durable real-time data streaming service with support for retry mechanism. 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 data collected is available in milliseconds to enable real-time analytics use cases such as real-time dashboards, real-time anomaly detection, dynamic pricing, and more.

KDS makes sure your streaming data is available to multiple real-time analytics applications, to Amazon S3, or AWS Lambda within 70 milliseconds of the data being collected. Amazon Kinesis data streams scale from megabytes to terabytes per hour and scale from thousands to millions of PUT records per second. You can dynamically adjust the throughput of your stream at any time based on the volume of your input data.

How Data Streams work:

via - https://aws.amazon.com/kinesis/data-streams/?nc=sn&loc=2&dn=2

Incorrect options:

Use Amazon Simple Notification Service (Amazon SNS) for data ingestion and configure AWS Lambda to trigger logic for downstream processing - Amazon Simple Notification Service (Amazon SNS) is a fully managed messaging service for both application-to-application (A2A) and application-to-person (A2P) communication. Amazon SNS is a push mechanism that does not support robust retry mechanisms, as is needed in the current use case.

Use Amazon Simple Queue Service (Amazon SQS) for data ingestion and configure AWS Lambda to trigger logic for downstream processing - Amazon Simple Queue Service (Amazon SQS) is a messaging service that helps in decoupling systems and reducing the complexity of architecture. Amazon SQS can still work but Amazon Kinesis Data streams is custom made for streaming real-time data.

Use Amazon API Gateway with the existing REST-based interface to create a high performing architecture - Amazon API Gateway is not meant for handling real-time streaming data.

Reference:

https://aws.amazon.com/kinesis/data-streams/?nc=sn&loc=2&dn=2

Domain

Design Resilient Architectures

Question 32Skipped

The database backend for a retail company's website is hosted on Amazon RDS for MySQL having a primary instance and three read replicas to support read scalability. The company has mandated that the read replicas should lag no more than 1 second behind the primary instance to provide the best possible user experience. The read replicas are falling further behind during periods of peak traffic spikes, resulting in a bad user experience as the searches produce inconsistent results.

You have been hired as an AWS Certified Solutions Architect - Associate to reduce the replication lag as much as possible with minimal changes to the application code or the effort required to manage the underlying resources.

Which of the following will you recommend?

Set up an Amazon ElastiCache for Redis cluster in front of the MySQL database. Update the website to check the cache before querying the read replicas

Host the MySQL primary database on a memory-optimized Amazon EC2 instance. Spin up additional compute-optimized Amazon EC2 instances to host the read replicas

Correct answer

Set up database migration from Amazon RDS MySQL to Amazon Aurora MySQL. Swap out the MySQL read replicas with Aurora Replicas. Configure Aurora Auto Scaling

Set up database migration from Amazon RDS MySQL to Amazon DynamoDB. Provision a large number of read capacity units (RCUs) to support the required throughput and enable Auto-Scaling

Overall explanation

Correct option:

Set up database migration from Amazon RDS MySQL to Amazon Aurora MySQL. Swap out the MySQL read replicas with Aurora Replicas. Configure Aurora Auto Scaling

Aurora features a distributed, fault-tolerant, and self-healing storage system that is decoupled from compute resources and auto-scales up to 128 TiB 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 Simple Storage Service (Amazon S3), and replication across three Availability Zones (AZs).

Since Amazon Aurora Replicas share the same data volume as the primary instance in the same AWS Region, there is virtually no replication lag. The replica lag times are in the 10s of milliseconds (compared to the replication lag of seconds in the case of MySQL read replicas). Therefore, this is the right option to ensure that the read replicas lag no more than 1 second behind the primary instance.

Aurora Replicas:

Feature	Amazon Aurora Replicas	MySQL Replicas
Number of replicas	Up to 15	Up to 5
Replication type	Asynchronous (milliseconds)	Asynchronous (seconds)
Performance impact on primary	Low	High
Replica location	In-region	Cross-region
Act as failover target	Yes (no data loss)	Yes (potentially minutes of data loss)
Automated failover	Yes	No
Support for user-defined replication delay	No	Yes
Support for different data or schema vs. primary	No	Yes

via - https://aws.amazon.com/rds/aurora/faqs/

Incorrect options:

Host the MySQL primary database on a memory-optimized Amazon EC2 instance. Spin up additional compute-optimized Amazon EC2 instances to host the read replicas - Hosting the MySQL primary database and the read replicas on the Amazon EC2 instances would result in significant overhead to manage the underlying resources such as OS patching, database patching, etc. So this option is incorrect.

Set up an Amazon ElastiCache for Redis cluster in front of the MySQL database. Update the website to check the cache before querying the read replicas - Introducing a caching layer would result in significant changes to the application code, so this option is incorrect.

Set up database migration from Amazon RDS MySQL to Amazon DynamoDB. Provision a large number of read capacity units (RCUs) to support the required throughput and enable Auto-Scaling - Introducing a NoSQL database, such as Amazon DynamoDB, would result in significant changes to the application code since the database queries would have to be re-written for Amazon DynamoDB. Therefore, this option is incorrect.

Reference:

https://aws.amazon.com/rds/aurora/faqs/

Domain

Design High-Performing Architectures

Question 33Skipped

The engineering team at an e-commerce company wants to migrate from Amazon Simple Queue Service (Amazon SQS) Standard queues to FIFO (First-In-First-Out) queues with batching.

As a solutions architect, which of the following steps would you have in the migration checklist? (Select three)

Convert the existing standard queue into a FIFO (First-In-First-Out) queue

Correct selection

Make sure that the throughput for the target FIFO (First-In-First-Out) queue does not exceed 3,000 messages per second

Make sure that the throughput for the target FIFO (First-In-First-Out) queue does not exceed 300 messages per second

Correct selection

Delete the existing standard queue and recreate it as a FIFO (First-In-First-Out) queue

Correct selection

Make sure that the name of the FIFO (First-In-First-Out) queue ends with the .fifo suffix

Make sure that the name of the FIFO (First-In-First-Out) queue is the same as the standard queue

Overall explanation

Correct options:

Delete the existing standard queue and recreate it as a FIFO (First-In-First-Out) queue

Make sure that the name of the FIFO (First-In-First-Out) queue ends with the .fifo suffix

Make sure that the throughput for the target FIFO (First-In-First-Out) queue does not exceed 3,000 messages per second

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 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 SQS offers two types of message queues. Standard queues offer maximum throughput, best-effort ordering, and at-least-once delivery. SQS FIFO queues are designed to guarantee that messages are processed exactly once, in the exact order that they are sent.

By default, FIFO queues support up to 3,000 messages per second with batching, or up to 300 messages per second (300 send, receive, or delete operations per second) without batching. Therefore, using batching you can meet a throughput requirement of upto 3,000 messages per second.

The name of a FIFO queue must end with the .fifo suffix. The suffix counts towards the 80-character queue name limit. To determine whether a queue is FIFO, you can check whether the queue name ends with the suffix.

If you have an existing application that uses standard queues and you want to take advantage of the ordering or exactly-once processing features of FIFO queues, you need to configure the queue and your application correctly. You can't convert an existing standard queue into a FIFO queue. To make the move, you must either create a new FIFO queue for your application or delete your existing standard queue and recreate it as a FIFO queue.

Incorrect options:

Convert the existing standard queue into a FIFO (First-In-First-Out) queue

Make sure that the name of the FIFO (First-In-First-Out) queue is the same as the standard queue - The name of a FIFO queue must end with the .fifo suffix.

Make sure that the throughput for the target FIFO (First-In-First-Out) queue does not exceed 300 messages per second - By default, FIFO queues support up to 3,000 messages per second with batching.

References:

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

https://docs.aws.amazon.com/AWSSimpleQueueService/latest/SQSDeveloperGuide/FIFO-queues.html

Domain

Design Resilient Architectures

Question 34Skipped

A company has set up AWS Organizations to manage several departments running their own AWS accounts. The departments operate from different countries and are spread across various AWS Regions. The company wants to set up a consistent resource provisioning process across departments so that each resource follows pre-defined configurations such as using a specific type of Amazon EC2 instances, specific IAM roles for AWS Lambda functions, etc.

As a solutions architect, which of the following options would you recommend for this use-case?

Use AWS CloudFormation templates to deploy the same template across AWS accounts and regions

Use AWS CloudFormation stacks to deploy the same template across AWS accounts and regions

Use AWS Resource Access Manager (AWS RAM) to deploy the same template across AWS accounts and regions

Correct answer

Use AWS CloudFormation StackSets to deploy the same template across AWS accounts and regions

Overall explanation

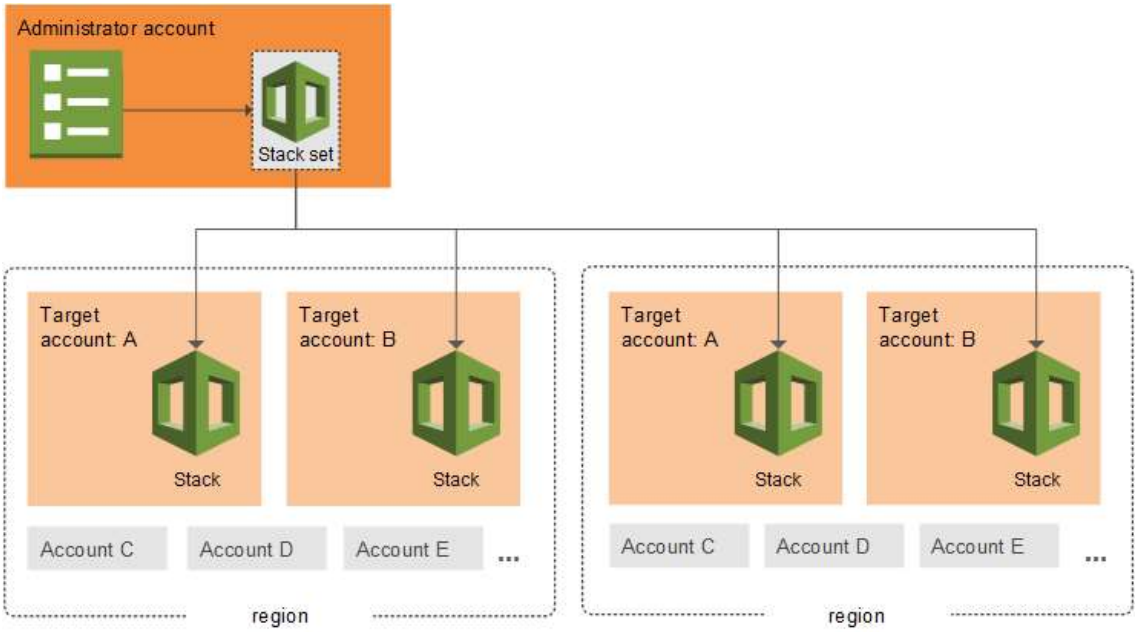
Correct option:

Use AWS CloudFormation StackSets to deploy the same template across AWS accounts and regions

AWS CloudFormation StackSet extends the functionality of stacks by enabling you to create, update, or delete stacks across multiple accounts and regions with a single operation. A stack set lets you create stacks in AWS accounts across regions by using a single AWS

CloudFormation template. Using an administrator account of an "AWS Organization", you define and manage an AWS CloudFormation template, and use the template as the basis for provisioning stacks into selected target accounts of an "AWS Organization" across specified regions.

AWS CloudFormation StackSets:



via - https://docs.aws.amazon.com/AWSCloudFormation/latest/UserGuide/stacksets-concepts.html

Incorrect options:

Use AWS CloudFormation templates to deploy the same template across AWS accounts and regions - AWS Cloudformation template is a JSON or YAML-format, text-based file that describes all the AWS resources you need to deploy to run your application. A template acts as a blueprint for a stack. AWS CloudFormation templates cannot be used to deploy the same template across AWS accounts and regions.

Use AWS CloudFormation stacks to deploy the same template across AWS accounts and regions - AWS CloudFormation stack is a set of AWS resources that are created and managed as a single unit when AWS CloudFormation instantiates a template. A stack cannot be used to deploy the same template across AWS accounts and regions.

Use AWS Resource Access Manager (AWS RAM) to deploy the same template across AWS accounts and regions - AWS Resource Access Manager (AWS RAM) is a service that enables you to easily and securely share AWS resources with any AWS account or within your AWS Organization. Resource Access Manager cannot be used to deploy the same template across AWS accounts and regions.

References:

https://docs.aws.amazon.com/AWSCloudFormation/latest/UserGuide/stacksets-concepts.html

https://docs.aws.amazon.com/AWSCloudFormation/latest/UserGuide/cfn-whatis-howdoesitwork.html

Domain

Design High-Performing Architectures

Question 35Skipped

A social media startup uses AWS Cloud to manage its IT infrastructure. The engineering team at the startup wants to perform weekly database rollovers for a MySQL database server using a serverless cron job that typically takes about 5 minutes to execute the database rollover script written in Python. The database rollover will archive the past week's data from the production database to keep the database small while still keeping its data accessible.

As a solutions architect, which of the following would you recommend as the MOST cost-efficient and reliable solution?

Create a time-based schedule option within an AWS Glue job to invoke itself every week and run the database rollover script

Provision an Amazon EC2 spot instance to run the database rollover script to be run via an OS-based weekly cron expression

Provision an Amazon EC2 scheduled reserved instance to run the database rollover script to be run via an OS-based weekly cron expression

Correct answer

Schedule a weekly Amazon EventBridge event cron expression to invoke an AWS Lambda function that runs the database rollover job

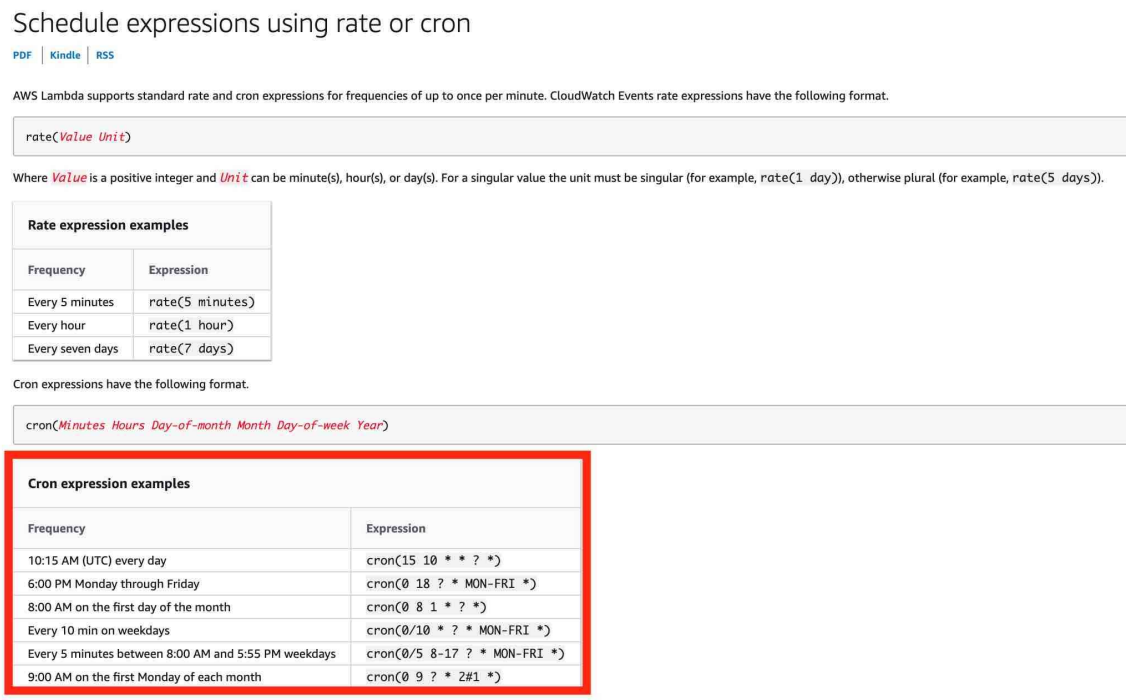
Overall explanation

Correct option:

Schedule a weekly Amazon EventBridge event cron expression to invoke an AWS Lambda function that runs the database rollover job

AWS Lambda lets you run code without provisioning or managing servers. You pay only for the compute time you consume. AWS Lambda supports standard rate and cron expressions for frequencies of up to once per minute.

Schedule expressions using rate or cron:



Schedule expressions using rate or cron

PDF | Kindle | RSS

AWS Lambda supports standard rate and cron expressions for frequencies of up to once per minute. CloudWatch Events rate expressions have the following format.

```
rate(Value Unit)
```

Where *Value* is a positive integer and *Unit* can be minute(s), hour(s), or day(s). For a singular value the unit must be singular (for example, `rate(1 day)`), otherwise plural (for example, `rate(5 days)`).

Rate expression examples

Frequency	Expression
Every 5 minutes	`rate(5 minutes)`
Every hour	`rate(1 hour)`
Every seven days	`rate(7 days)`

Cron expressions have the following format.

```
cron(Minutes Hours Day-of-month Month Day-of-week Year)
```

Cron expression examples

Frequency	Expression
10:15 AM (UTC) every day	`cron(15 10 * * ? *)`
6:00 PM Monday through Friday	`cron(0 18 ? * MON-FRI *)`
8:00 AM on the first day of the month	`cron(0 8 1 * ? *)`
Every 10 min on weekdays	`cron(0/10 * ? * MON-FRI *)`
Every 5 minutes between 8:00 AM and 5:55 PM weekdays	`cron(0/5 8-17 ? * MON-FRI *)`
9:00 AM on the first Monday of each month	`cron(0 9 ? * 2#1 *)`

Incorrect options:

Create a time-based schedule option within an AWS Glue job to invoke itself every week and run the database rollover script - 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 and it's not the right fit for running a database rollover script. Although AWS Glue is also serverless, AWS Lambda is a more cost-effective option compared to AWS Glue.

Provision an Amazon EC2 spot instance to run the database rollover script to be run via an OS-based weekly cron expression - 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 (up to 90% off the On-Demand price). As the Spot Instance runs whenever capacity is available, there is no guarantee that the weekly job will be executed during the defined time window. Additionally, the given use-case requires a serverless solution, therefore this option is incorrect.

Provision an Amazon EC2 scheduled reserved instance to run the database rollover script to be run via an OS-based weekly cron expression - Scheduled Reserved Instances run on a part-time basis. Scheduled Reserved Instances option allows you to use reserve capacity on a recurring daily, weekly, and monthly schedules. Scheduled Reserved Instances are available for one-year terms at 5-10% below On-Demand rates. As the given use-case requires a serverless solution, therefore this option is incorrect.

References:

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

https://docs.aws.amazon.com/lambda/latest/dg/services-cloudwatchevents-expressions.html

Domain

Design Cost-Optimized Architectures

Question 36Skipped

Your application is hosted by a provider on yourapp.provider.com. You would like to have your users access your application using www.your-domain.com, which you own and manage under Amazon Route 53.

Which Amazon Route 53 record should you create?

Create an A record

Create a PTR record

Create an Alias Record

Correct answer

Create a CNAME record

Overall explanation

Correct option:

Create a CNAME record

A CNAME record maps DNS queries for the name of the current record, such as acme.example.com, to another domain (example.com or example.net) or subdomain (acme.example.com or zenith.example.org).

CNAME records can be used to map one domain name to another. Although you should keep in mind that the DNS protocol does not allow you to create a CNAME record for 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 cannot create a CNAME record for example.com, but you can create CNAME records for www.example.com, newproduct.example.com, and so on.

Please review the major differences between CNAME and Alias Records:

Comparison of alias and CNAME records

Alias records are similar to CNAME records, but there are some important differences. The following list compares alias records and CNAME records.

Resources that you can redirect queries to

Alias records

An alias record can only redirect queries to selected AWS resources, such as the following:

- Amazon S3 buckets
- CloudFront distributions
- Another record in the same Route 53 hosted zone

For example, you can create an alias record named acme.example.com that redirects queries to an Amazon S3 bucket that is also named acme.example.com. You can also create an acme.example.com alias record that redirects queries to a record named zenith.example.com in the example.com hosted zone.

CNAME records

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 Route 53 as the DNS service for the domain that you're redirecting queries to.

Creating records that have the same name as the domain (records at the zone apex)

Alias records

In most configurations, you can create an alias record that has the same name as the hosted zone (the zone apex). The one exception is when you want to redirect queries from the zone apex (such as example.com) to a record in the same hosted zone that has a type of CNAME (such as zenith.example.com). The alias record must have the same type as the record you're routing traffic to, and creating a CNAME record for the zone apex isn't supported even for an alias record.

CNAME records

You can't create a CNAME record that has the same name as the hosted zone (the zone apex). This is true both for hosted zones for domain names (example.com) and for hosted zones for subdomains (zenith.example.com).

Pricing for DNS queries

Alias records

Route 53 doesn't charge for alias queries to AWS resources. For more information, see Amazon Route 53 Pricing.

CNAME records

Route 53 charges for CNAME queries.

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

Incorrect options:

Create an A record - Used to point a domain or subdomain to an IP address. 'A record' cannot be used to map one domain name to another.

Create a PTR record - A Pointer (PTR) record resolves an IP address to a fully-qualified domain name (FQDN) as an opposite to what A record does. PTR records are also called Reverse DNS records. 'PTR record' cannot be used to map one domain name to another.

Create an Alias Record - 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. 3rd party websites do not qualify for these as we have no control over those. 'Alias record' cannot be used to map one domain name to another.

Reference:

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

Domain

Design Cost-Optimized Architectures

Question 37Skipped

A biotechnology company has multiple High Performance Computing (HPC) workflows that quickly and accurately process and analyze genomes for hereditary diseases. The company is looking to migrate these workflows from their on-premises infrastructure to AWS Cloud.

As a solutions architect, which of the following networking components would you recommend on the Amazon EC2 instances running these HPC workflows?

Correct answer

Elastic Fabric Adapter (EFA)

Elastic IP Address (EIP)

Elastic Network Interface (ENI)

Elastic Network Adapter (ENA)

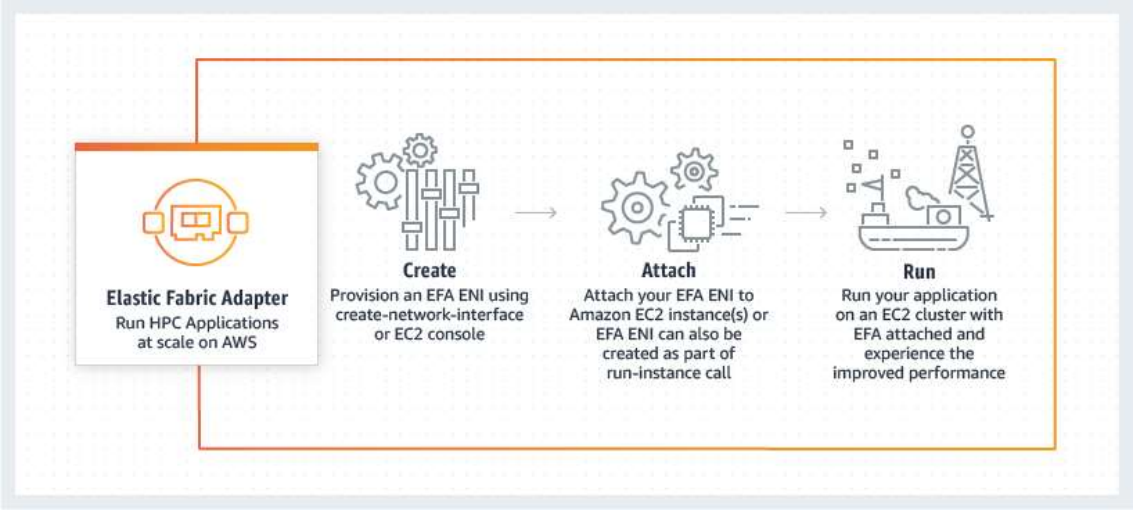
Overall explanation

Correct option:

Elastic Fabric Adapter (EFA)

An Elastic Fabric Adapter (EFA) is a network device that you can attach to your Amazon EC2 instance to accelerate High Performance Computing (HPC) and machine learning applications. It enhances the performance of inter-instance communication that is critical for scaling HPC and machine learning applications. EFA devices provide all Elastic Network Adapter (ENA) devices functionalities plus a new OS bypass hardware interface that allows user-space applications to communicate directly with the hardware-provided reliable transport functionality.

How Elastic Fabric Adapter Works:



via - https://aws.amazon.com/hpc/efa/

Incorrect options:

Elastic Network Interface (ENI) - An Elastic Network Interface (ENI) is a logical networking component in a VPC that represents a virtual network card. You can create a network interface, attach it to an instance, detach it from an instance, and attach it to another instance. The ENI is the simplest networking component available on AWS and is insufficient for HPC workflows.

Elastic Network Adapter (ENA) - Elastic Network Adapter (ENA) devices support enhanced networking via single root I/O virtualization (SR-IOV) to provide high-performance networking capabilities. Although enhanced networking provides higher bandwidth, higher packet per second (PPS) performance, and consistently lower inter-instance latencies, still EFA is a better fit for the given use-case because the EFA device provides all the functionality of an ENA device, plus hardware support for applications to communicate directly with the EFA device without involving the instance kernel (OS-bypass communication) using an extended programming interface.

Elastic IP Address (EIP) - An Elastic IP address (EIP) is a static IPv4 address associated with your AWS account. An Elastic IP address is a public IPv4 address, which is reachable from the internet. It is not a networking device that can be used to facilitate HPC workflows.

References:

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

https://aws.amazon.com/hpc/efa/

Domain

Design High-Performing Architectures

Question 38Skipped

An IT company hosts windows based applications on its on-premises data center. The company is looking at moving the business to the AWS Cloud. The cloud solution should offer shared storage space that multiple applications can access without a need for replication. Also, the solution should integrate with the company's self-managed Active Directory domain.

Which of the following solutions addresses these requirements with the minimal integration effort?

Use File Gateway of AWS Storage Gateway to create a hybrid storage solution

Use Amazon FSx for Lustre as a shared storage solution with millisecond latencies

Correct answer

Use Amazon FSx for Windows File Server as a shared storage solution

Use Amazon Elastic File System (Amazon EFS) as a shared storage solution

Overall explanation

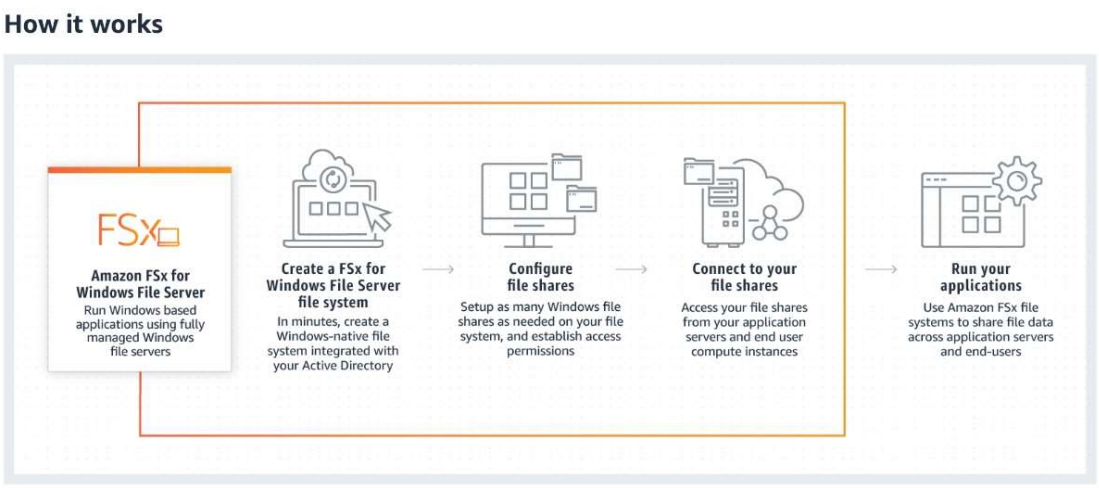
Correct option:

Use Amazon FSx for Windows File Server as a shared storage solution

Amazon FSx for Windows File Server provides fully managed, highly reliable, and scalable file storage that is accessible over the industry-standard Server Message Block (SMB) protocol. It is built on Windows Server, delivering a wide range of administrative features such as user quotas, end-user file restore, and Microsoft Active Directory (AD) integration. It offers single-AZ and multi-AZ deployment options, fully managed backups, and encryption of data at rest and in transit. You can optimize cost and performance for your workload needs with SSD and HDD storage options; and you can scale storage and change the throughput performance of your file system at any time.

With Amazon FSx, you get highly available and durable file storage starting from $0.013 per GB-month. Data deduplication enables you to optimize costs even further by removing redundant data. You can increase your file system storage and scale throughput capacity at any time, making it easy to respond to changing business needs. There are no upfront costs or licensing fees.

How Amazon FSx for Windows File Server works:



via - https://aws.amazon.com/fsx/windows/

Incorrect options:

Use File Gateway of AWS Storage Gateway to create a hybrid storage solution - AWS Storage Gateway connects an on-premises software appliance with cloud-based storage to provide seamless integration between your on-premises IT environment and the AWS storage infrastructure. Storage Gateway uses Amazon S3 to store data on AWS Cloud and from here the on-premises data can seamlessly integrate with Cloud services. It is not suited to be used as a shared storage space that multiple applications can access in parallel.

Use Amazon FSx for Lustre as a shared storage solution with millisecond latencies - Amazon FSx for Lustre is a fully managed service that provides cost-effective, high-performance storage for compute workloads. Many workloads such as machine learning, high performance computing (HPC), video rendering, and financial simulations depend on compute instances accessing the same set of data through high-performance shared storage. Lustre is Linux based, hence it is not the right choice since the use case is about Windows-based applications.

Use Amazon Elastic File System (Amazon EFS) as a shared storage solution - 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. Amazon EFS is a powerful, shared storage solution that would have been the right answer if the customer systems were Linux based. Amazon EFS is compatible with only Linux-based AMIs for Amazon EC2.

Reference:

https://aws.amazon.com/fsx/windows/

Domain

Design Resilient Architectures

Question 39Skipped

A financial services company has recently migrated from on-premises infrastructure to AWS Cloud. The DevOps team wants to implement a solution that allows all resource configurations to be reviewed and make sure that they meet compliance guidelines. Also, the solution should be able to offer the capability to look into the resource configuration history across the application stack.

As a solutions architect, which of the following solutions would you recommend to the team?

Use AWS Systems Manager to review resource configurations to meet compliance guidelines and maintain a history of resource configuration changes

Correct answer

Use AWS Config to review resource configurations to meet compliance guidelines and maintain a history of resource configuration changes

Use AWS CloudTrail to review resource configurations to meet compliance guidelines and maintain a history of resource configuration changes

Use Amazon CloudWatch to review resource configurations to meet compliance guidelines and maintain a history of resource configuration changes

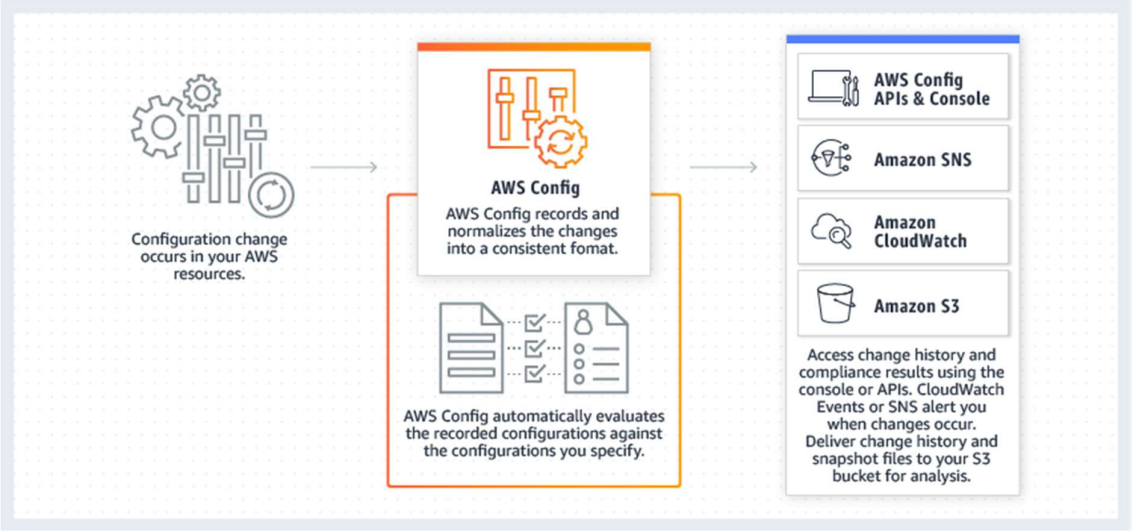
Overall explanation

Correct option:

Use AWS Config to review resource configurations to meet compliance guidelines and maintain a history of resource configuration changes

AWS Config is a service that enables you to assess, audit, and evaluate the configurations of your AWS resources. With Config, you can review changes in configurations and relationships between AWS resources, dive into detailed resource configuration histories, and determine your overall compliance against the configurations specified in your internal guidelines. You can use Config to answer questions such as - “What did my AWS resource look like at xyz point in time?”

How AWS Config Works:



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

Incorrect options:

Use Amazon CloudWatch to review resource configurations to meet compliance guidelines and maintain a history of resource configuration changes - AWS CloudWatch provides you with data and actionable insights to monitor your applications, respond to system-wide performance changes, optimize resource utilization, and get a unified view of operational health. You cannot use Amazon CloudWatch to maintain a history of resource configuration changes.

Use AWS CloudTrail to review resource configurations to meet compliance guidelines and maintain a history of resource configuration changes - With AWS CloudTrail, you can log, continuously monitor, and retain account activity related to actions across your AWS infrastructure. You can use AWS CloudTrail to answer questions such as - "Who made an API call to modify this resource?". AWS CloudTrail provides an event history of your AWS account activity thereby enabling governance, compliance, operational auditing, and risk auditing of your AWS account. You cannot use AWS CloudTrail to maintain a history of resource configuration changes.

Use AWS Systems Manager to review resource configurations to meet compliance guidelines and maintain a history of resource configuration changes - Using AWS Systems Manager, you can group resources, like Amazon EC2 instances, Amazon S3 buckets, or Amazon RDS instances, by application, view operational data for monitoring and troubleshooting, and take action on your groups of resources. You cannot use AWS Systems Manager to maintain a history of resource configuration changes.

Exam Alert:

You may see scenario-based questions asking you to select one of Amazon CloudWatch vs AWS CloudTrail vs AWS Config. Just remember this thumb rule -

Think resource performance monitoring, events, and alerts; think Amazon CloudWatch.

Think account-specific activity and audit; think AWS CloudTrail.

Think resource-specific history, audit, and compliance; think AWS Config.

References:

https://aws.amazon.com/config/

https://aws.amazon.com/cloudwatch/

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

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

Domain

Design Secure Architectures

Question 40Skipped

An IT training company hosted its website on Amazon S3 a couple of years ago. Due to COVID-19 related travel restrictions, the training website has suddenly gained traction. With an almost 300% increase in the requests served per day, the company's AWS costs have sky-rocketed for just the Amazon S3 outbound data costs.

As a Solutions Architect, can you suggest an alternate method to reduce costs while keeping the latency low?

Correct answer

Configure Amazon CloudFront to distribute the data hosted on Amazon S3 cost-effectively

Use Amazon EFS service, as it provides a shared, scalable, fully managed elastic NFS file system for storing AWS Cloud or on-premises data

To reduce Amazon S3 cost, the data can be saved on an Amazon EBS volume connected to an Amazon EC2 instance that can host the application

Configure Amazon S3 Batch Operations to read data in bulk at one go, to reduce the number of calls made to Amazon S3 buckets

Overall explanation

Correct option:

Configure Amazon CloudFront to distribute the data hosted on Amazon S3 cost-effectively

Storing content with Amazon S3 provides a lot of advantages. But to help optimize your application's performance and security while effectively managing cost, AWS recommends that you also set up Amazon CloudFront to work with your Amazon S3 bucket to serve and protect the content.

Amazon CloudFront is a content delivery network (CDN) service that delivers static and dynamic web content, video streams, and APIs around the world, securely and at scale. By design, delivering data out of Amazon CloudFront can be more cost-effective than delivering it from Amazon S3 directly to your users.

Amazon CloudFront serves content through a worldwide network of data centers called Edge Locations. Using edge servers to cache and serve content improves performance by providing

content closer to where viewers are located. Amazon CloudFront has edge servers in locations all around the world.

When a user requests content that you serve with Amazon CloudFront, their request is routed to a nearby Edge Location. If CloudFront has a cached copy of the requested file, CloudFront delivers it to the user, providing a fast (low-latency) response. If the file they've requested isn't yet cached, Amazon CloudFront retrieves it from your origin – for example, the S3 bucket where you've stored your content. Then, for the next local request for the same content, it's already cached nearby and can be served immediately.

By caching your content in Edge Locations, Amazon CloudFront reduces the load on your Amazon S3 bucket and helps ensure a faster response for your users when they request content. Also, data transfer out for content by using CloudFront is often more cost-effective than serving files directly from Amazon S3, and there is no data transfer fee from Amazon S3 to CloudFront. You only pay for what is delivered to the internet from Amazon CloudFront, plus request fees.

Incorrect options:

To reduce Amazon S3 cost, the data can be saved on an Amazon EBS volume connected to an Amazon EC2 instance that can host the application - Amazon EBS volumes are fast and are relatively cheap (though Amazon S3 is still a cheaper alternative). But, Amazon EBS volumes are accessible only through Amazon EC2 instances and are bound to a specific region.

Use Amazon EFS service, as it provides a shared, scalable, fully managed elastic NFS file system for storing AWS Cloud or on-premises data - Amazon EFS is a shareable file system that can be mounted onto Amazon EC2 instances. Amazon EFS is costlier than Amazon EBS and not a solution if the company is looking at reducing costs.

Configure Amazon S3 Batch Operations to read data in bulk at one go, to reduce the number of calls made to Amazon S3 buckets - This statement is incorrect and given only as a distractor. You can use Amazon S3 Batch Operations to perform large-scale batch operations on Amazon S3 objects, and it has nothing to do with content distribution.

Reference:

https://aws.amazon.com/blogs/networking-and-content-delivery/amazon-s3-amazon-cloudfront-a-match-made-in-the-cloud/

Domain

Design Cost-Optimized Architectures

Question 41Skipped

An IT consultant is helping a small business revamp their technology infrastructure on the AWS Cloud. The business has two AWS accounts and all resources are provisioned in the us-west-2 region. The IT consultant is trying to launch an Amazon EC2 instance in each of the two AWS accounts such that the instances are in the same Availability Zone (AZ) of the us-west-2 region. Even after selecting the same default subnet (us-west-2a) while launching the instances in each of the AWS accounts, the IT consultant notices that the Availability Zones (AZs) are still different.

As a solutions architect, which of the following would you suggest resolving this issue?

Use the default subnet to uniquely identify the Availability Zones across the two AWS Accounts

Use the default VPC to uniquely identify the Availability Zones across the two AWS Accounts

Reach out to AWS Support for creating the Amazon EC2 instances in the same Availability Zone (AZ) across the two AWS accounts

Correct answer

Use Availability Zone (AZ) ID to uniquely identify the Availability Zones across the two AWS Accounts

Overall explanation

Correct option:

Use Availability Zone (AZ) ID to uniquely identify the Availability Zones across the two AWS Accounts

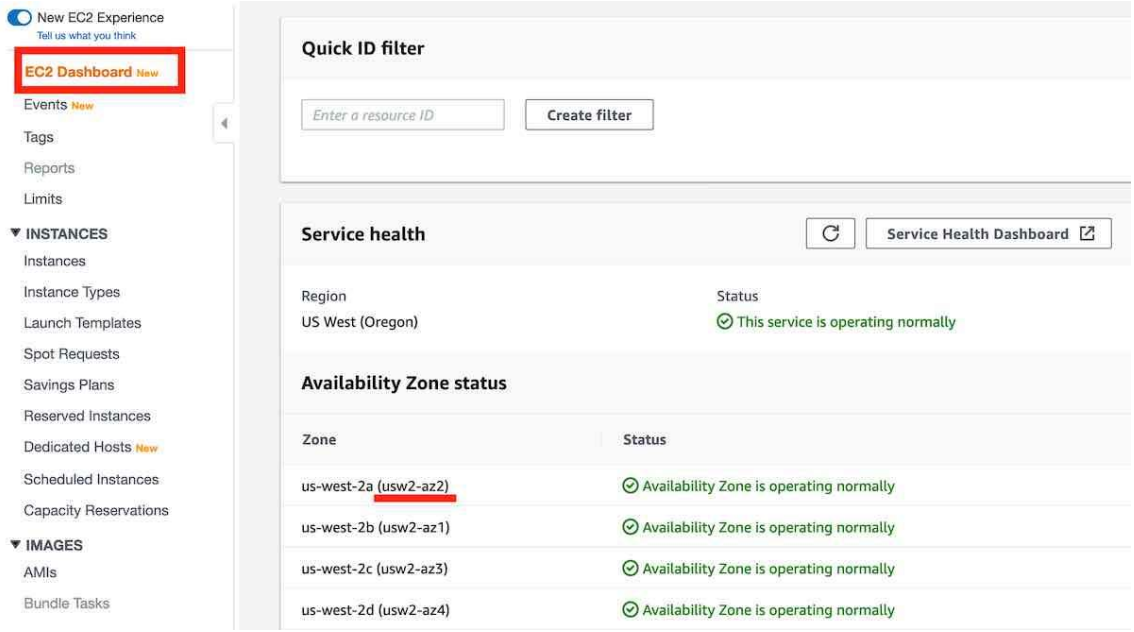
An Availability Zone is represented by a region code followed by a letter identifier; for example, us-east-1a. To ensure that resources are distributed across the Availability Zones for a region, AWS maps Availability Zones to names for each AWS account. For example, the Availability Zone us-west-2a for one AWS account might not be the same location as us-west-2a for another AWS account.

To coordinate Availability Zones across accounts, you must use the AZ ID, which is a unique and consistent identifier for an Availability Zone. For example, usw2-az2 is an AZ ID for the us-west-2 region and it has the same location in every AWS account.

Viewing AZ IDs enables you to determine the location of resources in one account relative to the resources in another account. For example, if you share a subnet in the Availability Zone with the AZ ID usw2-az2 with another account, this subnet is available to that account in the Availability Zone whose AZ ID is also usw2-az2.

You can view the AZ IDs by going to the service health section of the Amazon EC2 Dashboard via your AWS Management Console.

Availability Zone (AZ) IDs for Availability Zones:



Incorrect options:

Use the default VPC to uniquely identify the Availability Zones across the two AWS Accounts - A virtual private cloud (VPC) is a virtual network dedicated to your AWS account. It is logically isolated from other virtual networks in the AWS Cloud. Since a VPC spans an AWS region, it cannot be used to uniquely identify an Availability Zone. Therefore, this option is incorrect.

Use the default subnet to uniquely identify the Availability Zones across the two AWS Accounts - A subnet is a range of IP addresses in your VPC. A subnet spans an Availability Zone of an AWS region. The default subnet representing the Availability Zone us-west-2a for one AWS account might not be the same location as us-west-2a for another AWS account. Therefore, this option is incorrect.

Reach out to AWS Support for creating the Amazon EC2 instances in the same Availability Zone (AZ) across the two AWS accounts - Since the AZ ID is a unique and consistent identifier for an Availability Zone, there is no need to contact AWS Support. Therefore, this option is incorrect.

Reference:

https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-regions-availability-zones.html

Domain

Design High-Performing Architectures

Question 42Skipped

A financial services company wants to move the Windows file server clusters out of their datacenters. They are looking for cloud file storage offerings that provide full Windows compatibility. Can you identify the AWS storage services that provide highly reliable file storage

that is accessible over the industry-standard Server Message Block (SMB) protocol compatible with Windows systems? (Select two)

Amazon Simple Storage Service (Amazon S3)

Amazon Elastic File System (Amazon EFS)

Amazon Elastic Block Store (Amazon EBS)

Correct selection

Amazon FSx for Windows File Server

Correct selection

File Gateway Configuration of AWS Storage Gateway

Overall explanation

Correct options:

Amazon FSx for Windows File Server

Amazon FSx for Windows File Server is a fully managed, highly reliable file storage that is accessible over the industry-standard Server Message Block (SMB) protocol. It is built on Windows Server, delivering a wide range of administrative features such as user quotas, end-user file restore, and Microsoft Active Directory (AD) integration.

File Gateway Configuration of AWS Storage Gateway

Depending on the use case, AWS Storage Gateway provides 3 types of storage interfaces for on-premises applications: File, Volume, and Tape. The File Gateway enables you to store and retrieve objects in Amazon S3 using file protocols such as Network File System (NFS) and Server Message Block (SMB).

Incorrect options:

Amazon Elastic File System (Amazon EFS) - Amazon EFS is a file storage service for use with Amazon EC2. Amazon EFS provides a file system interface, file system access semantics, and concurrently-accessible storage for up to thousands of Amazon EC2 instances. Amazon EFS uses the Network File System protocol. EFS does not support SMB protocol.

Amazon Elastic Block Store (Amazon EBS) - Amazon EBS is a block-level storage service for use with Amazon EC2. Amazon EBS can deliver performance for workloads that require the lowest latency access to data from a single EC2 instance. EBS does not support SMB protocol.

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. Amazon S3 provides a simple, standards-based REST web services interface that is designed to work with any Internet-development toolkit. S3 does not support SMB protocol.

References:

https://aws.amazon.com/fsx/windows/

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

Domain

Design Resilient Architectures

Question 43Skipped

An e-commerce company is using Elastic Load Balancing (ELB) for its fleet of Amazon EC2 instances spread across two Availability Zones (AZs), with one instance as a target in Availability Zone A and four instances as targets in Availability Zone B. The company is doing benchmarking for server performance when cross-zone load balancing is enabled compared to the case when cross-zone load balancing is disabled.

As a solutions architect, which of the following traffic distribution outcomes would you identify as correct?

With cross-zone load balancing enabled, one instance in Availability Zone A receives 50% traffic and four instances in Availability Zone B receive 12.5% traffic each. With cross-zone load balancing disabled, one instance in Availability Zone A receives 20% traffic and four instances in Availability Zone B receive 20% traffic each

Correct answer

With cross-zone load balancing enabled, one instance in Availability Zone A receives 20% traffic and four instances in Availability Zone B receive 20% traffic each. With cross-zone load balancing disabled, one instance in Availability Zone A receives 50% traffic and four instances in Availability Zone B receive 12.5% traffic each

With cross-zone load balancing enabled, one instance in Availability Zone A receives 20% traffic and four instances in Availability Zone B receive 20% traffic each. With cross-zone load balancing disabled, one instance in Availability Zone A receives no traffic and four instances in Availability Zone B receive 25% traffic each

With cross-zone load balancing enabled, one instance in Availability Zone A receives no traffic and four instances in Availability Zone B receive 25% traffic each. With cross-zone load balancing disabled, one instance in Availability Zone A receives 50% traffic and four instances in Availability Zone B receive 12.5% traffic each

Overall explanation

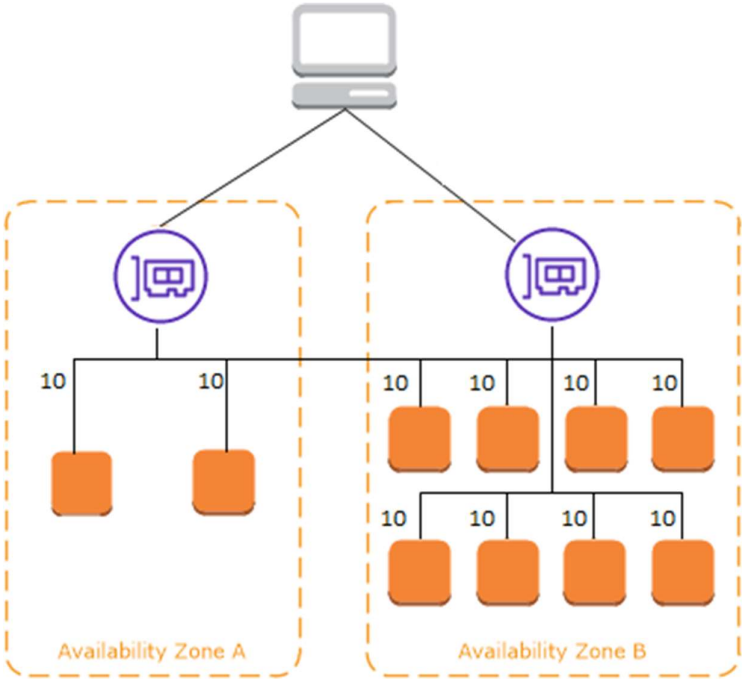
Correct option:

With cross-zone load balancing enabled, one instance in Availability Zone A receives 20% traffic and four instances in Availability Zone B receive 20% traffic each. With cross-zone load balancing disabled, one instance in Availability Zone A receives 50% traffic and four instances in Availability Zone B receive 12.5% traffic each

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. Therefore, one instance in Availability Zone A receives 20% traffic and four instances in Availability Zone B receive 20% traffic each. When cross-zone load balancing is disabled, each load balancer node distributes traffic only across the registered targets in its Availability Zone. Therefore, one instance in Availability Zone A receives 50% traffic and four instances in Availability Zone B receive 12.5% traffic each.

Consider the following diagrams (the scenario illustrated in the diagrams involves 10 target instances split across 2 AZs) to understand the effect of cross-zone load balancing.

If cross-zone load balancing is enabled, each of the 10 targets receives 10% of the traffic. This is because each load balancer node can route its 50% of the client traffic to all 10 targets.



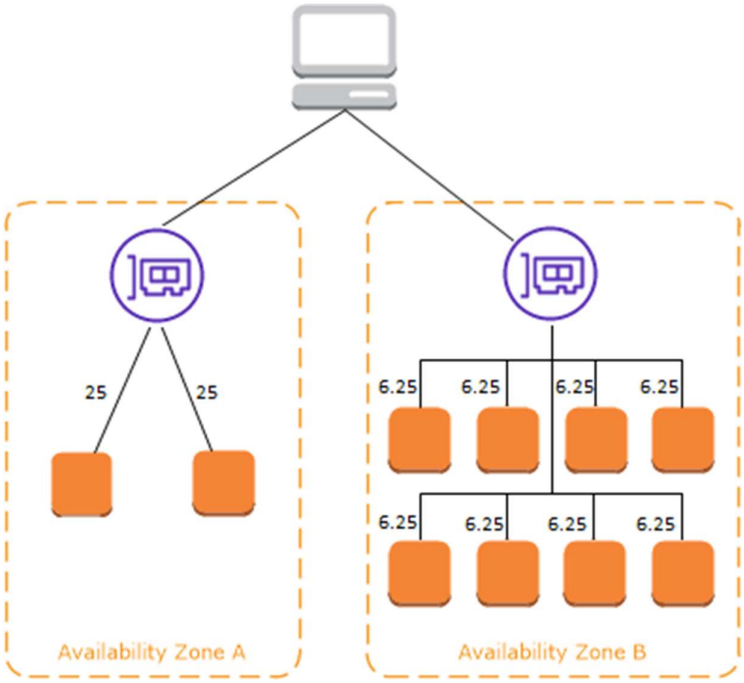
via - https://docs.aws.amazon.com/elasticloadbalancing/latest/userguide/how-elastic-load-balancing-works.html

If cross-zone load balancing is disabled:

Each of the two targets in Availability Zone A receives 25% of the traffic.

Each of the eight targets in Availability Zone B receives 6.25% 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



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

Incorrect options:

With cross-zone load balancing enabled, one instance in Availability Zone A receives 50% traffic and four instances in Availability Zone B receive 12.5% traffic each. With cross-zone load balancing disabled, one instance in Availability Zone A receives 20% traffic and four instances in Availability Zone B receive 20% traffic each

With cross-zone load balancing enabled, one instance in Availability Zone A receives no traffic and four instances in Availability Zone B receive 25% traffic each. With cross-zone load balancing disabled, one instance in Availability Zone A receives 50% traffic and four instances in Availability Zone B receive 12.5% traffic each

With cross-zone load balancing enabled, one instance in Availability Zone A receives 20% traffic and four instances in Availability Zone B receive 20% traffic each. With cross-zone load balancing disabled, one instance in Availability Zone A receives no traffic and four instances in Availability Zone B receive 25% traffic each

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

Reference:

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

Domain

Design Resilient Architectures

Question 44Skipped

An engineering lead is designing a VPC with public and private subnets. The VPC and subnets use IPv4 CIDR blocks. There is one public subnet and one private subnet in each of three Availability Zones (AZs) for high availability. An internet gateway is used to provide internet access for the public subnets. The private subnets require access to the internet to allow Amazon EC2 instances to download software updates.

Which of the following options represents the correct solution to set up internet access for the private subnets?

Set up three NAT gateways, one in each private subnet in each AZ. Create a custom route table for each AZ that forwards non-local traffic to the NAT gateway in its AZ

Set up three Internet gateways, one in each private subnet in each AZ. Create a custom route table for each AZ that forwards non-local traffic to the Internet gateway in its AZ

Set up three egress-only internet gateways, one in each public subnet in each AZ. Create a custom route table for each AZ that forwards non-local traffic to the egress-only internet gateway in its AZ

Correct answer

Set up three NAT gateways, one in each public subnet in each AZ. Create a custom route table for each AZ that forwards non-local traffic to the NAT gateway in its AZ

Overall explanation

Correct option:

Set up three NAT gateways, one in each public subnet in each AZ. Create a custom route table for each AZ that forwards non-local traffic to the NAT gateway in its AZ

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.

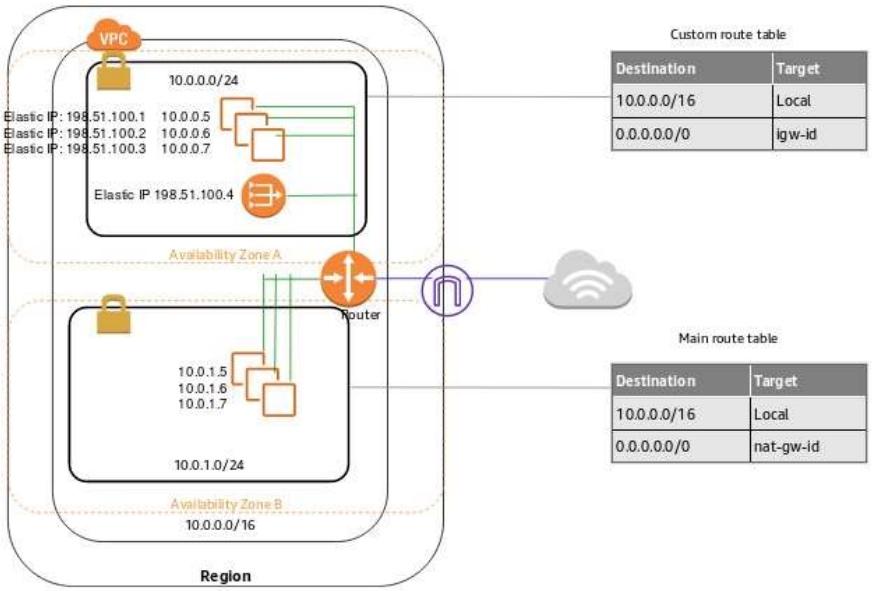
To create a NAT gateway, you must specify the public subnet in which the NAT gateway should reside. You must also specify an Elastic IP address to associate with the NAT gateway when you create it. The Elastic IP address cannot be changed after you associate it with the NAT Gateway. After you've created a NAT gateway, you must update the route table associated with one or more of your private subnets to point internet-bound traffic to the NAT gateway. This enables instances in your private subnets to communicate with the internet.

Each NAT gateway is created in a specific Availability Zone and implemented with redundancy in that zone.

If you have resources in multiple Availability Zones and they share one NAT gateway, and if the NAT gateway's Availability Zone is down, resources in the other Availability Zones lose internet access. To create an Availability Zone-independent architecture, create a NAT gateway in each Availability Zone and configure your routing to ensure that resources use the NAT gateway in the same Availability Zone.

How NAT gateway works:

The following diagram illustrates the architecture of a VPC with a NAT gateway. The main route table sends internet traffic from the instances in the private subnet to the NAT gateway. The NAT gateway sends the traffic to the internet gateway using the NAT gateway's Elastic IP address as the source IP address.



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

Incorrect options:

Set up three NAT gateways, one in each private subnet in each AZ. Create a custom route table for each AZ that forwards non-local traffic to the NAT gateway in its AZ - NAT gateways need to be set up in public subnets, so this option is incorrect.

Set up three Internet gateways, one in each private subnet in each AZ. Create a custom route table for each AZ that forwards non-local traffic to the Internet gateway in its AZ - Internet gateways cannot be provisioned in private subnets of a VPC.

Set up three egress-only internet gateways, one in each public subnet in each AZ. Create a custom route table for each AZ that forwards non-local traffic to the egress-only internet gateway in its AZ - 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. The given use-case is for IPv4 traffic, hence an Egress-only Internet gateway is not an option.

Reference:

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

Domain

Design Secure Architectures

Question 45Skipped

A company has a hybrid cloud structure for its on-premises data center and AWS Cloud infrastructure. The company wants to build a web log archival solution such that only the most frequently accessed logs are available as cached data locally while backing up all logs on Amazon S3.

As a solutions architect, which of the following solutions would you recommend for this use-case?

Use AWS Volume Gateway - Stored Volume - to store the most frequently accessed logs locally for low-latency access while storing the full volume with all logs in its Amazon S3 service bucket

Correct answer

Use AWS Volume Gateway - Cached Volume - to store the most frequently accessed logs locally for low-latency access while storing the full volume with all logs in its Amazon S3 service bucket

Use AWS Direct Connect to store the most frequently accessed logs locally for low-latency access while storing the full backup of logs in an Amazon S3 bucket

Use AWS Snowball Edge Storage Optimized device to store the most frequently accessed logs locally for low-latency access while storing the full backup of logs in an Amazon S3 bucket

Overall explanation

Correct option:

Use AWS Volume Gateway - Cached Volume - to store the most frequently accessed logs locally for low-latency access while storing the full volume with all logs in its Amazon S3 service bucket

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. With cached volumes, the AWS Volume Gateway stores the full volume in its Amazon S3 service bucket, and just the recently accessed data is retained in the gateway's local cache for low-latency access.

Incorrect options:

Use AWS Direct Connect to store the most frequently accessed logs locally for low-latency access while storing the full backup of logs in an Amazon S3 bucket - AWS Direct Connect lets you establish a dedicated network connection between your network and one of the AWS Direct Connect locations. AWS Direct connect cannot be used to store the most frequently accessed logs locally for low-latency access.

Use AWS Volume Gateway - Stored Volume - to store the most frequently accessed logs locally for low-latency access while storing the full volume with all logs in its Amazon S3 service bucket - With stored volumes, your entire data volume is available locally in the gateway, for fast read access. Volume Gateway also maintains an asynchronous copy of your stored volume in the service's Amazon S3 bucket. This does not fit the requirements per the given use-case, hence this option is not correct.

Use AWS Snowball Edge Storage Optimized device to store the most frequently accessed logs locally for low-latency access while storing the full backup of logs in an Amazon S3 bucket - You can use Snowball Edge Storage Optimized device to securely and quickly transfer dozens of terabytes to petabytes of data to AWS. Snowball Edge Storage Optimized device cannot be used to store the most frequently accessed logs locally for low-latency access.

Reference:

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

Domain

Design High-Performing Architectures

Question 46Skipped

An e-commerce company runs its web application on Amazon EC2 instances in an Auto Scaling group and it's configured to handle consumer orders in an Amazon Simple Queue Service (Amazon SQS) queue for downstream processing. The DevOps team has observed that the performance of the application goes down in case of a sudden spike in orders received.

As a solutions architect, which of the following solutions would you recommend to address this use-case?

Correct answer

Use a target tracking scaling policy based on a custom Amazon SQS queue metric

Use a step scaling policy based on a custom Amazon SQS queue metric

Use a scheduled scaling policy based on a custom Amazon SQS queue metric

Use a simple scaling policy based on a custom Amazon SQS queue metric

Overall explanation

Correct option:

Use a target tracking scaling policy based on a custom Amazon SQS queue metric

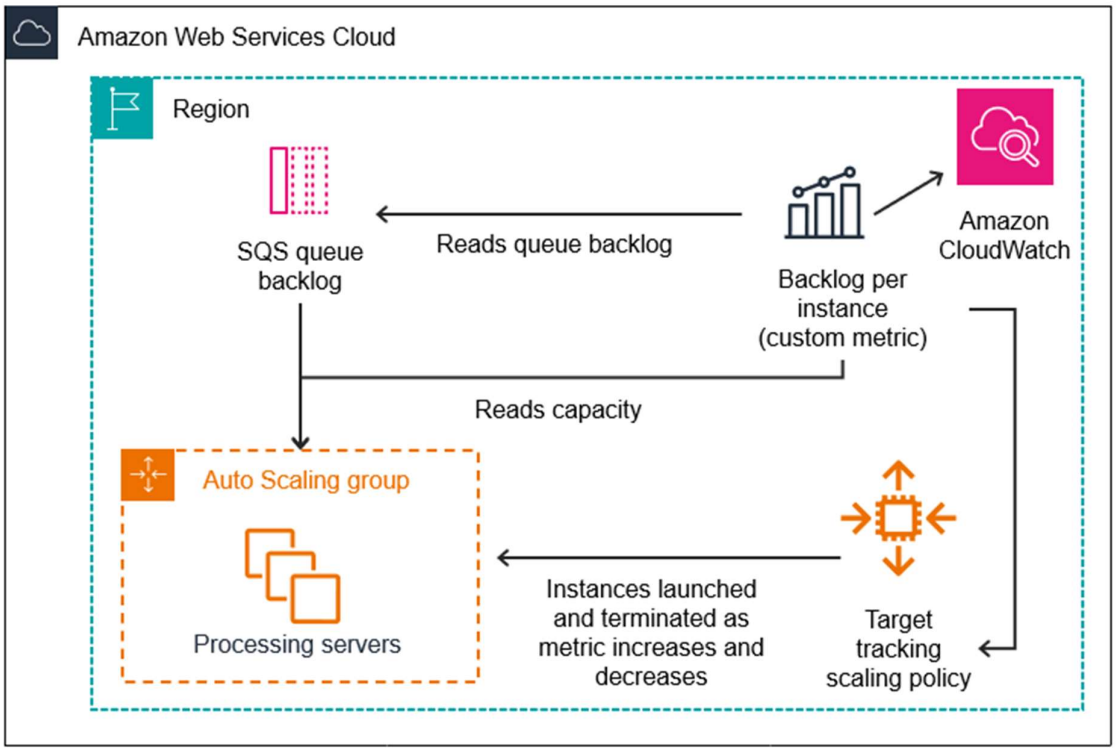
If you use a target tracking scaling policy based on a custom Amazon SQS queue metric, dynamic scaling can adjust to the demand curve of your application more effectively. You may use an existing CloudWatch Amazon SQS metric like ApproximateNumberOfMessagesVisible for target tracking but you could still face an issue so that the number of messages in the queue might not change proportionally to the size of the Auto Scaling group that processes messages from the queue. The solution is to use a backlog per instance metric with the target value being the acceptable backlog per instance to maintain.

To calculate your backlog per instance, divide the ApproximateNumberOfMessages queue attribute by the number of instances in the InService state for the Auto Scaling group. Then set a target value for the Acceptable backlog per instance.

To illustrate with an example, let's say that the current ApproximateNumberOfMessages is 1500 and the fleet's running capacity is 10. If the average processing time is 0.1 seconds for each message and the longest acceptable latency is 10 seconds, then the acceptable backlog per instance is 10 / 0.1, which equals 100. This means that 100 is the target value for your target

tracking policy. If the backlog per instance is currently at 150 (1500 / 10), your fleet scales out, and it scales out by five instances to maintain proportion to the target value.

Scaling Based on Amazon SQS:



via - https://docs.aws.amazon.com/autoscaling/ec2/userguide/as-using-sqs-queue.html

Incorrect options:

Use a simple scaling policy based on a custom Amazon SQS queue metric - With simple scaling, you choose scaling metrics and threshold values for the Amazon CloudWatch alarms that trigger the scaling process. The main issue with simple scaling is that after a scaling activity is started, the policy must wait for the scaling activity or health check replacement to complete and the cooldown period to expire before responding to additional alarms. This implies that the application would not be able to react quickly to sudden spikes in orders.

Use a step scaling policy based on a custom Amazon SQS queue metric - With step scaling, you choose scaling metrics and threshold values for the Amazon CloudWatch alarms that trigger the scaling process. When step adjustments are applied, they increase or decrease the current capacity of your Auto Scaling group, and the adjustments vary based on the size of the alarm breach. For the given use-case, step scaling would try to approximate the correct number of instances by increasing/decreasing the steps as per the policy. This is not as efficient as the target tracking policy where you can calculate the exact number of instances required to handle the spike in orders.

Use a scheduled scaling policy based on a custom Amazon SQS queue metric - Scheduled scaling allows you to set your scaling schedule. For example, let's say that every week the traffic to your web application starts to increase on Wednesday, remains high on Thursday, and starts to decrease on Friday. You can plan your scaling actions based on the predictable traffic

patterns of your web application. Scaling actions are performed automatically as a function of time and date. You cannot use scheduled scaling policies to address the sudden spike in orders.

Reference:

https://docs.aws.amazon.com/autoscaling/ec2/userguide/as-using-sqs-queue.html

Domain

Design Secure Architectures

Question 47Skipped

A company has a license-based, expensive, legacy commercial database solution deployed at its on-premises data center. The company wants to migrate this database to a more efficient, open-source, and cost-effective option on AWS Cloud. The CTO at the company wants a solution that can handle complex database configurations such as secondary indexes, foreign keys, and stored procedures.

As a solutions architect, which of the following AWS services should be combined to handle this use-case? (Select two)

Correct selection

AWS Schema Conversion Tool (AWS SCT)

Basic Schema Copy

AWS Glue

AWS Snowball Edge

Correct selection

AWS Database Migration Service (AWS DMS)

Overall explanation

Correct options:

AWS Schema Conversion Tool (AWS SCT)

AWS Database Migration Service (AWS DMS)

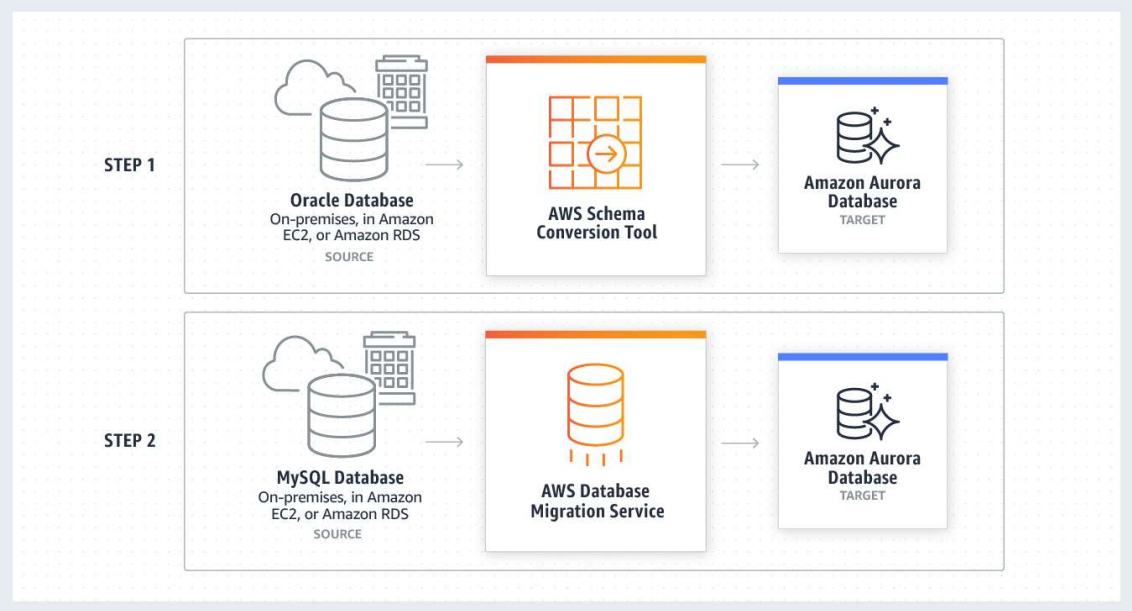
AWS Database Migration Service helps you migrate databases to AWS quickly and securely. The source database remains fully operational during the migration, minimizing downtime to applications that rely on the database. AWS Database Migration Service supports homogeneous migrations such as Oracle to Oracle, as well as heterogeneous migrations between different database platforms, such as Oracle or Microsoft SQL Server to Amazon Aurora.

Given the use-case where the CTO at the company wants to move away from license-based, expensive, legacy commercial database solutions deployed at the on-premises data center to more efficient, open-source, and cost-effective options on AWS Cloud, this is an example of heterogeneous database migrations.

For such a scenario, the source and target databases engines are different, like in the case of Oracle to Amazon Aurora, Oracle to PostgreSQL, or Microsoft SQL Server to MySQL migrations. In this case, the schema structure, data types, and database code of source and target databases can be quite different, requiring a schema and code transformation before the data migration starts.

That makes heterogeneous migrations a two-step process. First use the AWS Schema Conversion Tool to convert the source schema and code to match that of the target database, and then use the AWS Database Migration Service to migrate data from the source database to the target database. All the required data type conversions will automatically be done by the AWS Database Migration Service during the migration. The source database can be located on your on-premises environment outside of AWS, running on an Amazon EC2 instance, or it can be an Amazon RDS database. The target can be a database in Amazon EC2 or Amazon RDS.

Heterogeneous Database Migrations:



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

Incorrect options:

AWS Snowball Edge - 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 TB of usable HDD storage, 40 vCPUs, 1 TB of SATA SSD storage, and up to 40 Gb network connectivity to address large scale data transfer and pre-processing use cases. As each Snowball Edge Storage Optimized device can handle 80TB of data, you can order 10 such devices to take care of the data transfer for all applications. The original 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 Snowball device had 80TB of storage space. AWS Snowball Edge cannot be used for database migrations.

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. Therefore, it cannot be used for database migrations.

Basic Schema Copy - To quickly migrate a database schema to your target instance you can rely on the Basic Schema Copy feature of AWS Database Migration Service. Basic Schema Copy will automatically create tables and primary keys in the target instance if the target does not already contain tables with the same names. Basic Schema Copy is great for doing a test migration, or when you are migrating databases heterogeneously e.g. Oracle to MySQL or SQL Server to Oracle. Basic Schema Copy will not migrate secondary indexes, foreign keys or stored procedures. When you need to use a more customizable schema migration process (e.g. when you are migrating your production database and need to move your stored procedures and secondary database objects), you must use the AWS Schema Conversion Tool.

References:

https://aws.amazon.com/dms/

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

https://aws.amazon.com/dms/schema-conversion-tool/

Domain

Design Cost-Optimized Architectures

Question 48Skipped

The development team at a retail company wants to optimize the cost of Amazon EC2 instances. The team wants to move certain nightly batch jobs to spot instances. The team has hired you as a solutions architect to provide the initial guidance.

Which of the following would you identify as CORRECT regarding the capabilities of spot instances? (Select three)

Correct selection

Spot Fleets can maintain target capacity by launching replacement instances after Spot Instances in the fleet are terminated

Correct selection

When you cancel an active spot request, it does not terminate the associated instance

Correct selection

If a spot request is persistent, then it is opened again after your Spot Instance is interrupted

When you cancel an active spot request, it terminates the associated instance as well

If a spot request is persistent, then it is opened again after you stop the Spot Instance

Spot Fleets cannot maintain target capacity by launching replacement instances after Spot Instances in the fleet are terminated

Overall explanation

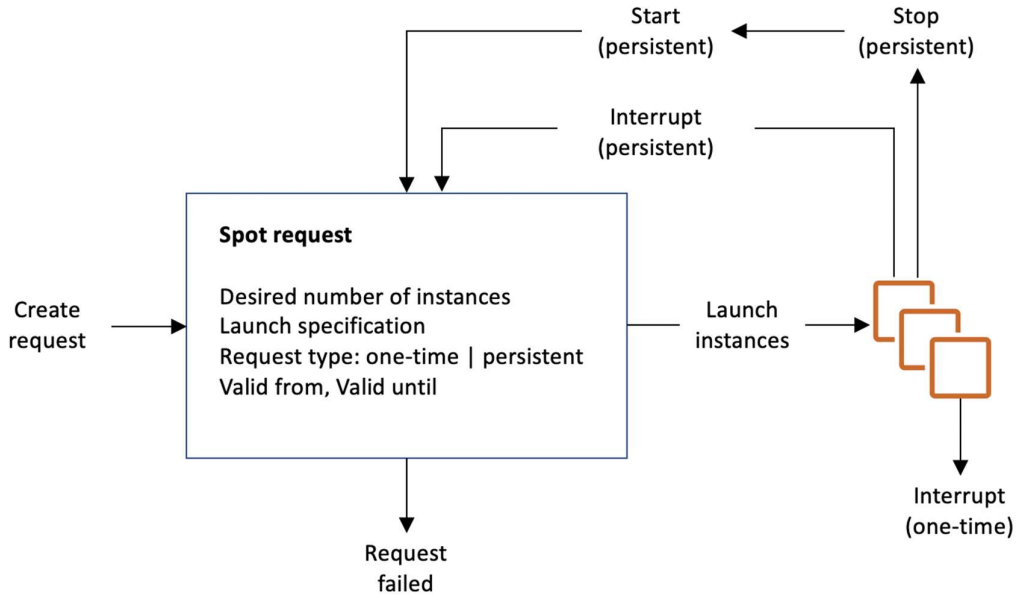
Correct options:

If a spot request is persistent, then it is opened again after your Spot Instance is interrupted

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 is set by Amazon EC2 and adjusted gradually based on the long-term supply of and demand for Spot Instances.

A Spot Instance request is either one-time or persistent. If the spot request is persistent, the request is opened again after your Spot Instance is interrupted. If the request is persistent and you stop your Spot Instance, the request only opens after you start your Spot Instance.

How Spot requests work:



via - https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/spot-requests.html

Spot Fleets can maintain target capacity by launching replacement instances after Spot Instances in the fleet are terminated

A Spot Fleet is a set of Spot Instances and optionally On-Demand Instances that is launched based on criteria that you specify. The Spot Fleet selects the Spot capacity pools that meet your needs and launches Spot Instances to meet the target capacity for the fleet. By default, Spot Fleets are set to maintain target capacity by launching replacement instances after Spot Instances in the fleet are terminated. You can submit a Spot Fleet as a one-time request, which does not persist after the instances have been terminated. You can include On-Demand Instance requests in a Spot Fleet request.

When you cancel an active spot request, it does not terminate the associated instance

If your Spot Instance request is active and has an associated running Spot Instance, or your Spot Instance request is disabled and has an associated stopped Spot Instance, canceling the request does not terminate the instance; you must terminate the running Spot Instance manually. Moreover, to cancel a persistent Spot request and terminate its Spot Instances, you must cancel the Spot request first and then terminate the Spot Instances.

Incorrect options:

When you cancel an active spot request, it terminates the associated instance as well - If your Spot Instance request is active and has an associated running Spot Instance, then canceling the request does not terminate the instance; you must terminate the running Spot Instance manually. So, this option is incorrect.

If a spot request is persistent, then it is opened again after you stop the Spot Instance - If the request is persistent and you stop your Spot Instance, the request only opens after you start your Spot Instance. So, this option is incorrect.

Spot Fleets cannot maintain target capacity by launching replacement instances after Spot Instances in the fleet are terminated - As mentioned above, Spot Fleets can maintain target capacity by launching replacement instances after Spot Instances in the fleet are terminated.

References:

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

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

Domain

Design Cost-Optimized Architectures

Question 49Skipped

A gaming company uses Application Load Balancers in front of Amazon EC2 instances for different services and microservices. The architecture has now become complex with too many Application Load Balancers in multiple AWS Regions. Security updates, firewall configurations, and traffic routing logic have become complex with too many IP addresses and configurations.

The company is looking at an easy and effective way to bring down the number of IP addresses allowed by the firewall and easily manage the entire network infrastructure. Which of these options represents an appropriate solution for this requirement?

Assign an Elastic IP to an Auto Scaling Group (ASG), and set up multiple Amazon EC2 instances to run behind the Auto Scaling Groups, for each of the Regions

Correct answer

Launch AWS Global Accelerator and create endpoints for all the Regions. Register the Application Load Balancers of each Region to the corresponding endpoints

Configure Elastic IPs for each of the Application Load Balancers in each Region

Set up a Network Load Balancer with elastic IP address. Register the private IPs of all the Application Load Balancers as targets of this Network Load Balancer

Overall explanation

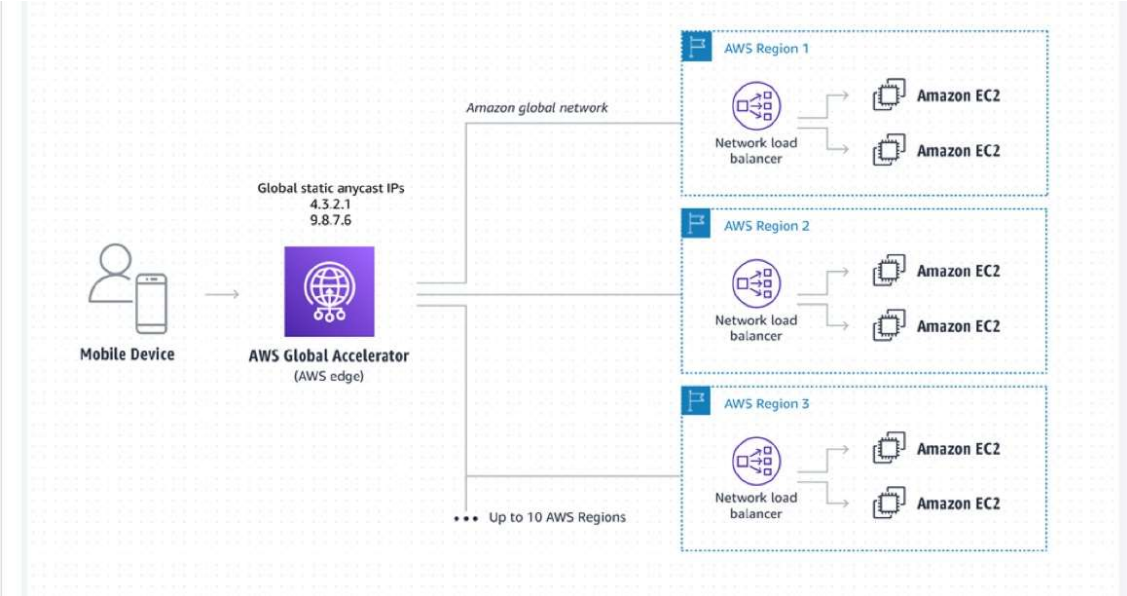
Correct option:

Launch AWS Global Accelerator and create endpoints for all the Regions. Register the Application Load Balancers of each Region to the corresponding endpoints

AWS Global Accelerator is a networking service that sends your user's traffic through Amazon Web Service's global network infrastructure, improving your internet user performance by up to 60%. When the internet is congested, Global Accelerator's automatic routing optimizations will help keep your packet loss, jitter, and latency consistently low.

With AWS Global Accelerator, you are provided two global static customer-facing IPs to simplify traffic management. On the back end, add or remove your AWS application origins, such as Network Load Balancers, Application Load Balancers, elastic IP address (EIP), and Amazon EC2 Instances, without making user-facing changes. To mitigate endpoint failure, AWS Global Accelerator automatically re-routes your traffic to your nearest healthy available endpoint.

Simplified and resilient traffic routing for multi-Region applications:



via - https://aws.amazon.com/global-accelerator/

Incorrect options:

Configure Elastic IPs for each of the Application Load Balancers in each Region - An Application Load Balancer cannot be assigned an Elastic IP address (static IP address).

Set up a Network Load Balancer with elastic IP address. Register the private IPs of all the Application Load Balancers as targets of this Network Load Balancer - A Network Load Balancer can be configured to take an Elastic IP address. However, with hundreds of Application Load Balancers and Network Load Balancers, the solution will be equally cumbersome to manage.

Assign an Elastic IP to an Auto Scaling Group (ASG), and set up multiple Amazon EC2 instances to run behind the Auto Scaling Groups, for each of the Regions - You cannot

assign an elastic IP address to an Auto Scaling Group (ASG), since ASG just manages a collection of Amazon EC2 instances.

References:

https://aws.amazon.com/global-accelerator/

https://aws.amazon.com/blogs/networking-and-content-delivery/using-static-ip-addresses-for-application-load-balancers/

Domain

Design High-Performing Architectures

Question 50Skipped

A DevOps engineer at an IT company just upgraded an Amazon EC2 instance type from t2.nano (0.5G of RAM, 1 vCPU) to u-12tb1.metal (12.3 TB of RAM, 448 vCPUs). How would you categorize this upgrade?

Correct answer

This is a scale up example of vertical scalability

This is a scale up example of horizontal scalability

This is a scale out example of vertical scalability

This is an example of high availability

Overall explanation

Correct option:

This is a scale up example of vertical scalability

Vertical scalability means increasing the size of the instance. For example, your application runs on a t2.micro. Scaling up that application vertically means running it on a larger instance such as t2.large. Scaling down that application vertically means running it on a smaller instance such as t2.nano. Scalability is very common for non-distributed systems, such as a database. There's usually a limit to how much you can vertically scale (hardware limit). In this case, as the instance type was upgraded from t2.nano to u-12tb1.metal, this is a scale up example of vertical scalability.

Incorrect options:

This is a scale up example of horizontal scalability - Horizontal Scalability means increasing the number of instances/systems for your application. When you increase the number of instances, it's called scale out whereas if you decrease the number of instances, it's called scale-in. Scale up is used in conjunction with vertical scaling and not with horizontal scaling. Hence this is incorrect.

This is a scale out example of vertical scalability - Scale out is used in conjunction with horizontal scaling and not with vertical scaling. Hence this is incorrect.

This is an example of high availability - High availability means running your application/system in at least 2 data centers (== Availability Zones). The goal of high availability

is to survive a data center loss. An example of High Availability is when you run instances for the same application across multi AZ. This option has been added as a distractor.

Domain

Design High-Performing Architectures

Question 51Skipped

A company recently experienced a database outage in its on-premises data center. The company now wants to migrate to a reliable database solution on AWS that minimizes data loss and stores every transaction on at least two nodes.

Which of the following solutions meets these requirements?

Set up an Amazon RDS MySQL DB instance and then create a read replica in another Availability Zone that synchronously replicates the data

Correct answer

Set up an Amazon RDS MySQL DB instance with Multi-AZ functionality enabled to synchronously replicate the data

Set up an Amazon RDS MySQL DB instance and then create a read replica in a separate AWS Region that synchronously replicates the data

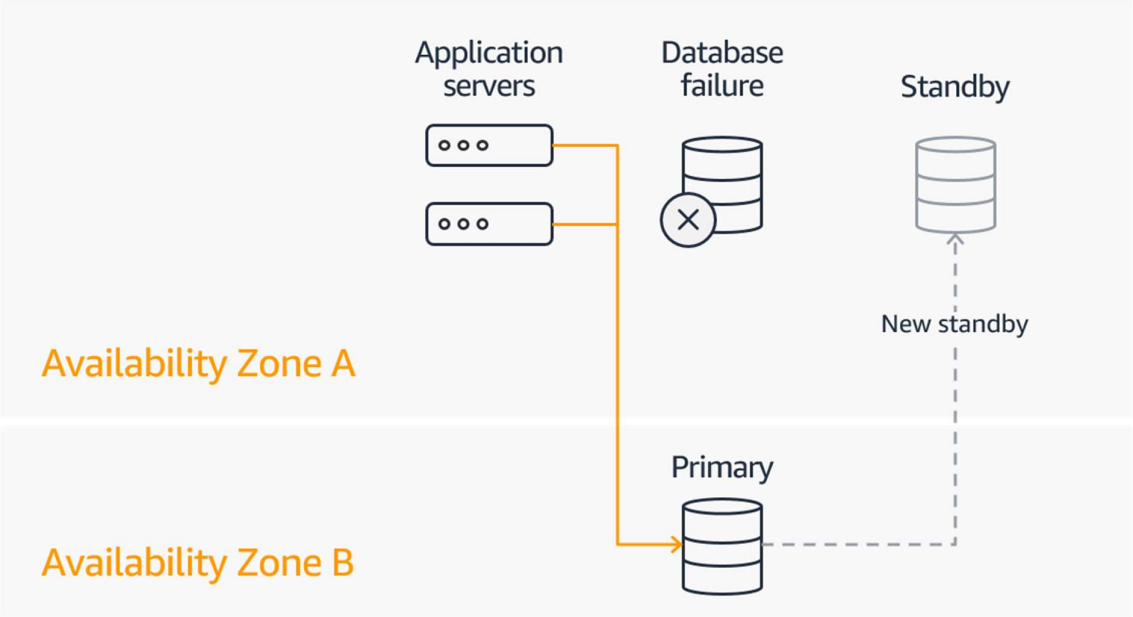
Set up an Amazon EC2 instance with a MySQL DB engine installed that triggers an AWS Lambda function to synchronously replicate the data to an Amazon RDS MySQL DB instance

Overall explanation

Correct option:

Set up an Amazon RDS MySQL DB instance with Multi-AZ functionality enabled to synchronously replicate the data

When you provision an RDS 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. Running a DB instance with high availability can enhance availability during planned system maintenance, and help protect your databases against DB instance failure and Availability Zone disruption. In the event of a planned or unplanned outage of your DB instance, Amazon RDS automatically switches to a standby replica in another Availability Zone if you have enabled Multi-AZ. The time it takes for the failover to complete depends on the database activity and other conditions at the time the primary DB instance became unavailable. Failover times are typically 60–120 seconds.



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

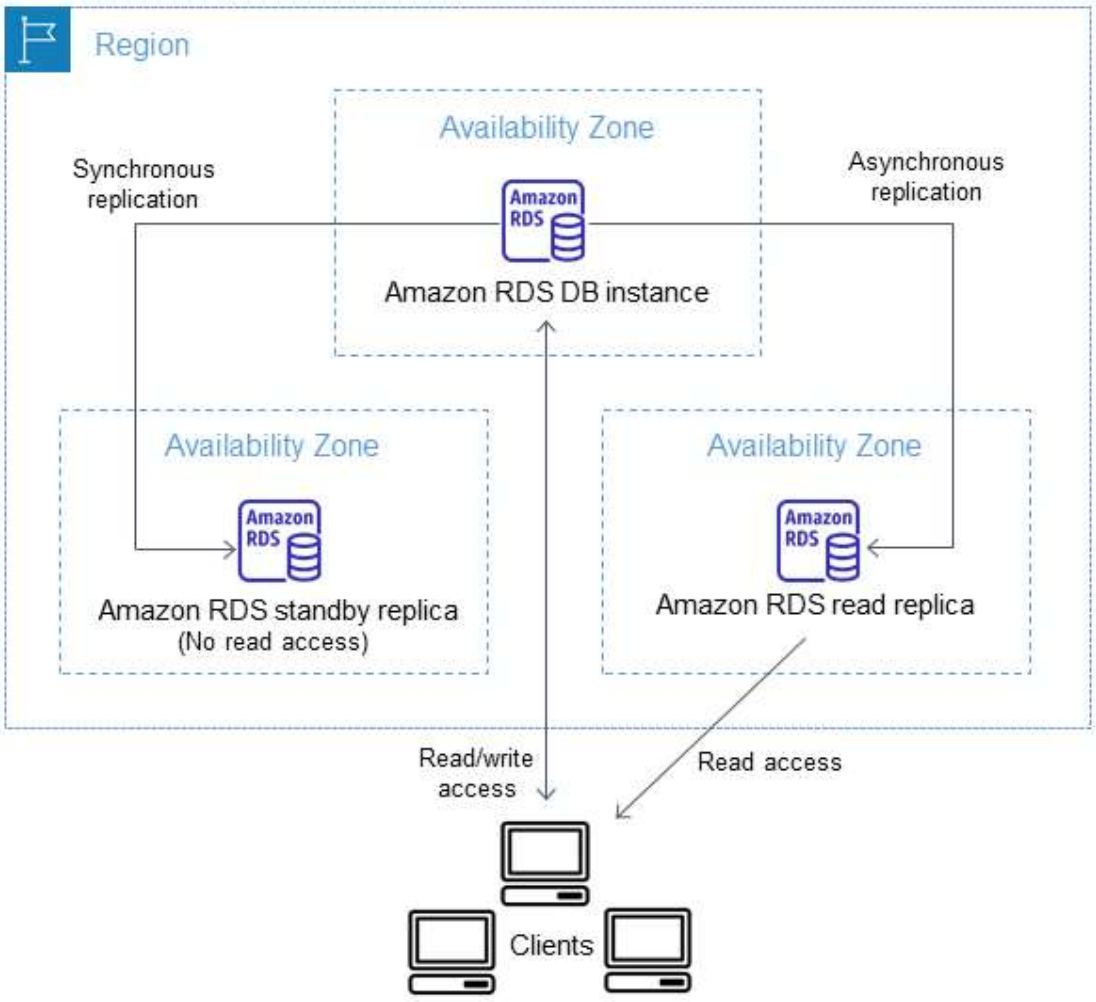
Incorrect options:

Set up an Amazon RDS MySQL DB instance and then create a read replica in another Availability Zone that synchronously replicates the data

Set up an Amazon RDS MySQL DB instance and then create a read replica in a separate AWS Region that synchronously replicates the data

Amazon RDS uses the MariaDB, Microsoft SQL Server, MySQL, Oracle, and PostgreSQL DB engines' built-in replication functionality to create a special type of DB instance called a read replica from a source DB instance. The source DB instance becomes the primary DB instance. Updates made to the primary DB instance are asynchronously copied to the read replica. You can reduce the load on your primary DB instance by routing read queries from your applications to the read replica. Using read replicas, you can elastically scale out beyond the capacity constraints of a single DB instance for read-heavy database workloads.

Both these options talk about creating a read replica that **synchronously** replicates the data, but in reality, any updates made to the primary DB instance are **asynchronously** copied to the read replica. So both these options are incorrect.



via - https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/USER_ReadRepl.html

Set up an Amazon EC2 instance with a MySQL DB engine installed that triggers an AWS Lambda function to synchronously replicate the data to an Amazon RDS MySQL DB instance - Setting up a database on an Amazon EC2 instance would not be reliable as you would have to monitor and manage the underlying Amazon EC2 instance for any issues or outages. In addition, using AWS Lambda to replicate the data from EC2 based MySQL DB to an Amazon RDS MySQL DB would make the solution really complex since the same functionality can be achieved out-of-the-box using RDS Multi-AZ configuration.

References:

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

https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/USER_ReadRepl.html

Domain

Design Resilient Architectures

Question 52Skipped

A retail organization is moving some of its on-premises data to AWS Cloud. The DevOps team at the organization has set up an AWS Managed IPSec VPN Connection between their remote on-premises network and their Amazon VPC over the internet.

Which of the following represents the correct configuration for the IPSec VPN Connection?

Create a Customer Gateway on both the AWS side of the VPN as well as the on-premises side of the VPN

Create a virtual private gateway (VGW) on the on-premises side of the VPN and a Customer Gateway on the AWS side of the VPN

Correct answer

Create a virtual private gateway (VGW) on the AWS side of the VPN and a Customer Gateway on the on-premises side of the VPN

Create a virtual private gateway (VGW) on both the AWS side of the VPN as well as the on-premises side of the VPN

Overall explanation

Correct option:

Create a virtual private gateway (VGW) on the AWS side of the VPN and a Customer Gateway on the on-premises side of the VPN

Amazon VPC provides the facility to create an IPsec VPN connection (also known as AWS site-to-site VPN) between remote customer networks and their Amazon VPC over the internet. The following are the key concepts for a site-to-site VPN:

Virtual private gateway: A virtual private gateway (VGW), also known as a VPN Gateway is the endpoint on the AWS VPC side of your VPN connection.

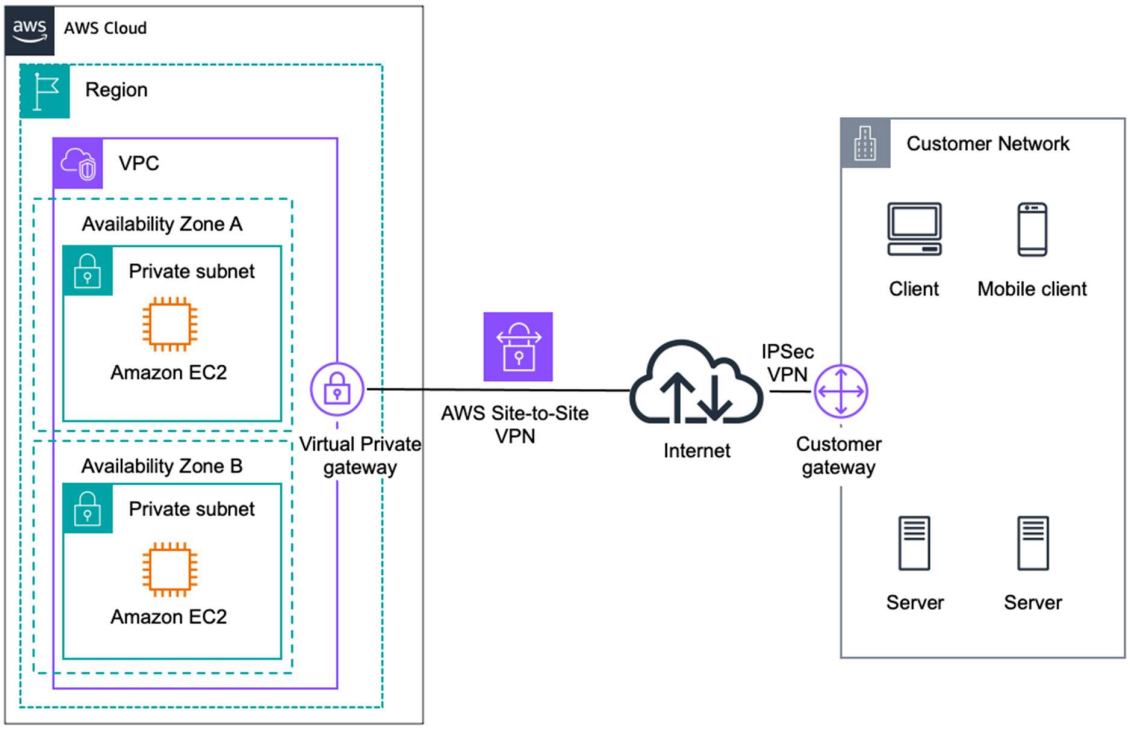
VPN connection: A secure connection between your on-premises equipment and your VPCs.

VPN tunnel: An encrypted link where data can pass from the customer network to or from AWS.

Customer Gateway: An AWS resource that provides information to AWS about your Customer Gateway device.

Customer Gateway device: A physical device or software application on the customer side of the Site-to-Site VPN connection.

AWS Managed IPSec VPN



via - https://docs.aws.amazon.com/whitepapers/latest/aws-vpc-connectivity-options/aws-managed-vpn-network-to-amazon.html

Incorrect options:

Create a virtual private gateway (VGW) on the on-premises side of the VPN and a Customer Gateway on the AWS side of the VPN - You need to create a virtual private gateway (VGW) on the AWS side of the VPN and a Customer Gateway on the on-premises side of the VPN. Therefore, this option is wrong.

Create a Customer Gateway on both the AWS side of the VPN as well as the on-premises side of the VPN - You need to create a virtual private gateway (VGW) on the AWS side of the VPN and a Customer Gateway on the on-premises side of the VPN. Therefore, this option is wrong.

Create a virtual private gateway (VGW) on both the AWS side of the VPN as well as the on-premises side of the VPN - You need to create a virtual private gateway (VGW) on the AWS side of the VPN and a Customer Gateway on the on-premises side of the VPN. Therefore, this option is wrong.

References:

https://docs.aws.amazon.com/whitepapers/latest/aws-vpc-connectivity-options/aws-managed-vpn-network-to-amazon.html

https://docs.aws.amazon.com/vpn/latest/s2svpn/VPC_VPN.html

Domain

Design Secure Architectures

Question 53Skipped

A financial services company wants to identify any sensitive data stored on its Amazon S3 buckets. The company also wants to monitor and protect all data stored on Amazon S3 against any malicious activity.

As a solutions architect, which of the following solutions would you recommend to help address the given requirements?

Use Amazon Macie to monitor any malicious activity on data stored in Amazon S3. Use Amazon GuardDuty to identify any sensitive data stored on Amazon S3

Correct answer

Use Amazon GuardDuty to monitor any malicious activity on data stored in Amazon S3. Use Amazon Macie to identify any sensitive data stored on Amazon S3

Use Amazon Macie to monitor any malicious activity on data stored in Amazon S3 as well as to identify any sensitive data stored on Amazon S3

Use Amazon GuardDuty to monitor any malicious activity on data stored in Amazon S3 as well as to identify any sensitive data stored on Amazon S3

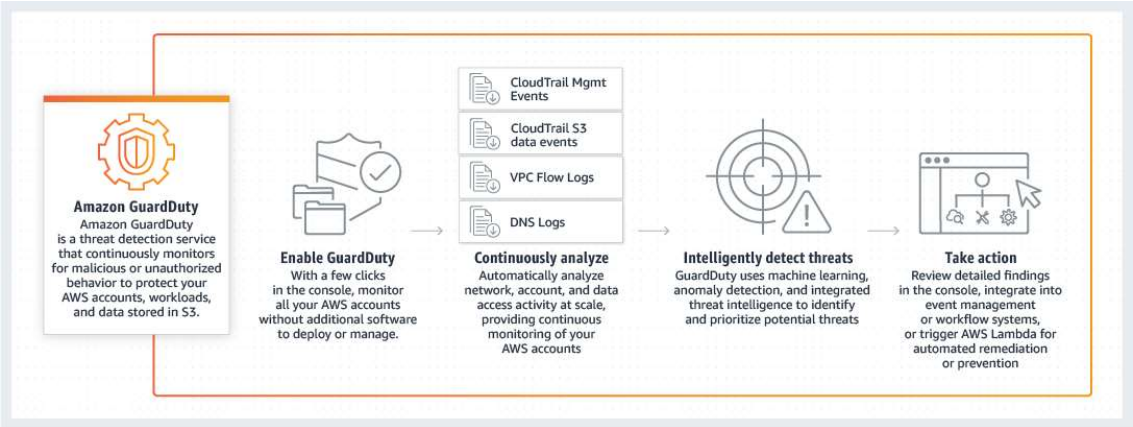
Overall explanation

Correct option:

Use Amazon GuardDuty to monitor any malicious activity on data stored in Amazon S3. Use Amazon Macie to identify any sensitive data stored on Amazon S3

Amazon GuardDuty offers threat detection that enables you to continuously monitor and protect your AWS accounts, workloads, and data stored in Amazon S3. GuardDuty analyzes continuous streams of meta-data generated from your account and network activity found in AWS CloudTrail Events, Amazon VPC Flow Logs, and DNS Logs. It also uses integrated threat intelligence such as known malicious IP addresses, anomaly detection, and machine learning to identify threats more accurately.

How Amazon GuardDuty works:

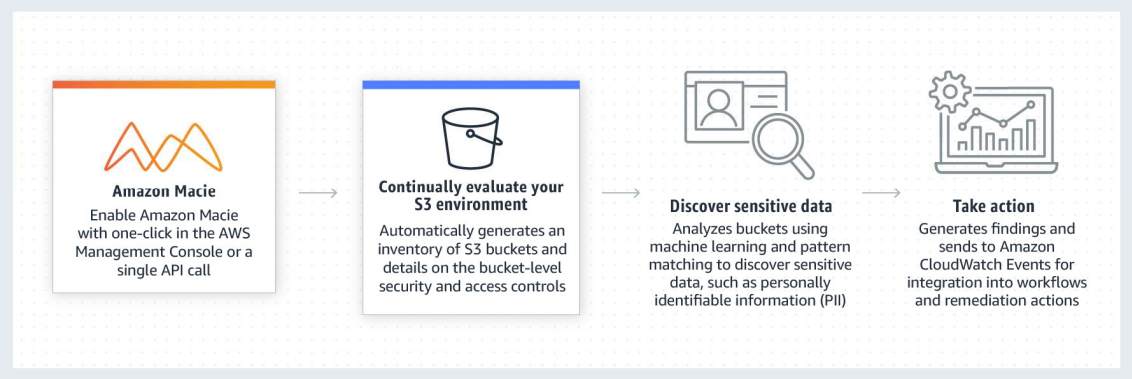


via - https://aws.amazon.com/guardduty/

Amazon Macie is a fully managed data security and data privacy service that uses machine learning and pattern matching to discover and protect your sensitive data on Amazon S3. Macie

automatically detects a large and growing list of sensitive data types, including personally identifiable information (PII) such as names, addresses, and credit card numbers. It also gives you constant visibility of the data security and data privacy of your data stored in Amazon S3.

How Amazon Macie works:



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

Incorrect options:

Use Amazon GuardDuty to monitor any malicious activity on data stored in Amazon S3 as well as to identify any sensitive data stored on Amazon S3

Use Amazon Macie to monitor any malicious activity on data stored in Amazon S3 as well as to identify any sensitive data stored on Amazon S3

Use Amazon Macie to monitor any malicious activity on data stored in Amazon S3. Use Amazon GuardDuty to identify any sensitive data stored on Amazon S3

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

References:

https://aws.amazon.com/guardduty/

https://aws.amazon.com/macie/

Domain

Design Secure Architectures

Question 54Skipped

A data analytics company manages an application that stores user data in a Amazon DynamoDB table. The development team has observed that once in a while, the application writes corrupted data in the Amazon DynamoDB table. As soon as the issue is detected, the team needs to remove the corrupted data at the earliest.

What do you recommend?

Use Amazon DynamoDB on-demand backup to restore the table to the state just before corrupted data was written

Use Amazon DynamoDB Streams to restore the table to the state just before corrupted data was written

Correct answer

Use Amazon DynamoDB point in time recovery to restore the table to the state just before corrupted data was written

Configure the Amazon DynamoDB table as a global table and point the application to use the table from another AWS region that has no corrupted data

Overall explanation

Correct option:

Use Amazon DynamoDB point in time recovery to restore the table to the state just before corrupted data was written

Amazon DynamoDB enables you to back up your table data continuously by using point-in-time recovery (PITR). When you enable PITR, DynamoDB backs up your table data automatically with per-second granularity so that you can restore to any given second in the preceding 35 days.

PITR helps protect you against accidental writes and deletes. For example, if a test script writes accidentally to a production DynamoDB table or someone mistakenly issues a "DeleteItem" call, PITR has you covered.

Incorrect options:

Use Amazon DynamoDB on-demand backup to restore the table to the state just before corrupted data was written - The on-demand backup and restore process scales without degrading the performance or availability of your applications. It uses a new and unique distributed technology that lets you complete backups in seconds regardless of table size. You can create backups that are consistent within seconds across thousands of partitions without worrying about schedules or long-running backup processes. All on-demand backups are cataloged, discoverable, and retained until they are explicitly deleted.

On-demand backup is created upon request. So this option is not correct since an on-demand backup cannot be created pre-emptively to handle data corruption issues that happen once in a while.

Configure the Amazon DynamoDB table as a global table and point the application to use the table from another AWS region that has no corrupted data - 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 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.

Any changes made to any item in any replica table are replicated to all the other replicas within the same global table. In a global table, a newly written item is usually propagated to all replica tables within a second. With a global table, each replica table stores the same set of data items.

Amazon DynamoDB does not support partial replication of only some of the items. If applications update the same item in different Regions at about the same time, conflicts can arise. To help ensure eventual consistency, Amazon DynamoDB global tables use a last-writer-wins reconciliation between concurrent updates, in which DynamoDB makes its best effort to determine the last writer. With this conflict resolution mechanism, all replicas agree on the latest update and converge toward a state in which they all have identical data.

Global tables replicate your Amazon DynamoDB tables automatically across your choice of AWS Regions. This option has been added as a distractor since you cannot point the application to use the table from another AWS region, since there is no "other" table in another region. It's just a single logical Global table.

Use Amazon DynamoDB Streams to restore the table to the state just before corrupted data was written - Amazon DynamoDB Streams captures a time-ordered sequence of item-level modifications in any Amazon DynamoDB table and stores this information in a log for up to 24 hours. Applications can access this log and view the data items as they appeared before and after they were modified, in near-real time. A DynamoDB stream is an ordered flow of information about changes to items in a Amazon DynamoDB table. When you enable a stream on a table, DynamoDB captures information about every modification to data items in the table.

Amazon DynamoDB Streams writes stream records in near-real time so that you can build applications that consume these streams and take action based on the contents. It will take considerable effort and custom coding to reliably rebuild table data to the state just before any corrupted data was written. So this option is not the best fit.

References:

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

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

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

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

Domain

Design Resilient Architectures

Question 55Skipped

A startup has created a new web application for users to complete a risk assessment survey for COVID-19 symptoms via a self-administered questionnaire. The startup has purchased the domain covid19survey.com using Amazon Route 53. The web development team would like to create Amazon Route 53 record so that all traffic for covid19survey.com is routed to www.covid19survey.com.

As a solutions architect, which of the following is the MOST cost-effective solution that you would recommend to the web development team?

Create an NS record for covid19survey.com that routes traffic to www.covid19survey.com

Create an MX record for covid19survey.com that routes traffic to www.covid19survey.com

Create a CNAME record for covid19survey.com that routes traffic to www.covid19survey.com

Correct answer

Create an alias record for covid19survey.com that routes traffic to www.covid19survey.com

Overall explanation

Correct option:

Create an alias record for covid19survey.com that routes traffic to www.covid19survey.com

Alias records provide Amazon Route 53–specific extension to DNS functionality. Alias records let you route traffic to selected AWS resources, such as Amazon CloudFront distributions and Amazon S3 buckets.

You can create an alias record at the top node of a DNS namespace, also known as the zone apex, however, you cannot create a CNAME record for the top node of the DNS namespace. So, if you register the DNS name covid19survey.com, the zone apex is covid19survey.com. You can't create a CNAME record for covid19survey.com, but you can create an alias record for covid19survey.com that routes traffic to www.covid19survey.com.

Exam Alert:

You should also note that Amazon Route 53 doesn't charge for alias queries to AWS resources but Route 53 does charge for CNAME queries. Additionally, an alias record can only redirect queries to selected AWS resources such as Amazon S3 buckets, Amazon CloudFront distributions, and another record in the same Amazon Route 53 hosted zone; however a CNAME record can redirect DNS queries to any DNS record. So, you can create a CNAME record that redirects queries from app.covid19survey.com to app.covid19survey.net.

Incorrect options:

Create a CNAME record for covid19survey.com that routes traffic to www.covid19survey.com - You cannot create a CNAME record for the top node of the DNS namespace, so this option is incorrect.

Create an MX record for covid19survey.com that routes traffic to www.covid19survey.com - An MX record specifies the names of your mail servers and, if you have two or more mail servers, the priority order. It cannot be used to create Amazon Route 53 record to route traffic for the top node of the DNS namespace, so this option is incorrect.

Create an NS record for covid19survey.com that routes traffic to www.covid19survey.com - An NS record identifies the name servers for the hosted zone. It cannot be used to create Amazon Route 53 record to route traffic for the top node of the DNS namespace, so this option is incorrect.

References:

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

https://docs.aws.amazon.com/Route53/latest/DeveloperGuide/ResourceRecordTypes.html

Domain

Design Cost-Optimized Architectures

Question 56Skipped

A legacy application is built using a tightly-coupled monolithic architecture. Due to a sharp increase in the number of users, the application performance has degraded. The company now wants to decouple the architecture and adopt AWS microservices architecture. Some of these microservices need to handle fast running processes whereas other microservices need to handle slower processes.

Which of these options would you identify as the right way of connecting these microservices?

Correct answer

Configure Amazon Simple Queue Service (Amazon SQS) queue to decouple microservices running faster processes from the microservices running slower ones

Add Amazon EventBridge to decouple the complex architecture

Use Amazon Simple Notification Service (Amazon SNS) to decouple microservices running faster processes from the microservices running slower ones

Configure Amazon Kinesis Data Streams to decouple microservices running faster processes from the microservices running slower ones

Overall explanation

Correct option:

Configure Amazon Simple Queue Service (Amazon SQS) queue to decouple microservices running faster processes from the microservices running slower ones

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. 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.

Use Amazon SQS to transmit any volume of data, at any level of throughput, without losing messages or requiring other services to be available. Amazon SQS lets you decouple application components so that they run and fail independently, increasing the overall fault tolerance of the system. Multiple copies of every message are stored redundantly across multiple availability zones so that they are available whenever needed. Being able to store the messages and replay them is a very important feature in decoupling the system architecture, as is needed in the current use case.

Incorrect options:

Use Amazon Simple Notification Service (Amazon SNS) to decouple microservices running faster processes from the microservices running slower ones - Amazon SNS follows the

"publish-subscribe" (pub-sub) messaging paradigm, with notifications being delivered to clients using a "push" mechanism. This is an important difference between Amazon SNS and Amazon SQS. Whereas Amazon SQS is a polling mechanism, that gives applications the chance to poll at their own comfort, the push mechanism assumes the other applications are present. For the current requirement, we need messages to be stored till they are processed by the downstream applications. Hence, Amazon SQS is the right choice.

Configure Amazon Kinesis Data Streams to decouple microservices running faster processes from the microservices running slower ones - Amazon Kinesis Data Streams are used for streaming real-time high-volume data. Amazon Kinesis is a publish-subscribe model, used when publisher applications need to publish the same data to different consumers in parallel. Amazon SQS is the right fit for the current use case.

Add Amazon EventBridge to decouple the complex architecture - This event-based service is extremely useful for connecting non-AWS SaaS (Software as a Service) services to AWS services. With Amazon Eventbridge, the downstream application would need to immediately process the events whenever they arrive, thereby making it a tightly coupled scenario. Hence, this option is not correct.

References:

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

Domain

Design Resilient Architectures

Question 57Skipped

A financial services company is looking to move its on-premises IT infrastructure to AWS Cloud. The company has multiple long-term server bound licenses across the application stack and the CTO wants to continue to utilize those licenses while moving to AWS.

As a solutions architect, which of the following would you recommend as the MOST cost-effective solution?

Correct answer

Use Amazon EC2 dedicated hosts

Use Amazon EC2 reserved instances (RI)

Use Amazon EC2 on-demand instances

Use Amazon EC2 dedicated instances

Overall explanation

Correct option:

Use Amazon EC2 dedicated hosts

You can use Dedicated Hosts to launch Amazon EC2 instances on physical servers that are dedicated for your use. Dedicated Hosts give you additional visibility and control over how instances are placed on a physical server, and you can reliably use the same physical server

over time. As a result, Dedicated Hosts enable you to use your existing server-bound software licenses like Windows Server and address corporate compliance and regulatory requirements.

Incorrect options:

Use Amazon EC2 dedicated 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 may share hardware with other instances from the same AWS account that are not dedicated instances. Dedicated instances cannot be used for existing server-bound software licenses.

Use Amazon EC2 on-demand instances

Use Amazon EC2 reserved instances (RI)

Amazon EC2 presents a virtual computing environment, allowing you to use web service interfaces to launch instances with a variety of operating systems, load them with your custom application environment, manage your network's access permissions, and run your image using as many or few systems as you desire.

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.

Reserved Instances (RI) – 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.

Neither on-demand instances nor reserved instances can be used for existing server-bound software licenses.

References:

https://aws.amazon.com/ec2/dedicated-hosts/

https://aws.amazon.com/ec2/dedicated-hosts/faqs/

https://aws.amazon.com/ec2/pricing/dedicated-instances/

Domain

Design Cost-Optimized Architectures

Question 58Skipped

The DevOps team at a multi-national company is helping its subsidiaries standardize Amazon EC2 instances by using the same Amazon Machine Image (AMI). Some of these subsidiaries are in the same AWS region but use different AWS accounts whereas others are in different AWS regions but use the same AWS account as the parent company. The DevOps team has hired you as a solutions architect for this project.

Which of the following would you identify as CORRECT regarding the capabilities of an Amazon Machine Image (AMI)? (Select three)

Correct selection

You can share an Amazon Machine Image (AMI) with another AWS account

Copying an Amazon Machine Image (AMI) backed by an encrypted snapshot results in an unencrypted target snapshot

Correct selection

Copying an Amazon Machine Image (AMI) backed by an encrypted snapshot cannot result in an unencrypted target snapshot

Correct selection

You can copy an Amazon Machine Image (AMI) across AWS Regions

You cannot share an Amazon Machine Image (AMI) with another AWS account

You cannot copy an Amazon Machine Image (AMI) across AWS Regions

Overall explanation

Correct options:

You can copy an Amazon Machine Image (AMI) across AWS Regions

You can share an Amazon Machine Image (AMI) with another AWS account

Copying an Amazon Machine Image (AMI) backed by an encrypted snapshot cannot result in an unencrypted target snapshot

An Amazon Machine Image (AMI) provides the information required to launch an instance. An AMI includes the following:

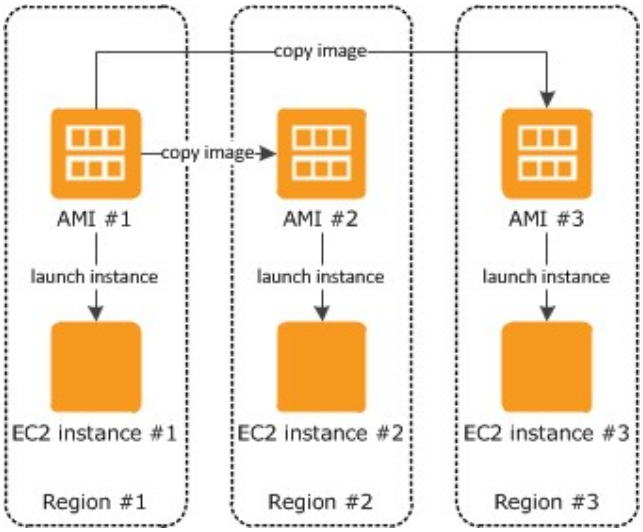
One or more Amazon EBS snapshots, or, for instance-store-backed AMIs, a template for the root volume of the instance.

Launch permissions that control which AWS accounts can use the AMI to launch instances.

A block device mapping that specifies the volumes to attach to the instance when it's launched.

You can copy an AMI within or across AWS Regions using the AWS Management Console, the AWS Command Line Interface or SDKs, or the Amazon EC2 API, all of which support the CopyImage action. You can copy both Amazon EBS-backed AMIs and instance-store-backed AMIs. You can copy AMIs with encrypted snapshots and also change encryption status during the copy process. Therefore, the option - "You can copy an AMI across AWS Regions" - is correct.

Copying AMIs across regions:



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

The following table shows encryption support for various AMI-copying scenarios. While it is possible to copy an unencrypted snapshot to yield an encrypted snapshot, you cannot copy an encrypted snapshot to yield an unencrypted one. Therefore, the option - "Copying an AMI backed by an encrypted snapshot cannot result in an unencrypted target snapshot" is correct.

Encryption and copying

The following table shows encryption support for various AMI-copying scenarios. While it is possible to copy an unencrypted snapshot to yield an encrypted snapshot, you cannot copy an encrypted snapshot to yield an unencrypted one.

Scenario	Description	Supported
1	Unencrypted-to-unencrypted	Yes
2	Encrypted-to-encrypted	Yes
3	Unencrypted-to-encrypted	Yes
4	Encrypted-to-unencrypted	No

Note

Encrypting during the `CopyImage` action applies only to Amazon EBS-backed AMIs. Because an instance store-backed AMI does not rely on snapshots, you cannot use copying to change its encryption status.

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

You can share an AMI with another AWS account. To copy an AMI that was shared with you from another account, the owner of the source AMI must grant you read permissions for the storage that backs the AMI, either the associated Amazon EBS snapshot (for an Amazon EBS-backed

AMI) or an associated S3 bucket (for an instance store-backed AMI). Therefore, the option - "You can share an AMI with another AWS account" - is correct.

Incorrect options:

You cannot copy an Amazon Machine Image (AMI) across AWS Regions

You cannot share an Amazon Machine Image (AMI) with another AWS account

Copying an Amazon Machine Image (AMI) backed by an encrypted snapshot results in an unencrypted target snapshot

These three options contradict the details provided in the explanation above.

Reference:

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

Domain

Design Secure Architectures

Question 59Skipped

A media streaming company is looking to migrate its on-premises infrastructure into the AWS Cloud. The engineering team is looking for a fully managed NoSQL persistent data store with in-memory caching to maintain low latency that is critical for real-time scenarios such as video streaming and interactive content. The team expects the number of concurrent users to touch up to a million so the database should be able to scale elastically.

As a solutions architect, which of the following AWS services would you recommend for this use-case?

Correct answer

Amazon DynamoDB

Amazon ElastiCache

Amazon RDS

Amazon DocumentDB

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 database with built-in security, backup and restore, and in-memory caching for internet-scale applications. DAX is a DynamoDB-compatible caching service that enables you to benefit from fast in-memory performance for demanding applications. Companies use caching through DynamoDB Accelerator (DAX) when they have high read volumes or need submillisecond read latency.

Incorrect options:

Amazon DocumentDB - Amazon DocumentDB is a fast, scalable, highly available, and fully managed document database service that supports MongoDB workloads. As a document database, Amazon DocumentDB makes it easy to store, query, and index JSON data. Although DocumentDB is fully managed, it does not have an in-memory caching layer.

Amazon ElastiCache - Amazon ElastiCache allows you to set up popular open-Source compatible in-memory data stores in the cloud. You can 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 such as Redis and Memcached. Elasticache is used as a caching layer. It's not a fully managed NoSQL database.

Amazon RDS - 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. It's not a NoSQL database.

References:

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

Domain

Design High-Performing Architectures

Question 60Skipped

An AWS Organization is using Service Control Policies (SCPs) for central control over the maximum available permissions for all accounts in their organization. This allows the organization to ensure that all accounts stay within the organization's access control guidelines.

Which of the given scenarios are correct regarding the permissions described below? (Select three)

Correct selection

Service control policy (SCP) affects all users and roles in the member accounts, including root user of the member accounts

If a user or role has an IAM permission policy that grants access to an action that is either not allowed or explicitly denied by the applicable service control policy (SCP), the user or role can still perform that action

Correct selection

Service control policy (SCP) does not affect service-linked role

Correct selection

If a user or role has an IAM permission policy that grants access to an action that is either not allowed or explicitly denied by the applicable service control policy (SCP), the user or role can't perform that action

Service control policy (SCP) affects service-linked roles

Service control policy (SCP) affects all users and roles in the member accounts, excluding root user of the member accounts

Overall explanation

Correct options:

If a user or role has an IAM permission policy that grants access to an action that is either not allowed or explicitly denied by the applicable service control policy (SCP), the user or role can't perform that action

Service control policy (SCP) affects all users and roles in the member accounts, including root user of the member accounts

Service control policy (SCP) does not affect service-linked role

Service control policy (SCP) are one type of policy that can be used to manage your organization. Service control policy (SCP) offers central control over the maximum available permissions for all accounts in your organization, allowing you to ensure your accounts stay within your organization's access control guidelines.

In service control policy (SCP), you can restrict which AWS services, resources, and individual API actions the users and roles in each member account can access. You can also define conditions for when to restrict access to AWS services, resources, and API actions. These restrictions even override the administrators of member accounts in the organization.

Please note the following effects on permissions vis-a-vis the service control policy (SCP):

If a user or role has an IAM permission policy that grants access to an action that is either not allowed or explicitly denied by the applicable service control policy (SCP), the user or role can't perform that action.

Service control policy (SCP) affects all users and roles in the member accounts, including root user of the member accounts.

Service control policy (SCP) does not affect any service-linked role.

Incorrect options:

If a user or role has an IAM permission policy that grants access to an action that is either not allowed or explicitly denied by the applicable service control policy (SCP), the user or role can still perform that action

Service control policy (SCP) affects all users and roles in the member accounts, excluding root user of the member accounts

Service control policy (SCP) affects service-linked roles

These three options contradict the details provided in the explanation above.

Reference:

https://docs.aws.amazon.com/organizations/latest/userguide/orgs_manage_policies_scp.html

Domain

Design Secure Architectures

Question 61Skipped

A financial services company is migrating their messaging queues from self-managed message-oriented middleware systems to Amazon Simple Queue Service (Amazon SQS). The development team at the company wants to minimize the costs of using Amazon SQS.

As a solutions architect, which of the following options would you recommend for the given use-case?

Use SQS visibility timeout to retrieve messages from your Amazon SQS queues

Use SQS message timer to retrieve messages from your Amazon SQS queues

Use SQS short polling to retrieve messages from your Amazon SQS queues

Correct answer

Use SQS long polling to retrieve messages from your Amazon SQS queues

Overall explanation

Correct option:

Use SQS long polling to retrieve messages from your Amazon SQS queues

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.

Amazon SQS provides short polling and long polling to receive messages from a queue. By default, queues use short polling. With short polling, Amazon SQS sends the response right away, even if the query found no messages. With long polling, Amazon SQS sends a response after it collects at least one available message, up to the maximum number of messages specified in the request. Amazon SQS sends an empty response only if the polling wait time expires.

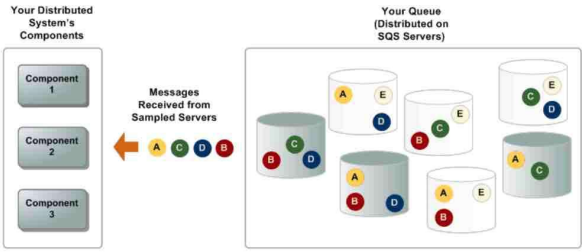
Long polling makes it inexpensive to retrieve messages from your Amazon SQS queue as soon as the messages are available. Using long polling can reduce the cost of using SQS because you can reduce the number of empty receives.

Short Polling vs Long Polling:

Consuming messages using short polling

When you consume messages from a queue using short polling, Amazon SQS samples a subset of its servers (based on a weighted random distribution) and returns messages from only those servers. Thus, a particular `ReceiveMessage` request might not return all of your messages. However, if you have fewer than 1,000 messages in your queue, a subsequent request will return your messages. If you keep consuming from your queues, Amazon SQS samples all of its servers, and you receive all of your messages.

The following diagram shows the short-polling behavior of messages returned from a standard queue after one of your system components makes a receive request. Amazon SQS samples several of its servers (in gray) and returns messages A, C, D, and B from these servers. Message E isn't returned for this request, but is returned for a subsequent request.



Consuming message using long polling

When the wait time for the `ReceiveMessage` API action is greater than 0, *long polling* is in effect. The maximum long polling wait time is 20 seconds. Long polling helps reduce the cost of using Amazon SQS by eliminating the number of empty responses (when there are no messages available for a `ReceiveMessage` request) and false empty responses (when messages are available but aren't included in a response). For information about enabling long polling for a new or existing queue using the AWS Management Console or the AWS SDK for Java (and the `CreateQueue`, `SetQueueAttributes`, and `ReceiveMessage` actions), see the Tutorial: Configuring long polling for an Amazon SQS queue tutorial. For best practices, see Setting up long polling.

via - https://aws.amazon.com/sqs/faqs/

Incorrect options:

Use SQS short polling to retrieve messages from your Amazon SQS queues - With short polling, Amazon SQS sends the response right away, even if the query found no messages. You end up paying more because of the increased number of empty receives.

Use SQS visibility timeout to retrieve messages from your Amazon SQS queues - Visibility timeout is a period during which Amazon SQS prevents other consumers from receiving and processing a given message. The default visibility timeout for a message is 30 seconds. The minimum is 0 seconds. The maximum is 12 hours. You cannot use visibility timeout to retrieve messages from your Amazon SQS queues. This option has been added as a distractor.

Use SQS message timer to retrieve messages from your Amazon SQS queues - You can use message timers to set an initial invisibility period for a message added to a queue. So, if you send a message with a 60-second timer, the message isn't visible to consumers for its first 60 seconds in the queue. The default (minimum) delay for a message is 0 seconds. The maximum is 15 minutes. You cannot use message timer to retrieve messages from your Amazon SQS queues. This option has been added as a distractor.

References:

https://docs.aws.amazon.com/AWSSimpleQueueService/latest/SQSDeveloperGuide/sqs-short-and-long-polling.html

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

Domain

Design Cost-Optimized Architectures

Question 62Skipped

A leading news aggregation company offers hundreds of digital products and services for customers ranging from law firms to banks to consumers. The company bills its clients based on per unit of clickstream data provided to the clients. As the company operates in a regulated industry, it needs to have the same ordered clickstream data available for auditing within a window of 7 days.

As a solutions architect, which of the following AWS services provides the ability to run the billing process and auditing process on the given clickstream data in the same order?

Amazon Kinesis Data Analytics

Amazon Kinesis Data Firehose

Correct answer

Amazon Kinesis Data Streams

Amazon Simple Queue Service (SQS)

Overall explanation

Correct option:

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 data collected is available in milliseconds to enable real-time analytics use cases such as real-time dashboards, real-time anomaly detection, dynamic pricing, and more.

Amazon Kinesis Data Streams enables real-time processing of streaming big data. It provides ordering of records, as well as the ability to read and/or replay records in the same order to multiple Amazon Kinesis Applications. The Amazon Kinesis Client Library (KCL) delivers all records for a given partition key to the same record processor, making it easier to build multiple applications reading from the same Amazon Kinesis data stream (for example, to perform counting, aggregation, and filtering). Amazon Kinesis Data Streams is recommended when you need the ability to consume records in the same order a few hours later.

For example, you have a billing application and an audit application that runs a few hours behind the billing application. Because Amazon Kinesis Data Streams stores data for a maximum of 365 days, you can easily run the audit application up to 7 days behind the billing application.

KDS provides the ability to consume records in the same order a few hours later

Q: When should I use Amazon Kinesis Data Streams, and when should I use Amazon SQS?

We recommend Amazon Kinesis Data Streams for use cases with requirements that are similar to the following:

- Routing related records to the same record processor (as in streaming MapReduce). For example, counting and aggregation are simpler when all records for a given key are routed to the same record processor.
- Ordering of records. For example, you want to transfer log data from the application host to the processing/archival host while maintaining the order of log statements.
- Ability for multiple applications to consume the same stream concurrently. For example, you have one application that updates a real-time dashboard and another that archives data to Amazon Redshift. You want both applications to consume data from the same stream concurrently and independently.
- Ability to consume records in the same order a few hours later. For example, you have a billing application and an audit application that runs a few hours behind the billing application. Because Amazon Kinesis Data Streams stores data for up to 7 days, you can run the audit application up to 7 days behind the billing application.

via - https://aws.amazon.com/kinesis/data-streams/faqs/

Incorrect options:

Amazon Kinesis Data Firehose - Amazon Kinesis Data Firehose is the easiest way to load streaming data into data stores and analytics tools. It can capture, transform, and load streaming data into Amazon S3, Amazon Redshift, Amazon Elasticsearch Service, and Splunk, enabling near real-time analytics with existing business intelligence tools and dashboards you're already using today. It is a fully managed service that automatically scales to match the throughput of your data and requires no ongoing administration. It can also batch, compress, and encrypt the data before loading it, minimizing the amount of storage used at the destination and increasing security. As Amazon Kinesis Data Firehose is used to load streaming data into data stores , therefore this option is incorrect.

Amazon Kinesis Data Analytics - Amazon Kinesis Data Analytics is the easiest way to analyze streaming data in real-time. You can quickly build SQL queries and sophisticated Java applications using built-in templates and operators for common processing functions to organize, transform, aggregate, and analyze data at any scale. Kinesis Data Analytics enables you to easily and quickly build queries and sophisticated streaming applications in three simple steps: setup your streaming data sources, write your queries or streaming applications and set up your destination for processed data. As Amazon Kinesis Data Analytics is used to build SQL queries and sophisticated Java applications, therefore this option is incorrect.

Amazon Simple Queue Service (SQS) - 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 (First-In-First-Out) queues are designed to guarantee that messages are processed exactly once, in the exact order that they are sent. For Amazon SQS, you cannot have the same message being consumed by multiple consumers in the same order a few hours later, therefore this option is incorrect.

References:

https://aws.amazon.com/kinesis/data-streams/faqs/

https://aws.amazon.com/kinesis/data-firehose/faqs/

https://aws.amazon.com/kinesis/data-analytics/faqs/

Domain

Design Resilient Architectures

Question 63Skipped

A leading bank has moved its IT infrastructure to AWS Cloud and they have been using Amazon EC2 Auto Scaling for their web servers. This has helped them deal with traffic spikes effectively. But, their MySQL relational database has now become a bottleneck and they urgently need a fully managed auto scaling solution for their relational database to address any unpredictable changes in the traffic.

Can you identify the AWS service that is best suited for this use-case?

Amazon Aurora

Correct answer

Amazon Aurora Serverless

Amazon DynamoDB

Amazon ElastiCache

Overall explanation

Correct options:

Amazon Aurora Serverless

Amazon Aurora Serverless is an on-demand, auto-scaling configuration for Amazon Aurora (MySQL-compatible and PostgreSQL-compatible editions), where the database will automatically start-up, shut down, and scale capacity up or down based on your application's needs. It enables you to run your database in the cloud without managing any database instances. It's a simple, cost-effective option for infrequent, intermittent, or unpredictable workloads. You pay on a per-second basis for the database capacity you use when the database is active and migrate between standard and serverless configurations with a few clicks in the Amazon RDS Management Console.

Incorrect options:

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. DynamoDB can handle more than 10 trillion requests per day and can support peaks of more than 20 million requests per second. But, it is a NoSQL database service and hence not a fit for the given use-case.

Amazon ElastiCache - Amazon ElastiCache allows you to set up popular open-Source compatible in-memory data stores in the cloud. You can 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 such as Redis and Memcached. Elasticache is used as a caching layer. It's not a fully managed MySQL database.

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 is up to five times faster than standard MySQL databases and three times faster than standard PostgreSQL databases. Amazon Aurora features a distributed, fault-tolerant, self-healing storage system that auto-scales up to 124 TB per database instance. But, its not a complete auto scaling solution and neither is it fully managed like Aurora serverless. Hence is not the right fit for the given use-case.

Reference:

https://aws.amazon.com/rds/aurora/serverless/

Domain

Design Resilient Architectures

Question 64Skipped

A company has hired you as an AWS Certified Solutions Architect – Associate to help with redesigning a real-time data processor. The company wants to build custom applications that process and analyze the streaming data for its specialized needs.

Which solution will you recommend to address this use-case?

Use Amazon Simple Queue Service (Amazon SQS) to process the data streams as well as decouple the producers and consumers for the real-time data processor

Use Amazon Kinesis Data Firehose to process the data streams as well as decouple the producers and consumers for the real-time data processor

Correct answer

Use Amazon Kinesis Data Streams to process the data streams as well as decouple the producers and consumers for the real-time data processor

Use Amazon Simple Notification Service (Amazon SNS) to process the data streams as well as decouple the producers and consumers for the real-time data processor

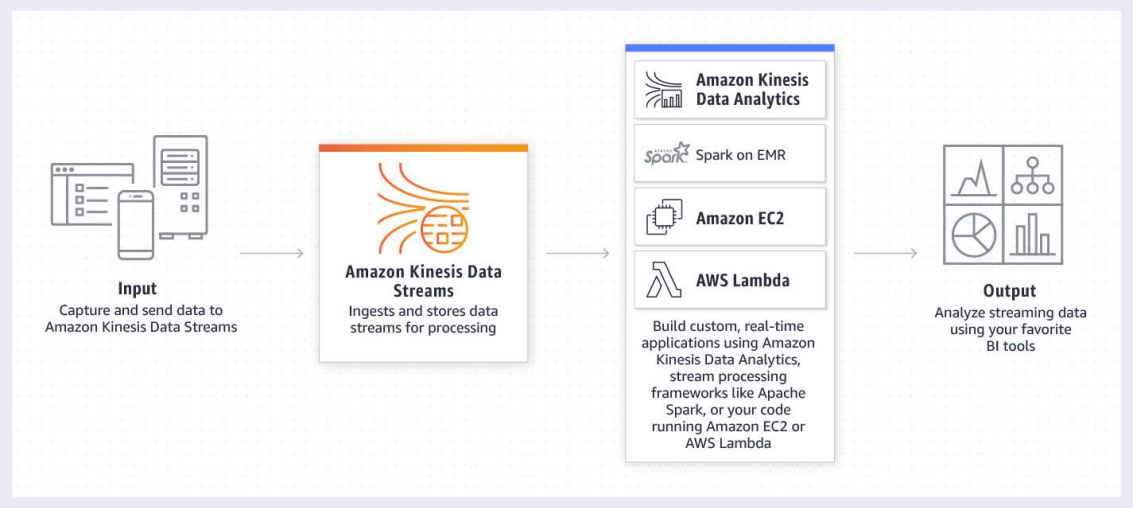
Overall explanation

Correct option:

Use Amazon Kinesis Data Streams to process the data streams as well as decouple the producers and consumers for the real-time data processor

Amazon Kinesis Data Streams is useful for rapidly moving data off data producers and then continuously processing the data, be it to transform the data before emitting to a data store, run real-time metrics and analytics, or derive more complex data streams for further processing. Kinesis data streams 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.

Kinesis Data Streams Overview:



via - https://aws.amazon.com/kinesis/data-streams/

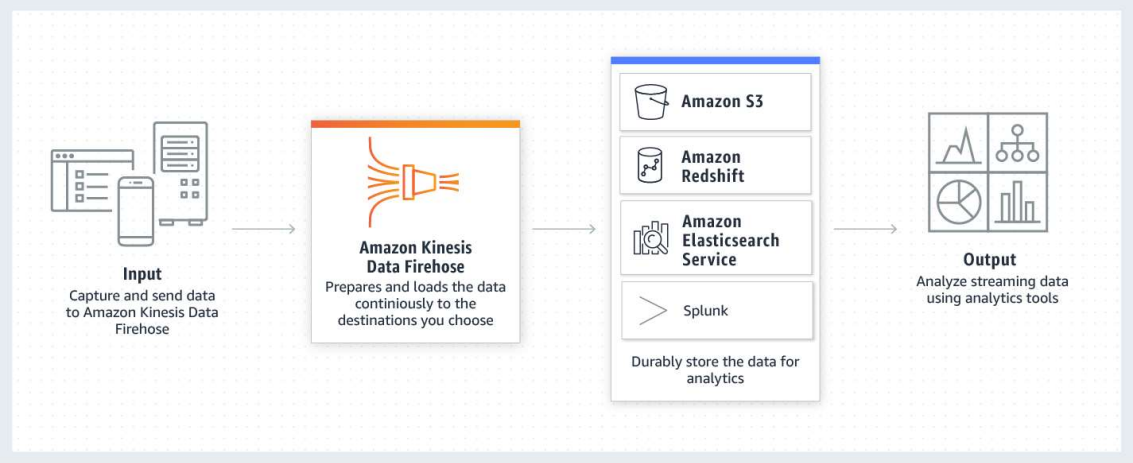
Incorrect options:

Use Amazon Simple Notification Service (Amazon SNS) to process the data streams as well as decouple the producers and consumers for the real-time data processor - 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. SNS cannot be used to decouple the producers and consumers for the real-time data processor as described in the given use-case.

Use Amazon Simple Queue Service (Amazon SQS) to process the data streams as well as decouple the producers and consumers for the real-time data processor - Amazon Simple Queue Service (Amazon SQS) offers a secure, durable, and available hosted queue that lets you integrate and decouple distributed software systems and components. SQS cannot be used to decouple the producers and consumers for the real-time data processor as described in the given use-case.

Use Amazon Kinesis Data Firehose to process the data streams as well as decouple the producers and consumers for the real-time data processor - Amazon Kinesis Data Firehose is the easiest way to load streaming data into data stores and analytics tools. Kinesis Firehose cannot be used to process and analyze the streaming data in custom applications. It can capture, transform, and load streaming data into Amazon S3, Amazon Redshift, Amazon Elasticsearch Service, and Splunk, enabling near real-time analytics.

Amazon Kinesis Data Firehose Overview



via - https://aws.amazon.com/kinesis/data-firehose/

References:

https://aws.amazon.com/kinesis/data-streams/

https://aws.amazon.com/kinesis/data-firehose/

Domain

Design Resilient Architectures

Question 65Skipped

The engineering team at a company wants to use Amazon Simple Queue Service (Amazon SQS) to decouple components of the underlying application architecture. However, the team is concerned about the VPC-bound components accessing Amazon Simple Queue Service (Amazon SQS) over the public internet.

As a solutions architect, which of the following solutions would you recommend to address this use-case?

Use Internet Gateway to access Amazon SQS

Use Network Address Translation (NAT) instance to access Amazon SQS

Use VPN connection to access Amazon SQS

Correct answer

Use VPC endpoint to access Amazon SQS

Overall explanation

Correct option:

Use VPC endpoint to access Amazon SQS

AWS customers can access Amazon Simple Queue Service (Amazon SQS) from their Amazon Virtual Private Cloud (Amazon VPC) using VPC endpoints, without using public IPs, and without needing to traverse the public internet. VPC endpoints for Amazon SQS are powered by AWS

PrivateLink, a highly available, scalable technology that enables you to privately connect your VPC to supported AWS services.

Amazon VPC endpoints are easy to configure. They also provide reliable connectivity to Amazon SQS without requiring an internet gateway, Network Address Translation (NAT) instance, VPN connection, or AWS Direct Connect connection. With VPC endpoints, the data between your Amazon VPC and Amazon SQS queue is transferred within the Amazon network, helping protect your instances from internet traffic.

AWS PrivateLink simplifies the security of data shared with cloud-based applications by eliminating the exposure of data to the public Internet. AWS PrivateLink provides private connectivity between VPCs, AWS services, and on-premises applications, securely on the Amazon network. AWS PrivateLink makes it easy to connect services across different accounts and VPCs to significantly simplify the network architecture.

Incorrect options:

Use Internet Gateway to access Amazon SQS - 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. This option is ruled out as the team does not want to use the public internet to access Amazon SQS.

Use VPN connection to access Amazon SQS - AWS Site-to-Site VPN (aka VPN Connection) 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. As the existing infrastructure is within AWS Cloud, therefore a VPN connection is not required.

Use Network Address Translation (NAT) instance to access Amazon SQS - You can use a network address translation (NAT) instance in a public subnet in your VPC to enable instances in the private subnet to initiate outbound IPv4 traffic to the Internet or other AWS services, but prevent the instances from receiving inbound traffic initiated by someone on the Internet. Amazon provides Amazon Linux AMIs that are configured to run as NAT instances. These AMIs include the string amzn-ami-vpc-nat in their names, so you can search for them in the Amazon EC2 console. This option is ruled out because NAT instances are used to provide internet access to any instances in a private subnet.

References:

https://aws.amazon.com/privatelink/

https://aws.amazon.com/about-aws/whats-new/2018/12/amazon-sqs-vpc-endpoints-aws-privatelink/

Domain

Design Secure Architectures