1. Question

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?

* **AWS Kinesis Data Firehose**
* Amazon SQS
* **AWS Kinesis Data Streams**
* AWS Kinesis Data Analytics

**Incorrect**

Correct option:  
AWS 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.

2. Question

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?

* AD Connector
* **AWS Managed Microsoft AD**
* Amazon Cloud Directory
* Simple AD

**Correct**

Correct option:  
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).

3. Question

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 Aurora Global Database**
* Provision Amazon RDS for PostgreSQL with a cross-Region read replica
* **Provision Amazon DynamoDB global tables**
* Provision Amazon RDS for MySQL with a cross-Region read replica

**Incorrect**

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.

4. Question

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
* You can use a range of IP addresses in CIDR block notation as the custom source for the inbound rule
* **You can use an Internet Gateway ID as the custom source for the inbound rule**
* You can use an IP address as the custom source for the inbound rule

**Correct**

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:

5. Question

A financial services company is migrating their messaging queues from self-managed message-oriented middleware systems to Amazon SQS. The development team at the company wants to minimize the costs of using SQS.  
As a solutions architect, which of the following options would you recommend for the given use-case?

* Use SQS short polling to retrieve messages from your Amazon SQS queues
* Use SQS message timer to retrieve messages from your Amazon SQS queues
* **Use SQS long polling to retrieve messages from your Amazon SQS queues**
* **Use SQS visibility timeout to retrieve messages from your Amazon SQS queues**

**Incorrect**

Correct option:  
Use SQS long polling to retrieve messages from your Amazon SQS queues  
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 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

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

* **Create an AWS WAF ACL and use an IP match condition to allow traffic only from those IPs that are allowed in the EC2 security group. Associate this new WAF ACL with the S3 bucket policy**
* Create a new NACL that allows traffic from the same IPs as specified in the current EC2 security group. Associate this new NACL with the CloudFront distribution
* **Configure an origin access identity (OAI) and associate it with the CloudFront distribution. Set up the permissions in the S3 bucket policy so that only the OAI can read the objects**
* **Create an AWS WAF ACL and use an IP match condition to allow traffic only from those IPs that are allowed in the EC2 security group. Associate this new WAF ACL with the CloudFront distribution**
* **Create a new security group that allows traffic from the same IPs as specified in the current EC2 security group. Associate this new security group with the CloudFront distribution**

**Incorrect**

Correct options:  
Configure an origin access identity (OAI) and associate it with the CloudFront distribution. Set up the permissions in the S3 bucket policy so that only the OAI can read the objects  
When you use CloudFront with an Amazon S3 bucket as the origin, you can configure 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 CloudFront distribution—that is, prevents them from accessing the content directly from the bucket, or through an unintended CloudFront distribution  
To do this, configure CloudFront to send authenticated requests to Amazon S3, and configure Amazon S3 to only allow access to authenticated requests from CloudFront. 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 EC2 security group. Associate this new WAF ACL with the 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 EC2 security group into the IP match condition.  
Incorrect options:  
Create an AWS WAF ACL and use an IP match condition to allow traffic only from those IPs that are allowed in the EC2 security group. Associate this new WAF ACL with the S3 bucket policy – You cannot associate a WAF ACL with an S3 bucket policy.  
Create a new NACL that allows traffic from the same IPs as specified in the current EC2 security group. Associate this new NACL with the CloudFront distribution – NACL is associated with a subnet within a VPC. CloudFront delivers your content through a worldwide network of data centers called edge locations. So a NACL cannot be associated with a CloudFront distribution.  
Create a new security group that allows traffic from the same IPs as specified in the current EC2 security group. Associate this new security group with the CloudFront distribution – A security group acts as a virtual firewall for your 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. CloudFront delivers your content through a worldwide network of data centers called edge locations. So a security group cannot be associated with a 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>

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 - 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
* 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

**Incorrect**

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. 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:

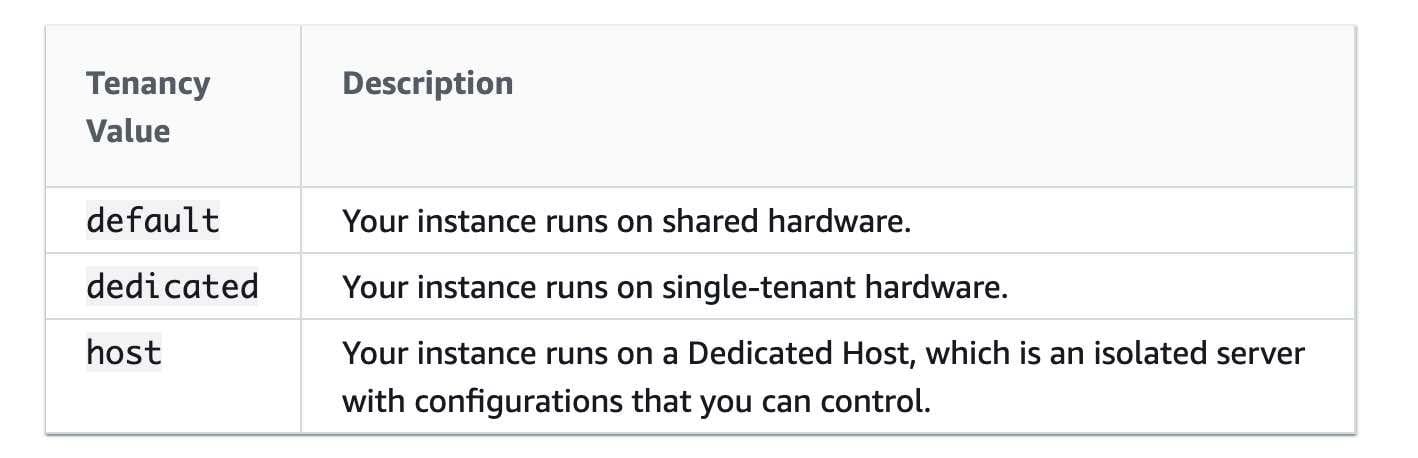
8. Question

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

**Incorrect**

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



9. Question

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 ElastiCache to meet the given requirements? (Select two)

* Use ElastiCache to improve latency and throughput for write-heavy application workloads
* **Use ElastiCache to improve the performance of Extract-Transform-Load (ETL) workloads**
* Use ElastiCache to run highly complex JOIN queries
* **Use ElastiCache to improve the performance of compute-intensive workloads**
* **Use ElastiCache to improve latency and throughput for read-heavy application workloads**

**Incorrect**

Correct option:  
Use ElastiCache to improve latency and throughput for read-heavy application workloads  
Use 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.

A screen shot of a computer

Description automatically generated

10. Question

A company has its application servers in the public subnet that connect to the RDS instances in the private subnet. For regular maintenance, the 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 an Egress-only internet gateway for the resources in the private 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 a NAT instance in the public subnet of the VPC
* **Configure a NAT Gateway in the public subnet of the VPC**

**Correct**

Correct option:  
Configure a NAT Gateway in the public subnet of the VPC – 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.

11. Question

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 EC2 instance and needs to support up to 25,000 IOPS per volume.  
As a solutions architect, which of the following EBS volume types would you recommend for this use-case?

* **Provisioned IOPS SSD (io1)**
* Cold HDD (sc1)
* General Purpose SSD (gp2)
* Throughput Optimized HDD (st1)

**Correct**

Correct option:  
Provisioned IOPS SSD (io1)  
Provisioned IOPS SSD (io1) is backed by solid-state drives (SSDs) and is a high-performance 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.

12. Question

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?

* **Schedule a weekly EventBridge event cron expression to invoke a Lambda function that runs the database rollover job**
* Provision an EC2 spot instance to run the database rollover script to be run via an OS-based weekly cron expression
* Create a time-based schedule option within an AWS Glue job to invoke itself every week and run the database rollover script
* **Provision an EC2 scheduled reserved instance to run the database rollover script to be run via an OS-based weekly cron expression**

**Incorrect**

Correct option:  
Schedule a weekly EventBridge event cron expression to invoke a 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:

##### A screenshot of a computer Description automatically generated 13. Question

A retail company has its flagship application running on a fleet of EC2 instances behind an Elastic Load Balancer (ELB). The engineering team has been seeing recurrent issues wherein the in-flight requests from the ELB to the EC2 instances are getting dropped when an instance becomes unhealthy.  
Which of the following features can be used to address this issue?

* Sticky Sessions
* **Cross Zone load balancing**
* Idle Timeout
* **Connection Draining**

**Incorrect**

Correct option:  
Connection Draining  
To ensure that an Elastic Load Balancer 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.

14. Question

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 (NLB) 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?

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

**Incorrect**

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.

15. Question

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 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 an inbound endpoint on Route 53 Resolver and then Route 53 Resolver can conditionally forward queries to resolvers on the on-premises network via this endpoint**
* Create a universal endpoint on Route 53 Resolver and then Route 53 Resolver can receive and forward queries to resolvers on the on-premises network via this endpoint
* **Create an inbound endpoint on Route 53 Resolver and then DNS resolvers on the on-premises network can forward DNS queries to Route 53 Resolver via this endpoint**
* **Create an outbound endpoint on Route 53 Resolver and then Route 53 Resolver can conditionally forward queries to resolvers on the on-premises network via this endpoint**
* Create an outbound endpoint on Route 53 Resolver and then DNS resolvers on the on-premises network can forward DNS queries to Route 53 Resolver via this endpoint

**Incorrect**

Correct options:  
Create an inbound endpoint on Route 53 Resolver and then DNS resolvers on the on-premises network can forward DNS queries to Route 53 Resolver via this endpoint  
Create an outbound endpoint on Route 53 Resolver and then 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, Route 53 Resolver automatically answers DNS queries for local VPC domain names for 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 Route 53 Resolver and then DNS resolvers on the on-premises network can forward DNS queries to Route 53 Resolver via this endpoint.  
Resolver Inbound Endpoint

A screenshot of a computer

Description automatically generated

16. Question

An e-commerce company is using an Elastic Load Balancer for its fleet of EC2 instances spread across two Availability Zones, 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 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
* **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 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**

**Incorrect**

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.

A diagram of a computer network

Description automatically generated  
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

17. Question

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 on the AWS side of the VPN and a Customer Gateway on the on-premises side of the VPN**
* **Create a Virtual Private Gateway on both the AWS side of the VPN as well as the on-premises side of the VPN**
* Create a Virtual Private Gateway on the on-premises side of the VPN and a Customer Gateway on the AWS side of the VPN

**Incorrect**

Correct option:  
Create a Virtual Private Gateway 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 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 (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 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 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 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 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 on the AWS side of the VPN and a Customer Gateway on the on-premises side of the VPN. Therefore, this option is wrong.

18. Question

A data analytics company manages an application that stores user data in a DynamoDB table. The development team has observed that once in a while, the application writes corrupted data in the 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 DynamoDB on-demand backup to restore the table to the state just before corrupted data was written
* Use DynamoDB Streams to restore the table to the state just before corrupted data was written
* **Configure the DynamoDB table as a global table and point the application to use the table from another AWS region that has no corrupted data**
* **Use DynamoDB point in time recovery to restore the table to the state just before corrupted data was written**

**Incorrect**

Correct option:  
Use 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.

19. Question

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 Kinesis Data Streams to ingest the data, process it using AWS Lambda or run analytics using Kinesis Data Analytics**
* Use Amazon API Gateway with the existing REST-based interface to create a high performing architecture
* Use Amazon SQS for data ingestion and configure Lambda to trigger logic for downstream processing
* Use Amazon SNS for data ingestion and configure Lambda to trigger logic for downstream processing

**Correct**

Correct option:  
Use Amazon Kinesis Data Streams to ingest the data, process it using AWS Lambda or run analytics using 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. 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.

20. Question

A gaming company uses Application Load Balancers (ALBs) in front of Amazon EC2 instances for different services and microservices. The architecture has now become complex with too many ALBs 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?

* Configure Elastic IPs for each of the ALBs in each Region
* **Launch AWS Global Accelerator and create endpoints for all the Regions. Register the ALBs of each Region to the corresponding endpoints**
* Assign an Elastic IP to an Auto Scaling Group (ASG), and set up multiple Amazon EC2 instances to run behind the ASGs, for each of the Regions
* **Set up a Network Load Balancer (NLB) with Elastic IPs. Register the private IPs of all the ALBs as targets of this NLB**

**Incorrect**

Correct option:  
Launch AWS Global Accelerator and create endpoints for all the Regions. Register the ALBs 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 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 EC2 Instances, without making user-facing changes. To mitigate endpoint failure, Global Accelerator automatically re-routes your traffic to your nearest healthy available endpoint.  
Simplified and resilient traffic routing for multi-Region applications:

21. Question

A startup has recently moved their monolithic web application to AWS Cloud. The application runs on a single 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 Amazon EventBridge events that can trigger the recovery of the EC2 instance, in case the instance or the application fails**
* Configure AWS Trusted Advisor to monitor the health check of EC2 instance and provide a remedial action in case an unhealthy flag is detected
* Configure an Amazon CloudWatch alarm that triggers the recovery of the EC2 instance, in case the instance fails. The instance can be configured with EBS volume or with instance store volumes
* **Configure an Amazon CloudWatch alarm that triggers the recovery of the EC2 instance, in case the instance fails. The instance, however, should only be configured with an EBS volume**

**Incorrect**

Correct option:  
Configure an Amazon CloudWatch alarm that triggers the recovery of the EC2 instance, in case the instance fails. The instance, however, should only be configured with an EBS volume – If your instance fails a system status check, you can use CloudWatch alarm actions to automatically recover it. The recover option is available for over 90% of deployed customer EC2 instances. The 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

22. Question

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 EC2 instance (E1) in the subnet S1 and assigns a public IPv4 address to this instance. Next the team also launches a 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 EC2 instance E1?

* Route Table (R1)
* **NAT instance (N1)**
* **Internet Gateway (I1)**
* Subnet (S1)

**Incorrect**

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:

23. Question

Your application is hosted by a provider on yourapp.provider.com. You would like to have your users access your application using [http://www.your-domain.com](http://www.your-domain.com/), which you own and manage under Route 53.  
What Route 53 record should you create?

* Create a PTR record
* Create an Alias Record
* **Create an A record**
* **Create a CNAME record**

**Incorrect**

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 [http://www.example.com](http://www.example.com/), newproduct.example.com, and so on.  
Please review the major differences between CNAME and Alias Records:

A screenshot of a computer

Description automatically generated

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 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:

24. Question

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 EC2 instances running these HPC workflows?

* Elastic IP Address
* **Elastic Fabric Adapter**
* Elastic Network Adapter
* **Elastic Network Interface**

**Incorrect**

Correct option:  
Elastic Fabric Adapter  
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

25. Question

An AWS Organization is using Service Control Policies (SCP) 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)

* **SCPs affect service-linked roles**
* **SCPs affect all users and roles in attached accounts, including the root user**
* SCPs affect all users and roles in attached accounts, excluding the root user
* **SCPs do not affect service-linked role**
* 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 SCPs, the user or role can still perform that action
* **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 SCPs, the user or role can‘t perform that action**

**Incorrect**

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 SCPs, the user or role can‘t perform that action  
SCPs affect all users and roles in attached accounts, including the root user

An e-commerce company runs its web application on EC2 instances in an Auto Scaling group and it‘s configured to handle consumer orders in an 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?

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

**Incorrect**

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.