AWS Solutions Architect Associate (SAA-C02)

Exam Guide

A Guaranteed Approach to Pass the AWS SAA-C02 Exam in 2020-21

Written by
Allison Cope

ISBN: 9798621775773

Table of contents

INTRODUCTION

One of the most in-demand qualifications for cloud engineers is the AWS Certified Solutions Architect Associate certification. Getting the AWS Solutions Architect certification and having enough knowledge to take up the job can surely place a hefty six-figure paycheck in your hands every month. Thus, we are super excited to bring you this book to grow and level up your career with this masterpiece.

This is the most advanced last-minute AWS Solutions Architect Associate's exam prep guide based on the C02 version. This book is designed for individuals preparing for the AWS Certified Solution Architect examination and who already have completed AWS SAA training and are ready to take the exam. The purpose of this preparation book is to recall every important concept in the AWS SAA course so that individuals have a better chance of passing the exam. Beginners who are unaware of the AWS architectural principles and services can also benefit from this book as a starting point.

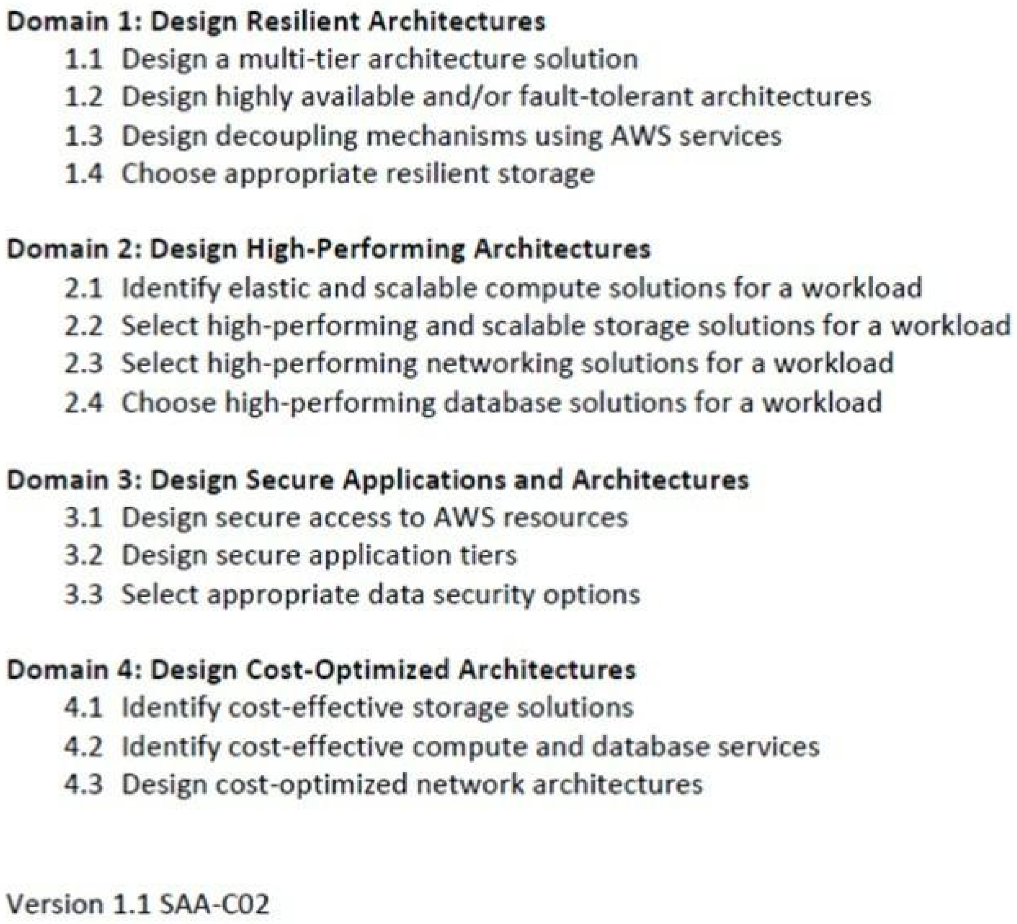
AWS writes their questions in such a way that only someone with real hands-on experience and an understanding of their services can pass. Through this book, you will be able to learn the key concepts asked in the examination and how they map the objectives of the examination. Some companies publish question banks on the AWS exams, and the content of these question banks is what we will see and learn here. To better prepare you for the exam itself, the chapters and content in this book are created to specifically target the AWS SAA-C02 exam questions, which will help you remember all the key concepts for the exam. These chapters include use-case scenarios and architectural diagrams to memorize all the key concepts for the exam quickly.

Some of the handy tips to pass the AWS Solutions Architect Associate exam are shared at the end of the book, taken from the experts and the people who passed the exam multiple times. The tips can help you get ideas on how to prepare and pass the exam. AWS Cheat Sheets are also presented in bullet points with easy to remember notes to give you a bird’s eye view of the important Amazon web services.

There is a free goodie of 60+ Practice Questions for our readers to identify their strengths and weaknesses in the AWS concepts. Practice Questions are the exam blueprint which test individual knowledge and skills before taking the final exam. Thus, it will add variety to your learning process and will give you more confidence before going into the actual test.

Before get started, first let me walk you through the AWS SAA-C02 exam domains. The exam itself is split into four domains or four sections. The examination is based on these four domains. Domain one is to design Resilient Architectures. Domain 2 is to design High-performance Architecture. In domain 3, we will look at how Secure Applications and Architectures can be built. Domain 4, the last section, is the development of Cost-Optimized Architectures.

Therefore, each of these domains is split into several percentages i.e. the questions will make up this amount of percentage. For example Domain one will have 30 percent of the examination, domain 2 will cover 28 percent, domain 3 will have 24 percent, and domain 4 will have 18 percent of the examination. You can see the domains and respective objectives in the following images. This is all about four domains. Let’s get started with our prep book.



Domain 1: Design Resilient Architectures

1.1 Design a multi-tier architecture solution
1.2 Design highly available and/or fault-tolerant architectures
1.3 Design decoupling mechanisms using AWS services
1.4 Choose appropriate resilient storage

Domain 2: Design High-Performing Architectures

2.1 Identify elastic and scalable compute solutions for a workload
2.2 Select high-performing and scalable storage solutions for a workload
2.3 Select high-performing networking solutions for a workload
2.4 Choose high-performing database solutions for a workload

Domain 3: Design Secure Applications and Architectures

3.1 Design secure access to AWS resources
3.2 Design secure application tiers
3.3 Select appropriate data security options

Domain 4: Design Cost-Optimized Architectures

4.1 Identify cost-effective storage solutions
4.2 Identify cost-effective compute and database services
4.3 Design cost-optimized network architectures

Version 1.1 SAA-C02

Amazon Web Services

Cloud technology has come out as one of the most important topics in the IT world. Both major and minor businesses are switching from old data centres to the new cloud infrastructure. These infrastructure models are frequently evolving, and the cloud providers are competing to provide new features and services to users to meet their challenges more effectively. Amazon web services remain a strong industry leader in the cloud computing sector with all the newly emerging models and service providers.

AWS offers more than 100 web services that enable organizations to launch virtual machines, storage systems, databases, security, and other resources. These resources then help to build complete application environments for companies to run applications, minimize cost, and scale up the operations. All AWS functionality is based on a pay-as- you-go model, where the user only pays for the type and number of services used. It entirely empowers its users to build and launch services, some of which run web and application servers in the cloud to host websites and internal applications.

Users are allowed to store their files and sensitive data securely on the cloud, which can be accessed later for backup and disaster recovery purposes. They can build managed databases to store data and execute reports on common DBMS such as MySQL, Oracle, PostgreSQL, or SQL Server. They can perform complete data analysis with services such as AWS Kinesis, AWS QuickSight, and AWS Glue. Furthermore, they can establish reliable content distribution networks using AWS CloudFront. With these networks, the users can send static files such as images and videos to different edge locations worldwide.

Why This Certification?

Certifications play a significant role in improving and measuring knowledge. AWS certification offers a guided approach of learning for beginners by keeping them on track and motivated. On the other hand, an expert might also want to level up and check his knowledge with certifications. Cloud Computing has never experienced such a great exposure before. Thus, it the best time to get certified in cloud technologies. AWS is one of the best platforms to start for beginners. It is the world's leading participant in cloud computing. Some other reasons to learn AWS certification are:

Cloud adoption statistics prove that cloud technology is the future.

The primary concern of a cloud professional is to stay up-to-date with the new IT trends. Business owners and experts from the industry have realized that cloud computing is the future of technology. Therefore, a person becomes important to an organization if he discovers the new technology trends earlier and adopts them. According to an assessment, nearly 85% of enterprise workload will be in the cloud by 2022. Thus, it is time to get AWS certification instead of watching the technology to pass by.

Speciality skills make us more desirable to business owners.

The only way to reach the desired level of success is to increase our skill set. The greater the knowledge about cloud computing and AWS's ins and outs, the easier it is to find the desired job. Keeping AWS certification in parallel, we must search for new skill sets to secure the career. The long journey to AWS certification is worth spending time and money.

AWS is the leader in the Infrastructure-as-a-service industry

The Infrastructure as a Service (IaaS) has grown by over 30 per cent in recent years. AWS is considered as one of the leading companies in the IaaS revolution. Therefore, an individual must get AWS certification to help different companies establish a secure and functional infrastructure. This certification certifies a person to learn the basic concepts of the AWS platform.

This knowledge facilitates the companies into the adoption of cloud. AWS certified professionals usually handle the far fewer mistakes in this process of adoption, which is why most of the businesses favour them.

Getting certified distinguishes resume from other qualified candidates.

The cloud industry is evolving every year. Therefore, to keep up with these industry changes and stay relevant in the jobs, most of the people will need to learn AWS technologies. Thus, getting AWS certified keeps an individual in the lead from most candidates in such a rapidly growing market.

An AWS Certification commands high paying jobs.

AWS certification is one of the highly paid certifications among the existing IT certifications. According to a ranking of popular IT certifications, AWS Certified-Developer is ranked fourth amongst the most common certifications in IT. It guarantees an average pay of $114,148 to its certified developers annually. AWS Solutions Architect Associate Certification secures an even higher ranking (2nd rank) as it ensures an average salary of $121,292 annually for its certified architects. Thus, these values make the AWS a clear choice for cloud certification.

An AWS certification demonstrates your credibility and expertise to customers and potential clients.

Trust and credibility play an important role in establishing and maintaining relationships while interacting with customers and providing AWS services to prospective clients. Mostly the customer expects an expert in the technology they are looking for. Therefore, certification is one of the best ways to demonstrate capability. Certification proves that the person has gone through tough training required to work with AWS technology. It can give you potential clients and peace of mind to current customers working with your organization.

AWS certifications open organizations up to partner program benefits

Companies qualify for different levels of the AWS partner program if they have AWS certified employees. Thus, the qualified organizations become AWS partners through the AWS partner program, enabling them to have access to various resources and provide training to support their customers working with AWS services in a better way. With each level of the partnership program, companies unlock greater benefits. The more certified resources a company has, the greater the number of benefits it can get.

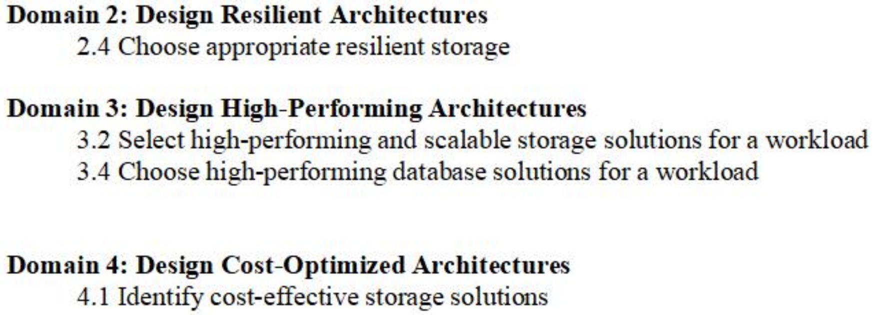
An AWS certification allows you to become a part of the AWS community.

It opens more networking opportunities for certified candidates. Amazon will provide official AWS certified logo usage and digital badges to display the credentials to the world. They have access to the AWS certified community, are invited to regional events, and free practise exams for other certifications.

AWS has emerged as one of the most rapidly growing technology in the IT industry. Different Companies are moving forward with AWS at remarkable rates. Getting AWS Certification means having the fundamental knowledge and learning about the required tools to work with technology's new aspect. Therefore, concerned individuals should seek certification because they can learn a lot in preparation for certification. Also, better jobs and earnings will be byproducts.

Storage Services

Let's begin with the first chapter of the book. In this chapter, we will see the Storage options available in AWS and link them to the exam objectives. So we have Domain 2 that is Designing Resilient Architectures when you look into the main objectives which relate to storage options. So here you have one objective to choose secure, robust storage. You also have Domain 3 where you have to select high- performing storage as well as database solutions for your underlying infrastructure.



Domain 2: Design Resilient Architectures

2.4 Choose appropriate resilient storage

Domain 3: Design High-Performing Architectures

3.2 Select high-performing and scalable storage solutions for a workload
3.4 Choose high-performing database solutions for a workload

Domain 4: Design Cost-Optimized Architectures

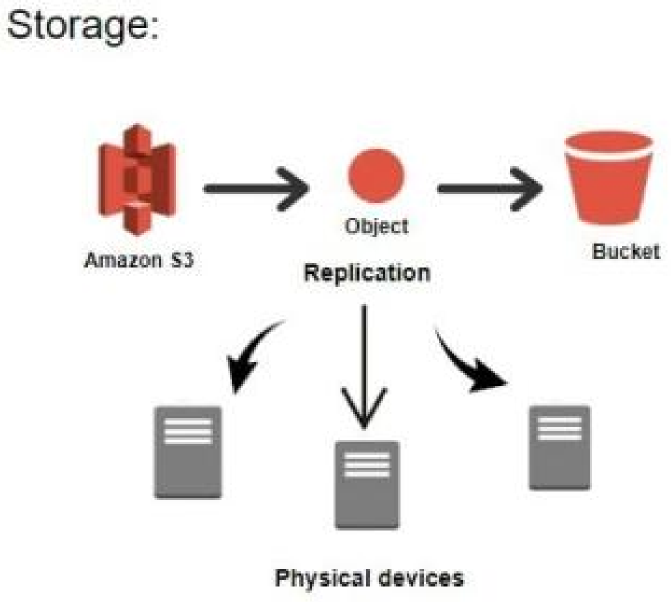
4.1 Identify cost-effective storage solutions

There is also an aspect of designing cost- optimized architecture that will reduce the costs of infrastructure, and here you'll choose to design them. You can, therefore, see that you have these storage options that are divided among these three domains. So, you should be prepared for questions that could ask you ideas behind these objectives and you can see that a significant portion of questions is dedicated to Domain 2 and Domain 3.

Let's continue to understand the various storage options. There are two types of storage in Amazon cloud infrastructure: Object-level storage and Block level storage. So, you have a Simple Storage Service called S3 and Amazon Glacier in object-level storage. On the other hand, the Elastic Block Storage volumes that are attachable to your EC2 instances are present in block-level storage. And then you have your databases such as Amazon Relational Database Service known as RDS, DynamoDB, Aurora, and Redshift. Now as you answer the questions, you should constantly observe the exam question by reading it again and again just to get some key terms in the question. This is one of the best ways to understand the question objective.

Sometimes you can see in questions: Design/build architecture with object-level storage in it. This key term for object-level storage should, therefore, steer your mind immediately to use either the Amazon S3 or Glacier. So sometimes, in the exam, search for these key terms in the question. It enables you to understand better and restricts the answer choices. In the end, it will lead you to choose the correct answer.

Now let's go to the object storage level which is S3. When mapping it to the first domain objective: Resilient storage, note that the Simple Storage Service S3 is highly durable and available. How is this done by the simple storage service? So when using S3 to upload an object to a bucket, the S3 copies the same object to several physical devices in the background. Only when multiple replicas are made on all physical devices then you get a response that the object is successfully uploaded in S3. If one of the physical entities crashes, the others are still available and thus, for your underlying objects, the S3 provides high availability and durability. So in the examination, if you're asked for object-level storage that has high durability and availability consider choosing the simple storage service (S3).



We have already discussed its advantages: if one physical device fails, the object remains present on another one. One important point is that it is only possible to replicate the objects within a specific region. Let's say in the exam if you've been given a scenario of disaster recovery in which you have a very critical data in a particular bucket which must be available all the time even if the region goes down, then in such case, S3 provides a feature called cross-region replication. By using cross-region replication, if you copy an object to a bucket in a given region, it will automatically be replicated to another bucket in a different region. It is done for you automatically.

The key concept here is to enable the bucket versioning because it is a requirement to make sure that objects are replicated into another bucket via cross-region replication. Remember, you can't replicate these objects if you have already existing ones in the bucket. So if you want to get your objects replication from the start, make sure you first enable cross-region replication and then continue adding objects to your primary bucket.

Other than Amazon S3, we have Amazon Glacier which is also used for archiving object storage. If you want to archive your data and your files, this is a very cost-effective option. Now AWS has this default process of data retrieval from the cold storage i.e., Glacier. You've uploaded your data onto Glacier, now you need to retrieve the data from it, you can't just download it directly. Instead, you have to submit something known as a job, and then after three to five hours, the objects you wish to retrieve would be available. So that's the trick when you have cold storage or archive storage.

It is relatively inexpensive, but one important thing to note is that it takes three to five hours to retrieve data. From an exam point of view, this is very significant. If you are asked a question that you can download your data/files after five hours or three hours and you are searching for a cost-effective option, you should choose Amazon Glacier. And there are two other ways in which you can download your objects even faster. The first thing in Amazon Glacier is called expedited retrieval. So in a couple of minutes, you can get your objects back by paying extra money. The second way is to use Amazon S3 Infrequent Access.

So in the exam, if one of the choices in the question states that you can use Amazon S3 Infrequent Access, then the expedited retrieval by Amazon Glacier is not the right option. You can then choose the Infrequent Access, which gives you cost efficiency and you can also access objects even quicker. Amazon Glacier is certainly much cheaper than Infrequent Access, but once again pick them according to the options available in the question. And then note that to send data to Amazon Glacier, you must either use the API Application Programming Interface or use the lifecycle policies specified in S3. Remember that you can't add objects directly to Amazon Glacier with the AWS console but either with API or life cycle policies.

Let's now go to EBS volumes. EBS volumes are not as durable as compared to S3. It's incredibly important to understand this concept. When you create an EBS volume in an Availability Zone, it is duplicated to different devices in that availability zone, but in case if that AZ fails for whatever reason, the volume will not be available anymore. Sometimes you can get a question like this: how you can secure your data on an EBS volume? As this is your responsibility not the job of AWS, you can take snapshots of EBS volumes. You can also make copies of these snapshots for disaster recovery and put them in another region. Remember that in another region, you can't create a snapshot directly. First, the snapshot must be created and then the option of copy snapshot must be used to copy the snapshots to another region. These steps are crucial to know from an examination point of view.

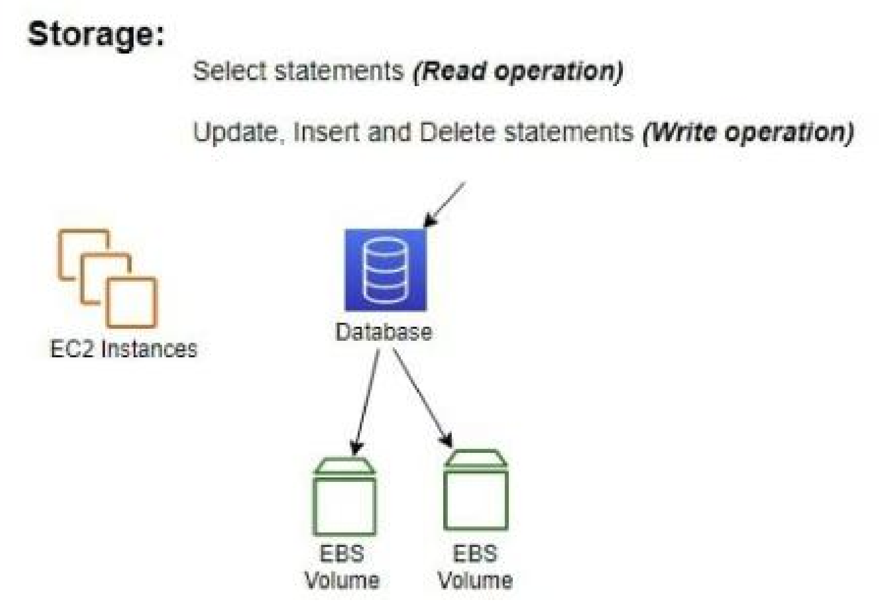
Now you'll have some questions regarding objectives of performance storage and cost-optimized storage when it comes to EBS volumes. Let's get a more detailed understanding of this. There are four types of EBS volumes so you've got the general-purpose SSD, you've got the provision IOPS, the throughput optimized HDD and lastly the cold HDD. Let's see when you'd use the various volume types.

Let's suppose you have an EC2 instance, and it hosts a web server with a predictable workload. In that case, the question can come as to what type of volume for the underlying EBS volumes will be most cost-effective? You should select the option of general-purpose SSD because this is ideal for critical workloads like web servers where for your EBS volumes, you don't need a lot of input/output operations. You can't select provision IOPS for such cases because that isn't cost-optimized. Provision IOPS is much expensive than general-purpose SSD. So choose the right option if you are being asked in the question about both performance storage and optimizing costs.

Let's now discuss the other EBS volume types. Let us assume you have an EC2 instance that has a database server, and it is resource-intensive. You have to find these types of keywords in the question. In this question, the workload on an EC2 instance is resource-intensive. There are many input and output operations, find all keywords in a question.

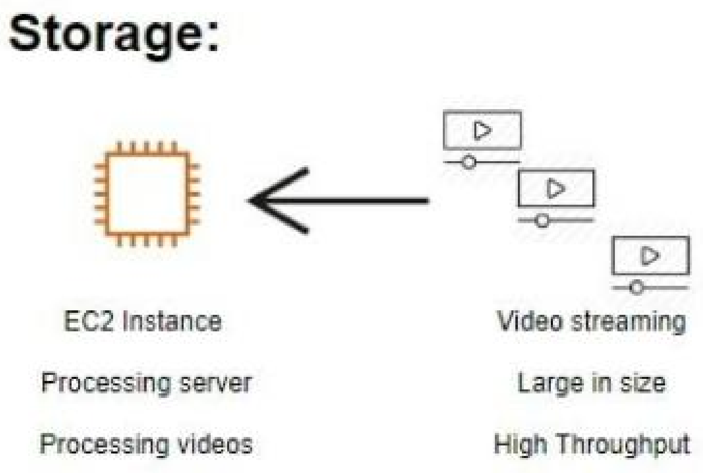
Before we can decide which EBS volume type is suitable for this scenario, I would like to describe what the input and output operations mean here. How does the database server differ from the webserver? A web server will usually include the web pages: you ask for a web page and it gathers the information, processes it and sends it to you. But what do you generally do in the databases? So usually in a very simple case of a database, you would fire a select statement which is similar to a read operation. You can also execute a write operation by updating, inserting and deleting statements.

Now, what's the database engine here? Now, let's assume its Oracle database, and what will be the internal process if you perform these read and write operations? Well, it will always go to its underlying volumes, which are stored on the EC2 instance then it will look for data on the volume when you fire the select statement. And it will write the data on the underlying volume on EC2 instance when you use an insert, delete or update statement. So, the database, therefore, needs to access the underlying


Storage:
Select statements (Read operation)
Update, Insert and Delete statements (Write operation)
EC2 Instances
Database
EBS Volume
EBS Volume

The read and write operations of data are therefore the responsibility of the input/output controller. If you do have multiple read & write operations, this controller may take a lot of time because it needs to search the disk from which data can be read, and to which data can be written. And if you have a large number of select, update, insert, and delete statements, then the controller begins to slow down and you get a drop in the system performance. In such a scenario, you should use Provision IOPS volumes because they are designed in such a way that the input and output operations are easier to perform. It is specifically designed for this purpose and it gives the underlying volumes better performance. As compared to the general-purpose SSD, you certainly need to pay extra, but in the examination, you would likely get a question based on performance, so in this case, you have to select Provision IOPS as the underlying volume type.

Consider the next question example. So we have an EC2 instance and we have some kind of video processing application on the server. There are several videos for processing. Some videos can be taken from an external source and uploaded to your EC2 instance. They may be large and the data is streaming from another application or another device. You need high throughput in such a case. Thus high throughput means turnout or the amount of data that comes in and goes out of your instance is quite large. It is neither the processing power nor the input and output operations perform on the disk. This means you aren't writing to the disk frequently but this time, it's the amount of data you write and read from the disk.



There are large portions of data here that's why it is referred to as high throughput. So the volume type, in this case, is Throughput optimized HDD. So, you would get a question like to select an EBS volume type with a high throughput of 400 megabytes per second. This is the amount of data that can be transmitted per second to the instance. This is ideal for large streaming of data like images, audio, and videos, etc. This is more cost-efficient than Provision IOPS.

Finally, we got the cold HDD for the infrequent access to storage. This is also designed for throughputs but mostly for infrequent access. Let us, therefore, consider a scenario in which you have a customer who uploads videos the way we did before. The videos are kept on the EBS volumes and frequently viewed over two months. But after a while, they begin to get access less often. What are you going to do in such a case? Well, you can have two separate EBS volumes if you need to store these videos on the EC2 instance. Thus, the videos can be stored on the EBS volume for two months where the EBS type is throughput optimized HDD. Two months later, you will be able to have a script that passes the video to another volume type which is cold HDD.

If performance storage is concerned overall, then it’s the perfect solution you can also optimize your cost at the same time. Because even after two months if you restore the videos on throughput optimized HDD, you will have to pay for extra storage when the data is not retrieved very often.

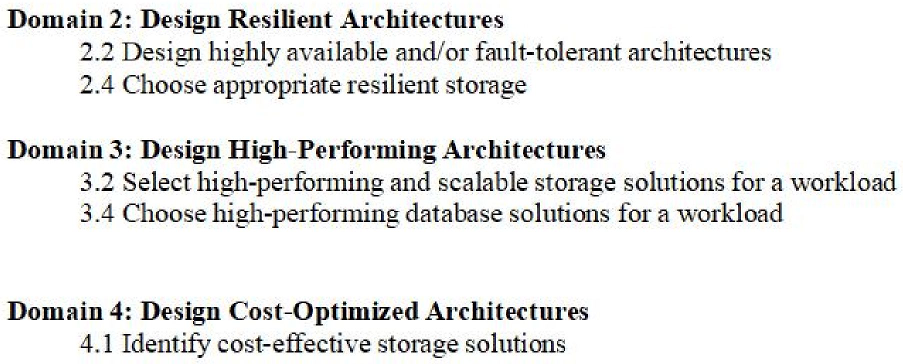
Let's see Amazon FSx and its types. Amazon FSx is a file system that gives high-performance data processing. So when I say the file system, we are using the file server to store our home directory or shared folder. AWS provides FSx service to use it as a file server to replace conventional storage. It uses SSD to give you high performance and low latency. So two types of storage are supported by AWS, one is Amazon FSx for Windows File server and the other is Amazon FSx for Lustre as a file server.

Amazon FSx for Lustre is mainly used for data-hungry applications such as machine learning, electronic design automation, financial modeling, and video processing. So you can use the Amazon FSx in so many cases. You can use it to share files and for home directories. You can use for lift-and-shift application workloads, and media and entertainment workflows. With data analytics as well: Luster can be used for business intelligence, data visualization app, high availability and deployments of the Microsoft SQL Server. This can also be used for content management and web services.

So, in this chapter, we've discussed the types of questions that will be asked in the examination about simple storage service and EBS volumes and how we can map its objectives.

Database Services

Let us now move to the next chapter of Database, and look at its options available in AWS in continuation of the objectives mapped in the exam. We will focus on the same objectives again, but we will add another one this time. So we have added objective 2.2 to see how high available architectures can be designed.



Domain 2: Design Resilient Architectures

2.2 Design highly available and/or fault-tolerant architectures

2.4 Choose appropriate resilient storage

Domain 3: Design High-Performing Architectures

3.2 Select high-performing and scalable storage solutions for a workload

3.4 Choose high-performing database solutions for a workload

Domain 4: Design Cost-Optimized Architectures

4.1 Identify cost-effective storage solutions

When it comes to databases in the examination, high availability is very essential. Therefore, you are going to see a couple of questions regarding high database availability, so let's take a look at the Amazon Relational Database Service. Database services such as Oracle, Microsoft SQL Server, MySQL, PostgreSQL, and MariaDB are offered. We also have Amazon Aurora; we have Amazon DynamoDB and Amazon Redshift. Let's talk about each of them in a bit of detail.

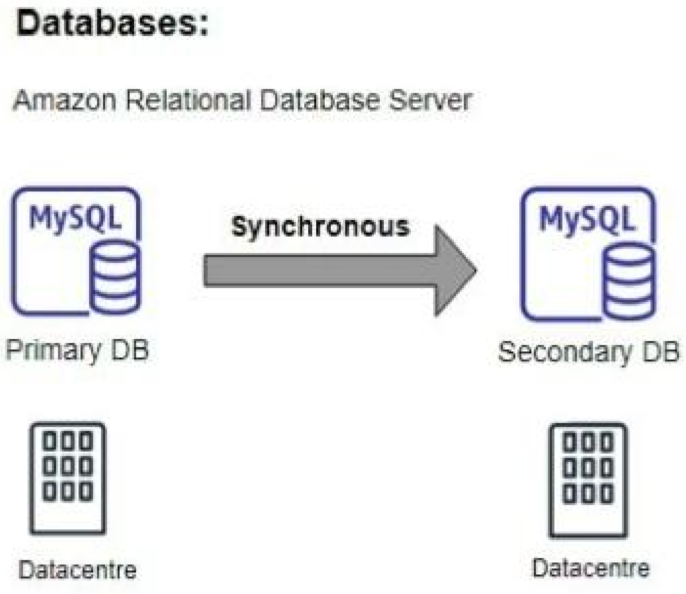
Where will the Amazon RDS be used? Let us assume that you have an existing database such as Oracle workload that you already have in your on-premise. Now many organizations sometimes have a large proportion of data on their on- premise locations. If they choose to switch to the cloud, they would prefer to transfer the data because they don't want the hassle of checking it on another database engine environment. Thus, they would normally migrate the Oracle data in the RDS to an Oracle database.

In this case, they could use a migration service in a database called DMS it might be an external tool or the already available service in AWS. So if something is already available on-premise, and you want it as a lift and shift then you can use the Oracle database service. Another reason you may want to use Amazon RDS is let's say you have an application from the third party and this relies on only working with certain database systems. So if you're using WordPress that operates more effectively with the PostgreSQL database, then in such a situation you could spin up the PostgreSQL database in the RDS.

The expertise of your developers in your organization could also be another factor. Let's say, they don't have the DynamoDB expertise. There is a lot of buzz about DynamoDB, so let's say you as a company is looking forward to using the DynamoDB service. You may want to shift your application from your on-premise location on to AWS, and then use DynamoDB, but then you need to look at the cost of the entire process.

Let's say your developers aren't ready to work along with DynamoDB, you don't have the money or the time to invest in upgrading DynamoDB developers. In such a scenario, if they prefer working with MySQL, you should choose a solution that is available in the RDS itself. Let us assume that you have structured data: if you have data that has schemas and contains a lot of table connections then you need to choose one of the available Relational Database Service and not DynamoDB. Remember that DynamoDB does not support table joints and a schema-less database. And note the key terms why I compare this to the context of DynamoDB, as you can get this contrast in terms of the multiple choices in the exam. You will have to decide sometimes whether you want to use DynamoDB or perhaps use a MySQL relation database solution. And this word structured data sometimes has a big influence on your judgment on what the right answer would be.

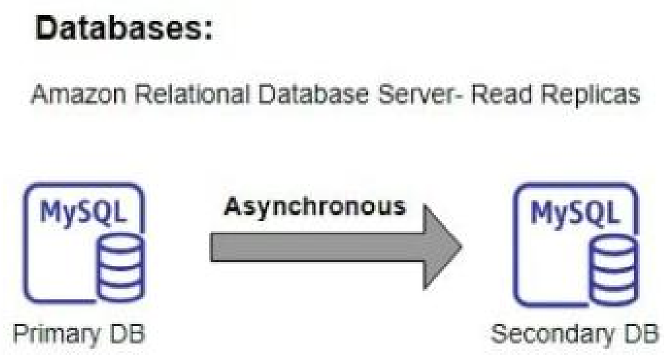
We will then discuss the high availability, so when it comes to RDS you will get a lot of questions on high availability. Let us say you want high availability for your MySQL Relational Database Solution, so enable the Multi-AZ feature. This feature is used when you have your primary database in one location, and you want to copy the data onto another location in case of infrastructure failure within a region. When Multi-AZ is enabled, this data is copied synchronously to a secondary database or another data center or availability zone.



Some of the important factors to remember are that only AWS has access to a secondary or backup database, you can't access it. In case your primary server crashes then AWS automatically moves your endpoint to a secondary or backup database, you cannot do it yourself. So there is also this option of Read Replica's in AWS for replication which we will talk about after multi-AZ. You should not select Read Replicas if it is an option for the high availability in the examination. Always enable multi-AZ, since multi-AZ is necessary for a relational database service for high availability.

The important point to note is that the feature of multi-AZ is the multiple availability zones available within a specific area or region so this does not duplicate the database in different regions. You will more likely to get questions on this topic in the exam. In that case, you can use your automated backups and snapshots in another region to make your database available for disaster recovery. And your snapshots can copy the snapshots of your database into another region and then use these snapshots to create a database in the primary region if it gets down. So this is all about AZ.

Let's discuss Read Replica's so the feature of Read Replicas is available with the PostgreSQL, MySQL, and MariaDB solutions. There's again your primary and secondary database and this time you've got asynchronous replication onto your secondary database. The important thing is that now you have access to the secondary database and both databases can be used side-by-side to give you an additional endpoint for the secondary database. Now if you have performance issues on your primary database, you should use Read Replicas in such a situation. This is one of the key points to remember.



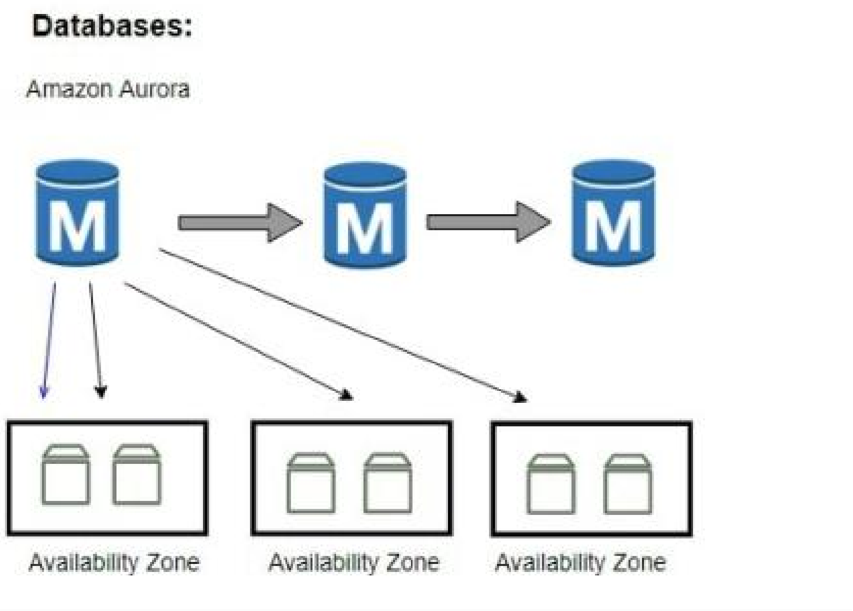
Let us then assume that you have a reporting application that takes static reports from your primary database. The number of users is on the rise and the delivery is affected. In this case, you should create an exact read replica of the database, and make it as a secondary database. After that, make the reporting application in such a way that it divides the workload between the two databases so that all the read operations do not go towards the entire primary database. They can also go to a secondary database.

Consider the same scenario in which you have an application but it reads as well as updates the data. The application can be created which updates the data in the primary database, and the reporting application can be used to receive queries from the secondary database. This is used to discharge the number of read operations on the primary database. We should use Replicas Read at that point.

Amazon Aurora: when should we use a database service of Amazon Aurora? Remember that Amazon Aurora is MySQL and PostgreSQL compatible database. If you want a faster MySQL engine, if you want improved performance, if you want to be fully managed, you do not want to manage anything on your own, choose Amazon Aurora. If your IT management team doesn't have time and maintenance resources or if you want high availability, then choose Amazon Aurora database service.

Now I want you to know what Amazon Aurora is doing under the cover when you say high availability. When you create a database of Amazon Aurora, you can create a cluster and you can also create volume clusters. When you write data on a table in the Amazon Aurora database, it duplicates the data or copies it onto six volumes. These six volumes are divided into three availability zones, which is a high number of volumes. It makes it more highly available in all places so if one AZ goes down the data in the other availability zones is still accessible. It, therefore, entitles all the volumes in all the availability zones and will perform a read operation to just one volume cluster in a single availability zone.

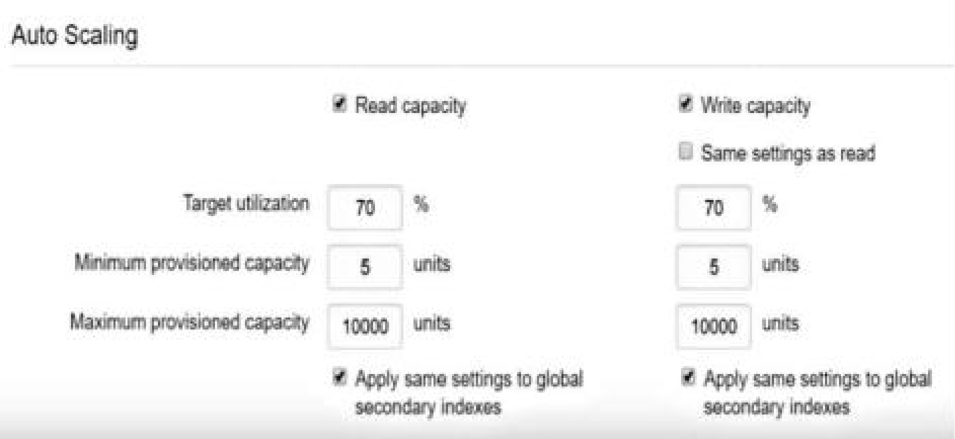
It also has this Read Replica's concept which we have learned in the RDS case. The same goes for Read Replicas in Amazon Aurora too. If you want to offioad the read operations onto the secondary database, the data in Amazon Aurora will be copied to these read replicas. The latency is less than 100 milliseconds. This is also a very important concept so, use the services of Amazon Aurora for less than 100 milliseconds latency rate to duplicate data across your different endpoints in the cluster.



Now let's see the Amazon DynamoDB database. You should use it to have a completely managed NoSQL database. You are storing files or JSON related data when you require fast data access so you want to quickly get and retrieve the data and begin using it and you have no joints in your table. You also have the index concept. You can define local and global secondary indexes in DynamoDB to search for the data on various attributes for quick access. It also has very high availability and durability. It is a fully managed service so it automatically scales the resources based on demand.

In the exam, if you'll get a question like you need a fully managed NoSQL database that could scale up or scale down beyond demand then you don't have to focus on demand. You just need to add the data in the DynamoDB tables and then it will be

the responsibility of AWS to ensure that the servers are scaled for your database table automatically. This new feature is known as Auto Scaling that is another very important concept. Remember that for your DynamoDB table, you must set the read and write capacity. It allows you to specify the number of input and output operations on your database table. The cost you pay for the DynamoDB table is also affected.



Auto Scaling

	☑ Read capacity	☑ Write capacity
		□ Same settings as read
Target utilization	70 %	70 %
Minimum provisioned capacity	5 units	5 units
Maximum provisioned capacity	10000 units	10000 units
	☑ Apply same settings to global secondary indexes	☑ Apply same settings to global secondary indexes

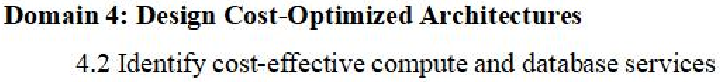
However, let's say that at the start you run an application that is not widely used so you set a low amount of read and write capacity for it. Then after a few months, you started to notice the unexpected throughput, you are getting absolute errors on your DynamoDB table. Here comes the Auto- Scaling service where AWS automatically scales your infrastructure instead of you manually adjusting the read and write capacity. It can be done when creating the database table where you have to specify that if the read/write capacity exceeds a specific threshold then automatically increase the units. As automation in the industry is a valuable thing so you may get questions on DynamoDB auto-scaling.

Now let's go onto Amazon Redshift: this database service is column-based and it is useful for aggregated data. So when you need to perform operations like sum, averages, minimum, and maximum on huge amounts of data, then think about Amazon Redshift in this case. Remember this is a petabyte database facility that can easily be scaled if you want data from multiple sources and it scales up to petabytes with only a few hundred gigabytes of data initially. So if you've got a lot of data then you can choose Redshift as it is a data warehousing solution.

In the traditional SQL-oriented databases such as MySQL, Oracle or Microsoft SQL, the data is stored on a disk horizontally, but since Redshift is a column-based database so the data is stored on the disk vertically. That is why it operates on your data faster in a column-based format. And this is why it's good if you've aggregated or want to aggregate your data. With Redshift, you can also use the business intelligence tool. All these points can be asked in the examination and then as an architect you have to choose which database service is suitable for your architecture.

And if you want disaster recovery, then you can use another service known as cross-region snapshots. With this snapshot data, you can recover the data onto a new cluster in any available region. It is a very important concept from exam perspective. This is the end of this chapter where we discussed the databases from the exam point of view. Now let's proceed to the next chapter in this book.

Compute Services



Domain 4: Design Cost-Optimized Architectures

4.2 Identify cost-effective compute and database services

Let's talk about the compute section. So in this chapter, I will discuss the concepts that align with objective 4.2 which is to determine how to design cost-optimized compute services. So let us first look at instance pricing. This is very important to know when you look for cost-optimized compute solutions in AWS. So there is a lot to the price of the EC2 instances. We, therefore, have on- demand instances, which is ideal for developing and testing environments. If you want the instances for a short period, let's assume that you have a test environment for one month and a development environment for two months then you should use on-demand instances. If you have batch processing operations, you will require spot instances where the operations will get through from interruptions and interruptions are because you place a bit price on that spot instance so when you lose the bit, you will lose the instance. There are other processes such as hibernation, but from an exam perspective understand that it's best to use a spot instance if you have batch processing operations.

Next, if you know you require a 24/7 facility and you need it all the time then you can save on costs by purchasing the reserved instances. Two more options are also available: first is a dedicated instance and the other one is a dedicated host. In the first case, the instance runs on hardware dedicated to a specific client. This might be a company with multiple AWS accounts and hardware dedicated to them. However, if the user has several AWS accounts, then instances that are launched on these accounts will use the same hardware.

On the other hand, there are some occasions when you require dedicated hosts, you want a physical server and you want full control over it. Dedicated hosts are used when you have a third party application in which the license is based on the number of cores. You cannot depend on the idea of AWS's virtual cores. You need this concept, and I've seen applications that have strict licensing policies in which you must have physical core as part of the contract. In this situation, you must use dedicated hosts or perhaps have a safety policy that specifies not to share your infrastructure with any other instance. You must use a dedicated host in such a scenario.

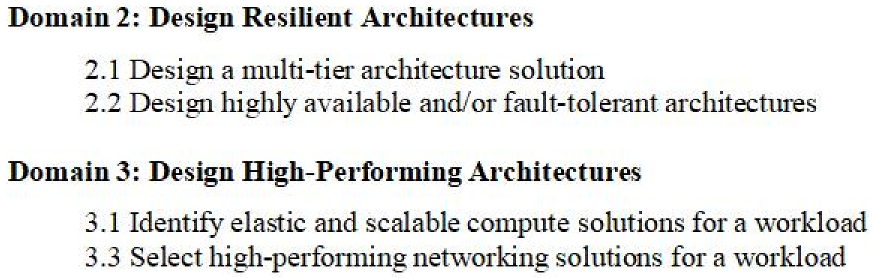
Now we will see our Serverless Computing option: so, when you want to manage the underlying infrastructure, then AWS Lambda is a good option. And you only get charged for the amount you're using in AWS Lambda. With AWS Lambda, porting your current code and saving on costs gets very easy since you need not be bothered about the expense of your underlying EC2 instance or your existing EBS volume. You only pay for the amount you use. This Lambda function is usually used together with the API Gateway, so it's always an API Gateway with the AWS Lambda if you get a question about service combination. APIs can be generated by customers in the API gateway service and you can then have the Lambda functions triggered by these APIs internally.

Then we have our Elastic Container Service and I want to discuss it here since we already have microservices and so many companies use microservices for building their architecture. It can, therefore, be used to orchestrate your containers. So rather than installing an orchestration service such as Kubernetes on EC2, you can use an elastic container service to handle all docker containers for you. Here's something you define which is known as types. In the type, you mention the image you want to pull down. This can be extracted from the S3 elastic container used in AWS and then containers are allocated to managed instances. All of this can be accessed by this service. All of this is automatically done for you in the elastic container service as it is a fully managed service. It also comes along with Auto-Scaling capability. Thus use the elastic container service to handle an orchestration entirely for you.

Let’s quickly see the difference between ENIs, ENAs, and EFAs from the exam perspective. You can use the ENI if you want to add a network interface for your instance or application. If you do not have any high- performance requirement, then you can use ENI as it is the basic adapter type which you can use with all types of instances. ENA is used in situations where higher bandwidth and lower inter-instance latency are needed. It supports limited instance types (HVM only). On the other hand, EFA is used for high-performance computing, MPI and ML use cases and tightly coupled applications. It can be used for all types of instances. This marks the end of this chapter.

Multi-Tier Applications

In this chapter, we are going to look at Multi- Tier applications. I want to discuss the following objectives which identify these concepts. So in the domain of Designing Resilient Architectures, we have the objectives of how to design a multi- tier architecture solution and how to design high availability or fault-tolerant Architecture. And you have elastic and scalable compute solutions in terms of performance, so we will talk about elasticity when considering Elastic Load Balancer. Then finally, this chapter covers a part of the objective of selecting high performing network solutions for your workload.



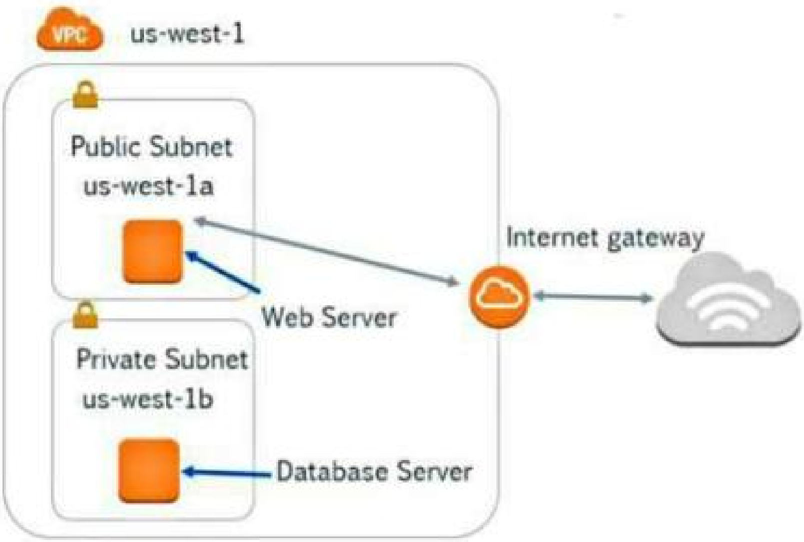
Domain 2: Design Resilient Architectures

2.1 Design a multi-tier architecture solution
2.2 Design highly available and/or fault-tolerant architectures

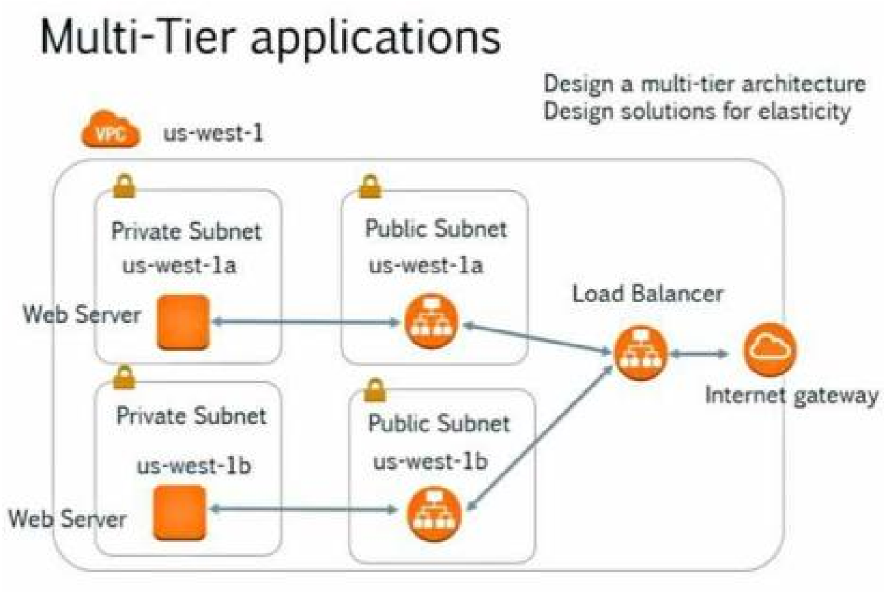
Domain 3: Design High-Performing Architectures

3.1 Identify elastic and scalable compute solutions for a workload
3.3 Select high-performing networking solutions for a workload

I would like to address the VPC in particular and how to build your Elastic Load Balancer coupled with instances. Let's discuss a use case scenario that outlines these objectives. Consider a very simple architecture: you have your VPC in which you have a public subnet and a private subnet. You may have your web server in a public subnet in one AZ, and you may have your database server in the private subnet. And internet access is available to the public subnet through the internet gateway.



Now let's assume that you want to build a multi-tier architecture and you want an elastic design with a high-availability and you will then want to place a load balancer in front of your web server. What you'd do is to have several Private Subnets, and your web servers will be put in these Private Subnets. Here we used Private Subnets because it's good from a security point of view. Note that you do not need to place the web servers in your Public Subnet if you are adding a load balancer in front of web servers. They can be in a Private Subnet because the load balancer communicates through the private IP's with the underlying web servers and not via the public IP's. So it's not your underlying web server that needs internet access but it's the load balancer actually that requires access to the Internet gateway. So we've got two subnets in this design. This is therefore great to use several subnets for your web servers.



Then when creating a load balancer, it should be formed in different public subnets. Now, what is the service of the load balancer? The load balancer service will make multiple load balancer nodes in each of those subnets when creating the load balancer in multiple subnets. Thus the load balancer itself is a highly available service. For high availability, you do not require multiple load balancers. This is a very common question that is asked during the examination. Do you need multiple load balancers for high availability if you already have one? Of Course not because the load balancer itself is a highly available solution so, you just need to do one

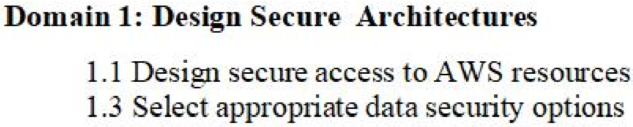
thing when creating your load balancer i.e., to specify multiple public subnets.

And what will AWS do if you have an Internet- facing load balancer? It creates a load balancing node that is used to distribute traffic and creates this node for each public subnet. The traffic is distributed to your underlying web servers by these nodes. That's how the underlying architecture works for you. This service generates multiple load balancing nodes in different subnets that correspond to multiple availability zones. You will only see a DNS name or one load balancer, but it happens underneath. Therefore, for your load balancer, you always have to create multiple subnets. This is the simplest way to design an architectural solution.

There are, however, several subnets for a high availability approach that map to different availability zones that are private to your Web servers. When you use an Internet-faced load balancer, we have several subnets for different areas of availability. If you use an Amazon RDS for your architecture, then your web servers must communicate with a back- end database server. And if you need high availability for your entire architectural solution, you can enable the feature of Multi-AZ for your relational database service.

Thus, I am putting this forward as an architect to recognize how you can achieve high availability for your entire infrastructure when you have an Elastic Load Balancer, have a web server and use the Relational Database Service. Now we shall move to the next chapter of this book.

Security Practices



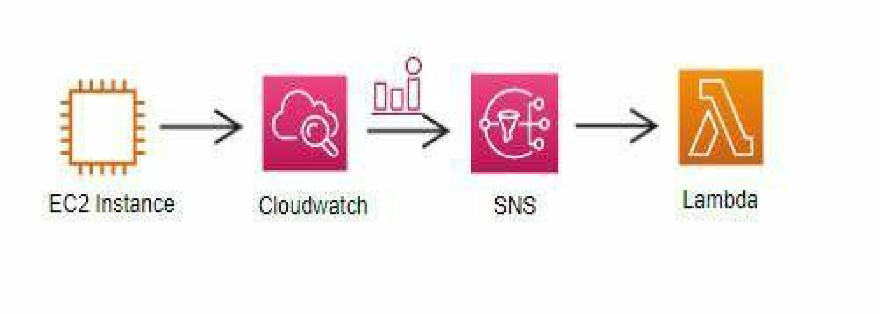
Domain 1: Design Secure Architectures

1.1 Design secure access to AWS resources
1.3 Select appropriate data security options

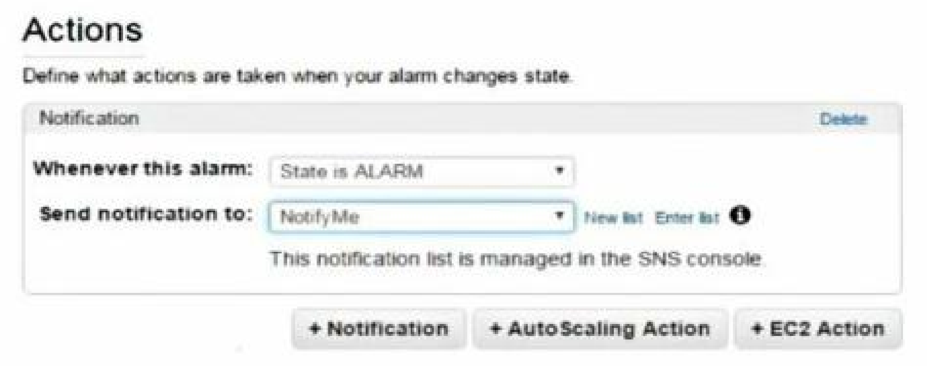
Now let's look at Security Practices which is very important from an architect's perspective for the latest exam. So in this chapter, I will discuss the concepts that align with objectives 1.1 and 1.3 of Secure Architectures Domain, which are to determine how to design secure access to AWS resources and select the appropriate data security options. Let’s discuss the AWS data security options first from the exam perspective.

So the first one is the Amazon CloudWatch which is very important for the CSAA exam. Now you can get many questions on the use case scenario related to CloudWatch in the examination but probably all the questions will come along with either CloudTrail logs used with it or AWS Lambda function invoked with some backup processes. So you must find that one hint to answer such type of questions in the exam. Don’t confuse yourself with these tricky questions.

Let’s discuss CloudWatch. So Amazon CloudWatch is simply a cloud monitoring tool for tracking your AWS resources in the cloud. You can monitor the traffic coming in and going out of your resources. You can set up alarms on your CloudWatch so it will trigger an alert for you when there is a deviation from regular system behavior so that you can go in and see what the problem is. You can also use CloudWatch logs and events. Let’s discuss a use case scenario from the exam perspective. Consider a use case that you want to launch a Lambda function wherever your CPU utilization of your instance goes beyond 40%. Suppose you want to configure the metric for CPU usage and when a 40 percent metric is crossed, you want to set an alarm so that it notifies the SNS NotifyMe about it. Now as you want to monitor the CPU usage of your EC2 instance, you will have to choose EC2 metric and in the list of matrices let’s say you want to view Window2 instance.



Then you are going to see the graph of the CPU utilization for your Window2 instance. Then go to Alarms to create an alarm. You can specify alarm details in the action tabs when creating an alarm. So once the alarm is triggered, then you can send your notification to your SNS Notify Me. For you to receive a notification, first you need to create a topic and subscribe to it either with your email ID or Lambda function in your SNS. If you have provided your e-mail address in SNS, then whenever a notification comes to the SNS NotifyMe service, it sends you an e-mail with that message.



Actions

Define what actions are taken when your alarm changes state

Notification

Whenever this alarm: State is ALARM

Send notification to: NotifyMe

This notification list is managed in the SNS console

So, with this alarm, you'll get a message that something happened in the CloudWatch. But in our use case, we need to execute a Lambda function, so I'll pick Lambda function in SNS. So wherever your CPU utilization of your Window2 instance in the graph goes beyond 40%, CloudWatch will trigger an alarm that will notify the SNS to launch a Lambda function. The Lambda function in the NotifyMe will be launched to clear all the background processes in the ECS instance and the utilization of the CPU will come down automatically. So in the exam, you will likely get questions on these types of CloudWatch use cases.

The Amazon CloudTrail is another security option. In terms of security, this is a very important service. The AWS cloud trail can be used to track all the API activities from your AWS account, either by making API calls from SDK or PowerShell or by using a console. All will be tracked with the help of your cloud trail service. It's very useful when you need to guarantee the company's compliance. You can also search the cloud trail logs when you detect unauthorized activity in your account. Let's assume that many resources exist in your account that should not be created. You can check these API calls in the CloudTrail logs. A solution architect should always activate CloudTrail logs for all regions. In case, if AWS creates any new region, then the API calls are automatically covered for that particular region as well. So when you enable CloudTrail logs for all regions it ensures that all API calls get covered.

Now let’s discuss the IAM from the security perspective as it covers both objectives in this chapter. When creating IAM users, make sure that the least privilege access is given to them so that they will only do certain tasks and show that you grant them access based on these tasks. If necessary use multi-factor authentication and update the password policy, do not enforce the same password policy. You can mention things in your password policy, like the characters you specify when the password is formed, the length of the password, and the password expiry, and all of these can be specified in your password policy. Deactivate the root access keys.

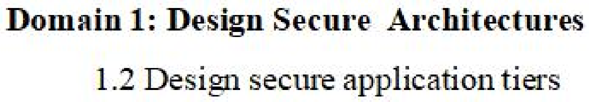
We have the bucket policy for the buckets in S3, which helps us to manage the access via the underlying objects even when allowing access to other AWS accounts. Note that you can do this with the help of the bucket policy. Now, if you do not want to allow public access to the whole bucket, but still want to allow certain users to access certain objects, Pre- signed URLs are one way of doing this. So you can specify the time limit in the pre-signed URL for a user to access an object.

Next, the IAM Roles are used to provide secure access to your AWS resources. So assume that you run an application on an EC2 instance and it needs to access S3 or DynamoDB. So make sure the instance has an IAM Role attached to it with the specific privilege. That is extremely important. It's all right to use access keys during development time, but make sure to use IAM Roles when you are going into the deployment phase for the safe and secure access. Even if a Lambda function is being used, always attach an IAM role to the Lambda function when it accesses an external service like DynamoDB or S3.

In terms of network security, we have seen before that a NAT instance or NAT gateway can be used to enable an instance to access the web in a private subnet. Remember that you cannot use a NAT gateway with an instance in a private subnet that accesses public resources such as DynamoDB, S3, or KMS. A special function known as VPC Endpoints is then used for this purpose. There are two VPC Endpoint types: first is Gateway Endpoint and the other is Interface Endpoint (Elastic Network Interface). If you want an instance to access S3 or DynamoDB, you can use gateway endpoint. And you can use interface endpoints to access other services like KMS. So you create a VPC endpoint for your application, connect it to the VPC, and then enable the instance to access the resource in your private subnet through the endpoint.

For Redshift, if you want the data that you need to load or copy to be privately held, then it should not go through the internet but only through the VPC. Then one function known as Redshift Enhanced VPC Routing can be used. You can use VPC Flow Logs to monitor traffic IP addresses in your VPC. And you use Bastion host as an administrator to administer instances in your private subnet. The bastion host is included in the public subnet to ensure that the right security groups are in place so that the user only has access to the admin workstation. That is the end of this chapter.

Security Of Virtual Private Cloud



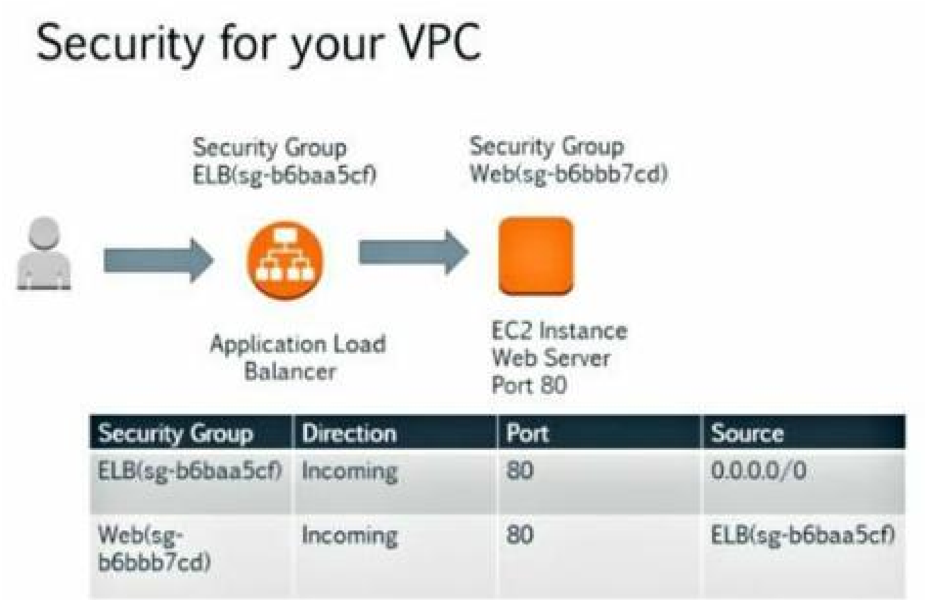
Domain 1: Design Secure Architectures

1.2 Design secure application tiers

I want to discuss the Security of the VPC in this chapter. This chapter mainly focuses on the objective of securing your application tiers, which is considered important for secure architectures. You can secure your VPC in two ways: First is Security Groups and the second is Network ACL. You can use Security Groups to manage the traffic to your EC2 instances. Also, note that all traffic is denied by default. And you can use Network ACLs if you want to restrict the traffic on your subnet. An important concept to remember for the exam is that when you introduce a rule into the Network ACL, it affects all instances that are included in your subnet. So when changing Network ACLs, you have to be very cautious.

Next, the question of having malicious traffic is a very common one that can be asked in the exam. So if you get suspicious traffic from a group of IP addresses to your EC2 instance in a specific subnet, you could use Network ACLs to put a denial traffic rule on these IPs. When setting up Security Groups for your VPC, let's assume that you have an Application Load Balancer so that users access this load balancer on your instance.

Let's say you host a web server on your EC2 instance that listens on port 80. As far as Security Groups are concerned, you must remember that your ELB and your webserver both have different security groups. This is very important from the exam point of view. So, remember to put a rule for the traffic coming on port 80 from a source for the ELB Security Group. Now let's assume that you have an internet-facing load balancer and user needs to come from the internet. Let's say port 80 only accepts HTTP traffic from the source anywhere and for the incoming traffic too. If you have an SSL certificate installed on your load balancer and you allow traffic on port 443, and then make sure that the ELB security group accepts on port 443.



Security for your VPC

Security Group
ELB(sg-b6baa5cf)

Security Group
Web(sg-b6bbb7cd)

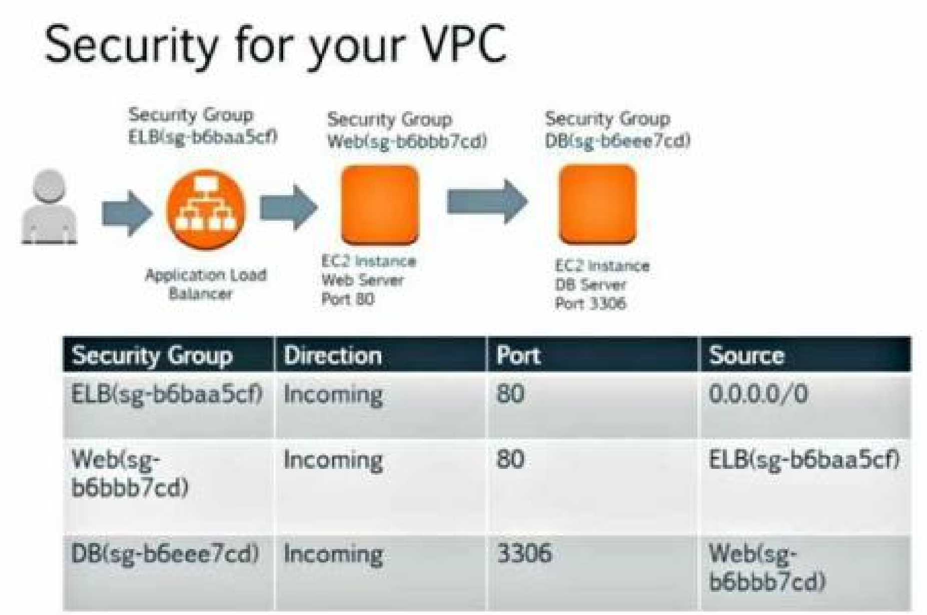
Application Load Balancer

EC2 Instance
Web Server
Port 80

Security Group	Direction	Port	Source
ELB(sg-b6baa5cf)	Incoming	80	0.0.0.0/0
Web(sg-b6bbb7cd)	Incoming	80	ELB(sg-b6baa5cf)

The next thing is the Security Group for your web instance so traffic from anywhere should not be permitted. The incoming traffic should be allowed only on port 80. Let's assume you only allow HTTP traffic so it is only accepted from the ELB source and not from anywhere. For the exam, this is a very important concept. In case your web instance starts to accept secure traffic then you should change your port to 443. If you have this setup, where you have an Application Load Balancer, web server, and a database server, then these are the rules that you would have to enforce.

Now the only difference is that the Database Security Group should accept incoming traffic based on the type of database server so let's say it's a MySQL database server over here. Thus it allows traffic on the port 3306 by default and here the webserver will be your source.



Security for your VPC

Security Group
ELB(sg-b6baa5cf)

Application Load Balancer

Security Group
Web(sg-b6bbb7cd)

EC2 Instance
Web Server
Port 80

Security Group
DB(sg-b6eee7cd)

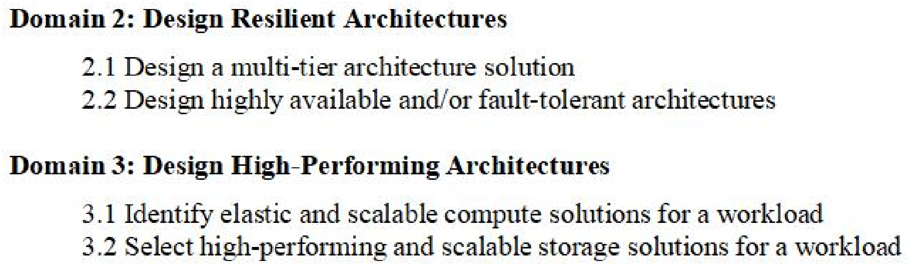
EC2 Instance
DB Server
Port 3306

Security Group	Direction	Port	Source
ELB(sg-b6baa5cf)	Incoming	80	0.0.0.0/0
Web(sg-b6bbb7cd)	Incoming	80	ELB(sg-b6baa5cf)
DB(sg-b6eee7cd)	Incoming	3306	Web(sg-b6bbb7cd)

You must understand this important concept of the incoming traffic path, port numbers, and source in the first place. When it comes to the exam, the source is considered to be very important. So this marks the end of this chapter.

Network Address Translation

I am going to discuss NAT or network address translation in this chapter. This falls under several layers: the design of multi-tier architecture solutions, high availability, elasticity and scalability and even the high-performance networking solutions for a workload. The questions you may be asked about NAT would then fall into any of those categories.



Domain 2: Design Resilient Architectures

2.1 Design a multi-tier architecture solution
2.2 Design highly available and/or fault-tolerant architectures

Domain 3: Design High-Performing Architectures

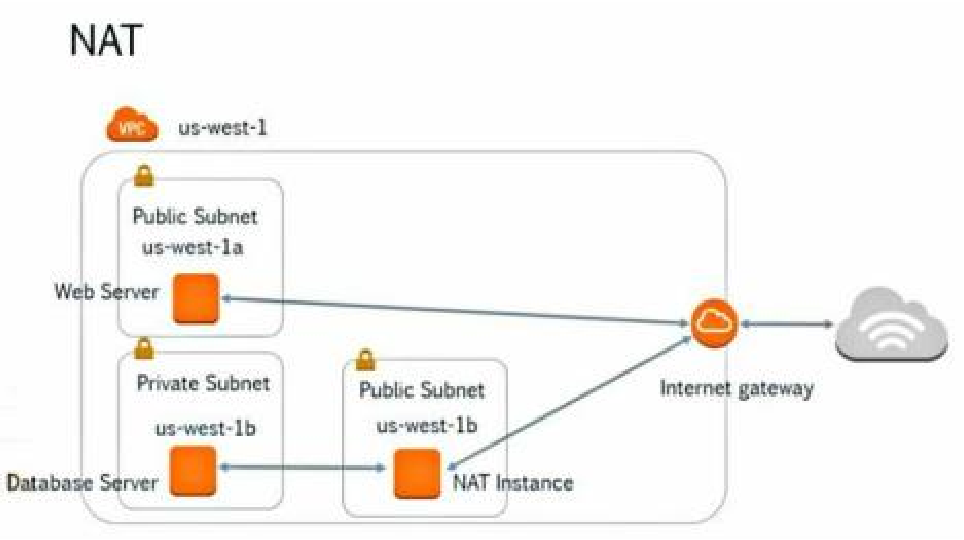
3.1 Identify elastic and scalable compute solutions for a workload
3.2 Select high-performing and scalable storage solutions for a workload

Let's try to figure out what could be asked in the examination. When considering NAT itself, you have a web server in a public subnet and a database server in a private subnet. Then you create the NAT instance to make sure the servers in the private subnet can connect with the internet securely without using internet gateway. Using NAT is useful because in this situation only the database server can access the internet, but no one from the internet can communicate with the database server. That's just one-way traffic. Therefore, if a message is sent via the NAT instance from your database server to the internet, the response would be received, but there is no way to send a request from the internet to the database server via that NAT instance.

This is not allowed when using a network address translation in which security factors exist. So now it's essential to ensure that the NAT instance is defined in the public subnet rather than in the private subnet. It is very important from an architectural perspective as the NAT instance should be able to communicate with the Internet gateway over the web. You have to remember that the NAT instance is used in your private subnet as an intermediate translator for the server.

You can now use a fully managed service called NAT gateway instead of the NAT instance. The NAT instance can easily be replaced with a NAT gateway, but you must make sure that when creating a NAT gateway it must be defined in public subnet. The NAT gateway is used where the NAT instance becomes a bottleneck. So let's assume that your private servers have many requests, it could be updated to be installed onto your servers. And then your NAT instance relies on the type of instance with a bandwidth limit which makes these NAT instances a bottleneck. The NAT gateway has high bandwidth and can allow a large amount of data flow to and from your subnets.

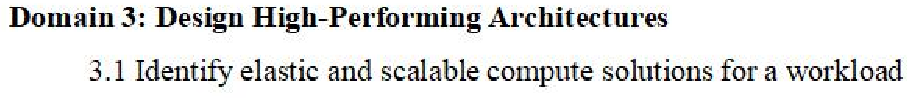
If you want to have a fully managed service, you can also use a NAT gateway. If you use a NAT instance, you have to manage the NAT instance and ensure that the instance is always functional, and you are also responsible for the instance security. But if you need a fully managed service, then use the NAT gateway service.



So when are you interested in using a NAT instance? You want to use this one as a proxy server. You can now create an Auto-Scaling Group in case you want the high availability for your NAT instance and then set the launch settings for your NAT instance. This point is very important from an exam perspective. And for high availability, you can also place multiple NAT instances in different availability zones. And if you want high availability for your NAT gateways, you can create multiple NAT gateways in different availability zones.

Thus we came at the end of this chapter. In this chapter, we discussed NAT and NAT high availability in your VPC from the exam perspective.

Elasticity & Scalability



Domain 3: Design High-Performing Architectures

3.1 Identify elastic and scalable compute solutions for a workload

Now we are going to have a quick review of elastic and scalable solutions for our workloads in this chapter. This is particularly appropriate to define High-Performing Architectures here to build elastic and scalable solutions.

Now we can use Elastic Load Balancer for elasticity as it can be used for distributing data to underlying EC2 instances. You can use an Auto-Scaling Group for scalability. You can now run instances of different metrics in an Auto-Scaling Group. You can also identify your custom metrics in CloudWatch in addition to normal metrics such as CPU utilization. And, since this isn't always the case, you can have function level metrics.

The reason that your application has issues and it needs to scale up is not just because the CPU is high and your disk input/output operations are high, but it can also be dependent on the performance of your application. You can, therefore, use CLI to upload your metrics onto CloudWatch from your application. It allows you to set up alarms and then enable instances in your Auto-Scaling unit to scale up or scale down depending on these metrics.

Now we will discuss the Auto Scaling Policies: I would like to explain the Schedule Scaling Policy from the viewpoint of the exam. This is required in a situation where you want to scale up or scale down an instance at a specific time, so the Auto-Scaling is scheduled here. Now some of the ways you can do this: imagine you have a special event for your application and you have to make sure that according to the requirement, your infrastructure is scaled prior to the event itself. At the last minute, you don't want to delay the scaling of your infrastructure that you want before the event can begin.

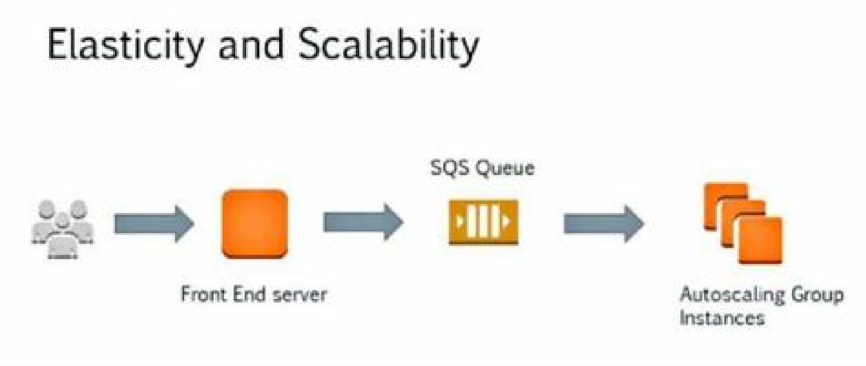
Another possible scenario says that a team uses an application that is described as part of an Auto- Scaling unit on a number of instances. So let's assume this program is pretty heavily utilized early in the day. Let's consider a case scenario in which people arrive at the workplace at 9 a.m. and during that time, everyone uses this particular application only at that point and then it's not used too much later in the day. What you can do is to make sure to provide consumers with the best experience early in the morning. Therefore, you can have a Schedule Scaling Strategy that adds instances early in the morning as part of an Auto-Scaling unit.

Then you might add another policy of dynamic scaling to minimize it after utilization gets back to usual. Since we have the Dynamic Scaling Policy so that the instances within the Auto-Scaling unit can be scaled dependent on these matrices, so this can be built on existing matrices such as the CPU utilization or you can use on your custom metrics.

The important aspect I want to address here is that of the cooldown time period from an exam viewpoint. This is also important for the Auto-Scaling group and what exactly is this time of cooling down? Let's assume that the Auto-Scaling unit spun three servers initially, so this is the minimum amount of servers you stated in the Auto-Scaling launch setup. Let's just assume now that you are starting to take a performance hit, your application has some issues, and that it has already set off a CloudWatch alarm at 9 a.m. Auto-Scaling now spins up to more servers as part of that alert so there are five servers now. So new software has to be installed for these two extra servers, and several scripts have to be run so that these two servers can join the first three servers and begin to accept the requests.

Let us assume that it lasts ten minutes because once you scale up new instances, you have to have something run on it and it has to be part of that group to begin to accept the requests. Then in another five minutes, as the new servers are not ready, your three servers have taken a hit and the alerts have been activated again in your CloudWatch. What's Autoscaling going to do? Auto-Scaling continues to unnecessarily spin up new servers again. Instead of five servers, you will tend to have seven servers and this could last for a certain amount of time. The reason behind this is that you didn't give those five servers sufficient time to settle down, that's where you get the cooldown timing period. What you can do is that you can extend this time duration to ensure the infrastructure has sufficient time to adjust. If alarms are activated by the CloudWatch during this time, they will be neglected. This provides sufficient time to properly scale your infrastructure and begin to accept requests. From an exam point of view, this is important.

Now the SQS service is one of the most popular decoupling services in terms of scalability. So it's very easy, you have consumers and they might go to a front end server. You can send messages to an SQS queue from your front end server, and then have an Auto- Scaling group that takes these messages and processes the necessary information. Now the scaling of the Auto-Scaling Group can be triggered even based on the number of messages in the queue.



Let's just assume that more requests are coming and the queue length increases, the number of servers in your scaling group does not process the messages in an ideal way. The scaling process can then be adjusted to increase the number of servers depending on the number of messages in the SQS queue. This, therefore, marks the end of this chapter.

Encryption

We'll look at Encryption in this chapter. Speaking of EBS volumes, note that the customer key that is specified in the AWS KMS feature enables encryption for your EBS volumes. This must be performed during the creation time of the volume. If the volume already exists, you can’t allow encryption for that volume but you can allow the encryption of your files on that volume by using OS-level tools like BitLocker for Windows.

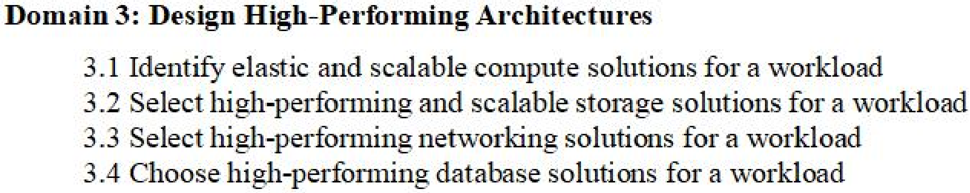
And when considering Amazon RDS, you can also enable encryption of the underlying database files. This applies to Amazon RDS and Amazon Aurora databases. One important thing to note is that when the encryption is enabled, all locks and snapshots are automatically encrypted. While creating the database, you have the screen to allow encryption and choose the master key to use for that encryption. It comes from the service of KMS. Even with DynamoDB, you can only allow DynamoDB table encryption at the time of development.

To enable encryption in S3, you can use Server-Side Encryption. However, in S3, AWS encrypts the object and stores it on the server before the object is stored on the physical server. If you send a get request to Amazon S3, it will automatically decode the object for you and return the object to you. Now you have three choices for enabling your Server-Side Encryption: the first option is to use AWS Managed keys. So, you don't have to think about anything here because everything is handled by AWS. This is perfect for businesses that don't want the key management problems. The second option you can use something in between is the AWS Key Management Service. You manage the keys here but you can't have full control over the key material but you still have the benefit of controlling the keys life cycle in the KMS service. Then, in KMS, you can also create customer keys to encrypt objects in the S3 bucket. The third option is using Customer Managed keys so you can use them to send the key as you import the object in your application. Then AWS picks up the key and encrypts and decrypts the object.

On the other hand, you can use Client-Side Encryption in which you first encrypt the object at the programming stage and then send it to the S3. The Key Management Service from AWS is now completely managed. It helps you to define your keys and handle your key lifecycle. For object encryption, you can use these keys. Thus, you use the customer keys for data key generation and then these data keys are used for encryption.

You even have the HSM (Hardware Security Module) service that is used when you are looking for full management of your keys. The keys you generate here do not have visibility to the AWS. However, you may want to use the service sometimes from a compliance point of view. If you use this service, your VPC is allocated a hardware device and you then access it through an IP. You will send a request from that IP address and then you will receive a key that can be used for the encryption and decryption of your data. This is the end of this chapter.

Performance



Domain 3: Design High-Performing Architectures

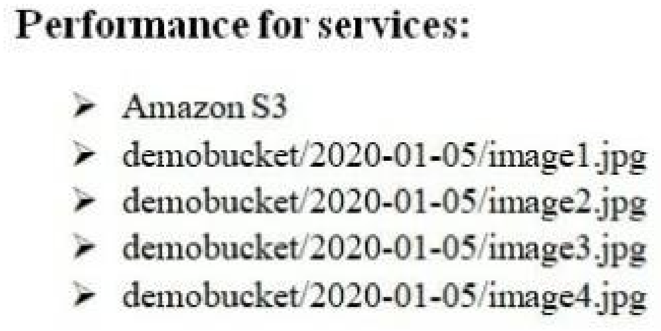
3.1 Identify elastic and scalable compute solutions for a workload
3.2 Select high-performing and scalable storage solutions for a workload
3.3 Select high-performing networking solutions for a workload
3.4 Choose high-performing database solutions for a workload

Let us discuss the performance of some AWS services in this chapter. This one comes under the performance objective, which from an exam viewpoint is very important. So when considering DynamoDB, one of the interesting things to get asked in the exam is DynamoDB acceleration (DAX). So the DAX is a DynamoDB in-memory cache. This is typically used when corporations use DynamoDB, and millions of requests are sent to the DynamoDB table per second. So they use this DynamoDB acceleration when they have to minimize latency. It generates an in-memory cache for DynamoDB to store the data needed quite often in the cache. It is a kind of elastic cache in front of a relational database service, but here there is an in-memory cache that is part of the DynamoDB service. This is useful if you have millions of DynamoDB requests and want to minimize the latency of access to that database.

The Amazon S3 is another important aspect in terms of performance from an exam point of view. So if you have workloads, let's assume you run an application of over 100 requests per second. This can be get or put requests for objects in the Amazon S3 bucket and then you need to worry about performance. Amazon S3 offers a service that scales accordingly to deal with the performance hit when the demands are below 100.

Let’s say if you as the consumer think that the number of requests can exceed a certain threshold, like 100 here, then AWS suggests you perform some practices when you upload your objects for high performance. One important thing to note is that you have the whole key name as you upload objects. This key name is used by Amazon S3 to create partitions and then save certain objects based on these partitions. And then in all the objects in a given partition, it creates an index. This is how Amazon S3 gets and puts an object in this service. So it's always useful when you have keys spread over several partitions to improve performance.

Let's take an example so let's say that you have a demo bucket and you save images in that bucket as per a specific date. So here all dates are fixed and constant, but the image's name or image key is changing so because you use the date for the underlying folder, what is the Amazon S3 service going to do? It is going to create a partition depending on the date and then store all of the image objects in that specific partition.

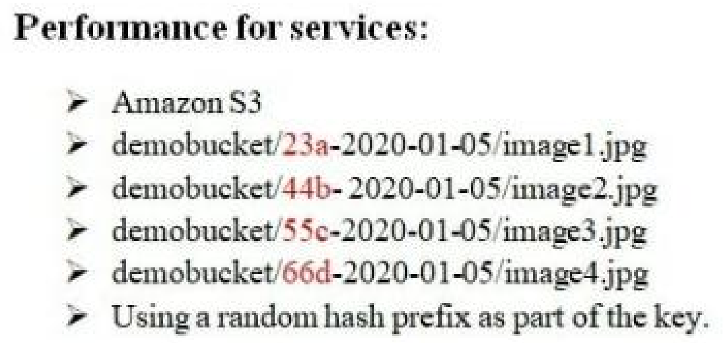


Performance for services:

- Amazon S3
- demobucket/2020-01-05/image1.jpg
- demobucket/2020-01-05/image2.jpg
- demobucket/2020-01-05/image3.jpg
- demobucket/2020-01-05/image4.jpg

Now, it's alright if you have a few objects, but once the number of items begins to grow, that's when you will get a performance hit because all inputs and outputs are focused on that single partition. So when creating keys, you must be careful to distribute them across several partitions. Through these various partitions, it gets simpler to fetch the required data.

So how could you change the key name to assure that you create multiple partitions when using the S3 objects? Let's take the same example, but this time we will add another key-value before each date, and these will be special key values. So AWS recommends that during the uploading process, a random hash prefix should become part of the key. This will improve the performance of both your get and put requests.



Performance for services:

- Amazon S3
- demobucket/23a-2020-01-05/image1.jpg
- demobucket/44b- 2020-01-05/image2.jpg
- demobucket/55c-2020-01-05/image3.jpg
- demobucket/66d-2020-01-05/image4.jpg
- Using a random hash prefix as part of the key.

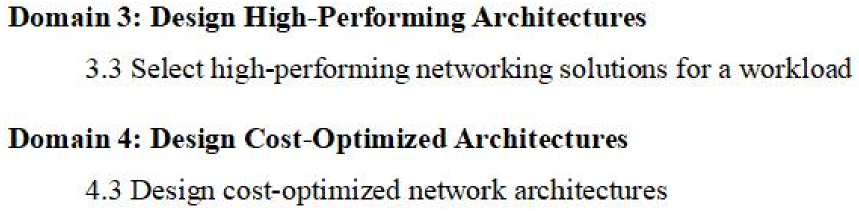
Let's move on to the performance of networking solutions. If you want to improve the networking capability of your EC2 instance, consider using EC2 instance with Enhanced Networking. You need to know it as an architect. If you want the data to be transferable between instances at low latency, think about putting them in a placement group. The only thing to keep in mind is that the instances in the Placement Group must be in the same availability zone because it is a prime prerequisite.

The next option is EBS Optimized instance, so you can choose this sort of instance if you want to manage the underlying volumes better. This type of instance will show that it is EBS optimized. If you want better links between your on-premises infrastructure and AWS, and a dedicated low-latency connection then you need AWS Direct Connect. If you want traffic encryption between AWS and your on-premise, you should use Amazon Managed Virtual Private Network Connections. Please note that AWS Direct Connect alone provides low data latency, but when it moves between the two places, then there will be no data encryption. So, if you are looking for both the encryption and low latency, then you can first create a Direct Connect link and then set up an AWS Virtual Private Network connection over the same direct connect link.

When you are an architect, it's very important to remember that you have to understand various aspects of the exam. So it's not only about environment management; it's also about security management, and network management as well. All these aspects are therefore very important for the examination. It marks the end of this chapter.

Network Architectures

So in this chapter, we are going to look at Networking Architectures from an architect's perspective for the exam. I want to discuss here the networking objectives of the AWS SAA-C02 exam. These objectives are selecting high-performing network solutions and designing cost-optimized network architectures for your workload. Now you can get some questions on the use case scenario of high-performing networking solutions in the exam so you need to understand these concepts in detail. Let's try to figure out what could be asked in the examination.



Domain 3: Design High-Performing Architectures

3.3 Select high-performing networking solutions for a workload

Domain 4: Design Cost-Optimized Architectures

4.3 Design cost-optimized network architectures

When considering Network Architectures, there are three types of network solutions for your workload. They are flat network architectures, segmented network architectures, and hybrid connectivity. Let's discuss flat network architectures first. So, in the flat network architecture, you have a single VPC architecture which means a single account in a single VPC. This is how Virtual Private Cloud was developed; you select a CIDR for a VPC and split it into subnets and individual availability zones. Then the subnets are placed into routing tables where they are subdivided into Public and Private Subnets. You can have five CIDR blocks attached to one VPC.

So in single VPC architecture, limited data transfer is to be considered unless you have dependencies that reach AZ boundaries. You can even have more than 300,000 IP addresses so your workloads can start to scale. And you can also split native constructs into Subnets Route Tables, NACLs and Security Groups. So if you have a VPC architecture in which you can have several hyper-scale customers who can operate hundreds of thousands of services like microservices within a single account in a single VPC, then the network architecture that is suitable for you is single VPC architecture.

Let’s say you have a single VPC that's Multi-Tenant, then you need to ensure that one of these tenants does not consume all your resources and not sharing with other users. So it can get rather difficult to have everyone in the same bucket when you talk of cost allocation or policy enforcement. Thus, VPC Sharing is the ideal approach to use single VPC with multiple accounts. It uses the Amazon Resource Access Manager to share your subnets and resources with other accounts in your AWS organization like the Transit Gateway, Route 53, and Resolving Rules within your VPC.

So in the single VPC multiple accounts or shared VPC model, you share a subnet to participants, and then they may launch their resources like EC2 instances or databases or ELBs within these subnets that you own. This is important from the exam perspective. VPC sharing has account-specific cost allocation, a single VPC blast radius, and shared DNS, but there is limited access control for application owners. Policy control and isolation that the individual AWS account provides can be implemented on that the single VPC layer. This method of centralizing and reusing VPC components reduce costs for the management and maintenance of your environment.

Ideally, when users look at these designs, they seek to optimize the costs, and that's why they place them into one single VPC because they have a lot of dependencies between workloads. Thus, users don't require mediums like the Transit Gateway, VPC peering, and private link for the VPC management with one single VPC. Let's move onto the segmented network architecture.

In Segmented network architecture, we have multiple accounts in multiple VPCs. So, this architecture is preferred for large companies that can control and push beyond the AWS account limits. So this architecture is required when users want isolation between their dev tests and production workloads or they want to isolate their business unit and workload category.

And another reason that it is designed is to completely isolate the individual microservices. For example, if you have a PCI or HIPAA compliance workload, then you can use it, or to simply separate the software testing and production environments. You can also isolate your blast radius if your workload crosses the VPC as well as the account boundaries. It provides you with the distributed service limitations that you can use to scale on AWS.

But, this approach increases the complexity in terms of IP management, like what type of IP range or CIDR do I assign for a bunch of my VPCs? How can I communicate with lots of other VPCs? You also need to consider how you handle access control between resources across different VPCs as well as IAM and networking accounts.

So here come several patterns of VPC networking connectivity. First, we have VPC Peering. Let's say you wanted to connect all your VPCs. In this case, you use VPC Peering to create a complete communication network between the VPCs. VPC Peering is easy to set up and has no bandwidth limitations. VPC Peering can be enabled for VPCs between the same regions, across regions or between different accounts. But, let's say you have four VPCs so you need to set up six peering links with the four VPCs and approve and configure routing for each of them. This could be asked in the examination.

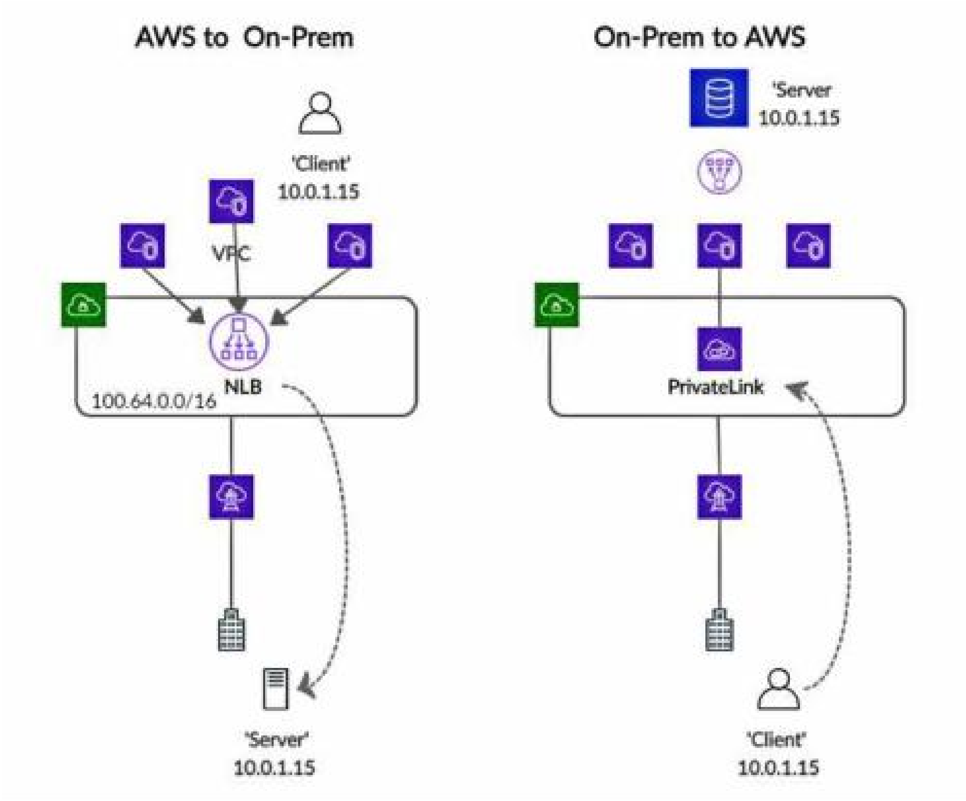
If you want to manage multiple networks and infrastructures, you can use Transit Gateway for that purpose. When you have hundreds of VPCs and connections, Transit Gateway eliminates the time- consuming process of linking individual VPCs with each other through VPC peering, as well as creating VPN tunnels between the on-premise and each VPC to allow on-premise connectivity. It uses either a VPN connection or Direct Connect connection. It's like you try to merge your edge connectivity with AWS. It scales up to 5,000 VPCs by default and supports equal- cost multipath VPN. And it also allows you routing flexibility and offers you multiple route tables. You must know the difference between VPC Peering and Transit Gateway from an exam perspective.

Let's see the concept of PrivateLink which is very important for the exam. Now every application needs to communicate with one another through three-way TCP handshakes. So PrivateLink is specially built for traditional TCP client-server relation where you have a client on one side and a server lives in a different VPC. You create a hole in another VPC to provide a service that's on a certain port along with an IP address to another VPC. Now, this is different from the VPC Peering and Transit Gateway, since it is just punching a hole for that service. It does not provide two-way connectivity for the entire VPC's. This is used because it entirely reduces the visibility to and from shared services. It also supports IAM policy on the endpoint itself and it solves large and complex network address translation.

Let's consider a use case in which you use PrivateLink to access dependencies on-premise that has overlapping IPs. In this architecture, you have on- premise services and you have cloud-based services and you want to share these resources. One option is to use a mediator or NAT VPC in this model because it allows you to set up bi-directional private links to expose endpoints in the cloud and on-premise. These are a kind of virtual interface that you are not targeting like a single server or maybe a load balancer on-prem.

So the IP address of your client in the VPC is 10.0.1.15 and your server on-premise has the same IP address. So here you put PrivateLink in the client's VPC. You have a Network Load Balancer in the mediator or NAT VPC and it has an IP target passes through Direct Connect to the on-premise database server. When the server receives the request, the request comes from the IP address of the mediator.

Now the reverse is that I have a private link in the mediator or NAT VPC, and I expose my service via a private link in the NLB within the VPC itself. The same IP addresses of the client-server solve the NAT issue as I mentioned earlier. You can see the difference in the diagram below:



Now how can you use PrivateLink for cost optimization? Thus, Centralizing interface VPC endpoint is one of the best ways of deploying PrivateLink to optimize costs. Some customers choose to host private link endpoints in shared services VPC because it reduces the cost of AWS PrivateLink endpoints based on the traffic profile. But you may end up paying for the data transfer costs when you have a Kinesis endpoint or something that uses a lot of bandwidth. Then you would just host the endpoint yourself.

Another thing to consider in this situation is when you create a PrivateLink; the service creates a private hosted zone for you. The PrivateLink service owns this private hosted zone. You don't own it so you can't connect it to other VPCs. Therefore, the privately hosted endpoints outside the VPC cannot be attached. Data processing costs can also increase according to how consumers communicate with a centralized endpoint. Thus, if you want to reduce costs and private link endpoints, that might be a good option for you.

Now, let's see the hybrid connectivity models that you need to know generally for the exam. First is AWS Site-to-Site VPN connectivity. The first option is to terminate the site-to-site VPN connection on a Transit Gateway. And then Transit Gateway spreads connectivity to thousands of VPCs. The second model involves terminating your Site-to-Site VPN connection on a Virtual Private Gateway attached to a single VPC. The third option is to terminate the VPN on an EC2 instance that runs VPN software from the AWS marketplace. You should choose to use the Transit Gateway for your VPN termination because it simplifies management by reducing the number of VPN tunnels and BGP sessions to handle. And it also scales horizontally in terms of throughput as well as the scale of the number of VPCs.

While VPN is a good option to start up, but may not be ideal for some production traffic. So, therefore choose AWS direct connect which gives high bandwidth connectivity between your data center and the AWS. This is all from hybrid connectivity that could be asked in the examination.

Let's discuss some of the differences between AWS Global Accelerator and CloudFront which is very important from the exam perspective. Both offer you the AWS backbone network to decrease latency for end-users. So if you have a website that is deployed just to one region, use CloudFront. It will also give you the capability of Response Caching and Lambda Edges. In Response Caching, you can cache static assets or full pages of your website close to end-users, which is very handy. And Lambda edge allows you to execute small pieces of software in a region close to the user in edge location. Due to this reason, CloudFront is a better option for minimizing latency for end-users but then, Global Accelerator offers you two public IPs, and because of that, it can work with any DNS system.

And in terms of failover between regions, the global accelerator is best because it offers you Multi-region failover. So, in case your website is deployed in more than one region, then it's good to use a combination of CloudFront and Global Accelerator to deliver content to end-users as well as to split traffic between regions.

In case you use API, which would never need caching then a global accelerator is a good choice because caching is not required. So global accelerator can distribute content and distribute API responses to end-users and also it can also switch traffic between regions. It also distributes traffic between multiple regions. Another feature of Global Accelerator is that its IP address is resolved regionally. This marks the end of this chapter.

Other Important Services

In the last chapter of this prep book, I am going to discuss a few more important services from the exam perspective. The keywords used here are going to help you in recalling the AWS architectural services and their operations during the exam. When you are looking at AWS services, the most important things to always look for are Ease-of-Use and Automation.

Let’s first discuss some services which are good in ease of use. If you have an application that focuses on micro-services and needs orchestration, choose Elastic Container Service. So you can organize the containers for your microservices program in the elastic container service which also supports Kubernetes. It has the Auto-Scaling feature as well.

Next, you can use Elastic Beanstalk if you want to provision development environments fast. Let's say you may have your dev team, you are the app's architect, and you want the dev team to work on your application's prototype and you might want to design a fast development environment. This can be done with Elastic Beanstalk. You can still use Elastic Beanstalk, even if you have a customized environment that uses Docker containers. If you get a question in the exam in which you are asked about the quick provisioning of development environments or if you intend to create customized environments in Docker, then choose the option of Elastic Beanstalk.

In the exam, if you're asked that you want your infrastructure to be automated then make sure to choose the CloudFormation. You can create templates in CloudFormation. These templates can be used for spinning up the AWS resources. Finally, if you have the necessary tools for configuration such as Chef that businesses use for their on-premise environments and you want to use AWS services, then consider using AWS Opswork as it provides chef recipes support.

If you want scalability, then use Autoscaling. So you must start using Auto-Scaling to manage the scalability of these different underlying layers whether it's your web tier, your application tier, your proxy servers, or NAT instances. Always remember to put the right conditions for scale in and scale-out. So when you are asked questions regarding scalability in the exam, then you as an architect shouldn't limit yourself to the only scalability of a certain tier such as the web tier. You can use the Auto Scaling Group (ASG) to scale different levels or layers in your application.

The next service is for managing deployments so the Blue-Green Deployments is considered to be the best way to handle deployments. You learn more about this when you're doing the DevOps engineering, but if you get a question on Blue-Green Deployments, consider using Route53 for the traffic control.

With the Weighted Routing policy, Route 53 can be used to route traffic on various environments in Blue- Green deployments.

Next when we talk about failure, remember that when designing an application as an architect, you must have a design with the fault in mind. If your web application crashes and you want an affordable failover website then, in this case, you can use the Simple Storage Service to get a static website. If your primary web instance or main website goes down, you can then have a failover policy in Route 53 to redirect your traffic to the static website in the Amazon S3. Note that Route 53 health checks can identify failures not only in AWS services but can also identify flaws of the services in your on-premises. These health checks can be based on the TCP protocol in which these health checks are done to check if your target services are operational and up-to-date.

If you have users worldwide and you want excellence in traffic management, so you can use CloudFront. From an examination perspective, it is very important. You may have an S3 bucket hosting a static website and you may also have an EC2 instance hosting a web application, all of this can be used as a source, then you can use CloudFront to deliver your traffic.

Now, let's discuss the services related to Automation. So the AWS Lambda is the service for automation. Think of AWS Lambda if you are asked about any form of automation. Some use cases are as follows: you can use CloudWatch Events to find if your instance is damaged. Then the instance can be terminated with the Lambda function. Take another use case in which you have an application that stores objects in S3. Now let's say that some metadata also gets stored together with the object in your application.

Let's assume that you have your privileged customers in an application who pays money for your service and that you want their objects to get more priority. You can then add metadata in your uploaded objects, say it's the privileged user, but how can you find out who the privileged users are and what are their objects? You can do it by creating a Lambda function that can be linked to an S3 event to read the metadata into the object when the object is uploaded and then create a row in a DynamoDB table. So for this kind of automation, you can use Lambda. So this marks the end of this chapter.

Handy Tips To Pass The AWS SAA Exam

Generally, best AWS training offers both theory and hands-on experience of the essential services on the platform. Three factors, including detailed information, exam guides, and sample questions, make it possible to pass any certification. Let us look at the detailed process of AWS Solutions Architect exam preparation even without basic AWS training.

Earn The Aws Certified Cloud Practitioner Certification

It is usually best for people to pass the AWS Cloud Practitioner exam who have no early AWS knowledge or experience. You can clear the exam in just two weeks by daily preparation. However, even with prior AWS cloud experience, it is still a wise decision to get it because this certification exam covers AWS service that the person may not have used in the past. Therefore, take notes all over the course and arrange them so that you can use it for the Solutions Architect exam as part of your study guide.

Complete Aws Certified Solutions Architect Associate Course

The learning process of AWS Solutions Architect Associate naturally builds from AWS basics to more advanced services. Start with the comprehensive course at the beginning and proceed step by step for the best approach. Complete each chapter/module to ensure that you're familiar with many of the foundational services covering Compute, Databases, Storage, Networking and Security.

While moving along the learning journey, hands- on labs are provided at the same time to apply the knowledge you have gained from the course. You need to take detailed notes and read Cheatsheets very often. Do not miss the hands-on labs. Create a free tier AWS account and perform hands-on in advance to learn about any topic. The easiest way to understand AWS is to practically use AWS. It helps to internalize the content of the lectures.

Although this book is not enough for a beginner to pass the exam, however, it is still an excellent resource that will be the foundation of your exam preparation. It is necessary to complete AWS training to pass the exam and get a job. Therefore, we will be publishing the AWS Solutions Architect Associate detailed course in October 2020. If you are interested, go and check it out in the same series (Road to AWS) on Amazon KDP.

Deep Dive On Specific AWS Services

There are a few services that will often appear in the exam. Furthermore, the user needs a strong understanding of them to pass. The easiest way to understand these services is to use them practically. The good thing about this prep book is that it covers all these important topics, so you do not need to go anywhere for the theory. Thus, the following areas need deep understanding:

AMAZON VPC

- The exam is very tough regarding VPC questions. Learn to build one from memory.
- Acknowledge the differences between using a VPN and using a Direct Connect.

AMAZON S3

Learn in detail about

- Bucket policies and ACLs
- S3 encryption options
- Storage types
- Lifecycle policies

AMAZON RDS

Learn RDS with a focus on areas that come up often, which include:

- Multi-AZ Replication
- Read Replicas

- Understand snapshots

AMAZON KINESIS

Be sure that you understand AWS Analytics and data warehousing tools at a high level:

- Kinesis: Streams vs Firehose vs Analytics
- Kinesis vs Redshift vs EMR (you will get scenario- based questions with these topics)
- Athena's differences from Kinesis

AMAZON LAMBDA AND SERVERLESS COMPONENTS

Learn Lambda but focus on:

- Limits/timeouts
- Events and what services are generally used with Lambda triggers.

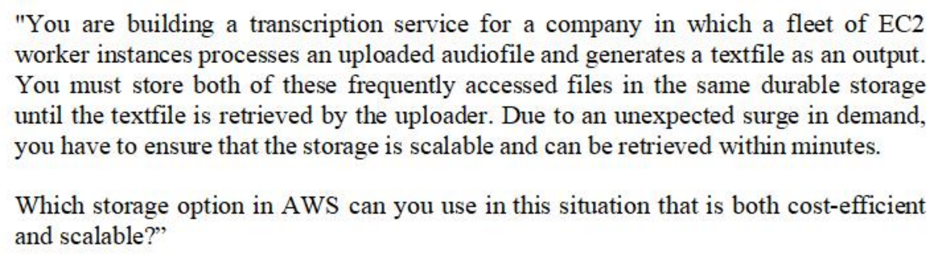
Serverless is likely to appear more and more on AWS exams over time, so you need to consider it as well.

Read The Aws Whitepapers & Faqs

Whitepapers are valuable information for all the services that Amazon publishes on its website. While studying for AWS certifications, it is essential to use some of the highly suggested whitepapers to read before writing the test. Try reading the white paper for the five pillars of the well-architected framework along with FAQs from AWS. There are also several scenario-based questions about the kind of storage. You must know the differences between them and when to use them (like EFS vs EBS, S3, and DynamoDB vs RDS). Test yourself by asking yourself questions.

Learn The Test-Taking Strategies For This Exam

Read the kicker first when dealing with a scenario-based question. Scenario-based questions often start with a detailed paragraph; followed by a "kicker" that asks you a question. Read the kicker then read the scenario. Consider the following use case.



"You are building a transcription service for a company in which a fleet of EC2 worker instances processes an uploaded audiofile and generates a textfile as an output. You must store both of these frequently accessed files in the same durable storage until the textfile is retrieved by the uploader. Due to an unexpected surge in demand, you have to ensure that the storage is scalable and can be retrieved within minutes.

Which storage option in AWS can you use in this situation that is both cost-efficient and scalable?"

The kicker is the last line in this use case. Pay attention to their exact demand. They often present multiple solutions that would work. However, you need to identify the most accurate one (in this case cost efficiency and scalability is the answer).

Use the process of elimination for every question:

Amazon updates the SAA-CO3 exams to include the AWS services introduced in the last few years. Questions are almost going to have one or two choices that are not the right answer. If you can get rid of these hurdles, you can improve your chances of choosing the right one. A very useful tip is to look for the obvious tricks and solve them in your mind. Although this might seem easy, in most cases, they will try to confuse you. Thus, it is essential to use it in the AWS SAA exam. The good news is that most of the updated topics have been covered in this book. So make sure to prepare for those questions.

Read questions twice:

Most of the exam questions are scenario-based with two or more answers. However, they are made to be tricky. Read the questions carefully and scan for keywords such as "cost-effective," or "reproducible data". Do not just look and instantly answer. A slight variation in the wording easily changes the correct answer, and you can select the wrong choice if you go too fast. First, eliminate the incorrect answers and then focus on the possible answers. The answer often sounds almost right, but looking at it word by word, you will find a word that invalidates the answer.

Use the "Mark for Review" feature:

The real exam allows it to mark answers for review and then go back at the end to recheck them. Select the best choice when a question is tough, and then mark it for review at the end based on the remaining time.

Use questions to answer other questions on the real exam:

It is one of the main reasons to use the feature of "Mark for Review." In many cases, one question can give you an answer to another question. You will note this a few times in your test, and you can therefore reverse this question and change the previous answer. That is why one would want to "mark for review" any doubtful question.

Look for certain keywords:

Questions in the exam may confuse you with the data provided as they are lengthy. Therefore, it is quite important to look for specific keywords and relate them with suitable answers. Here are some keywords and possible solutions:

Highly Available= It uses Multiple Availability zones.

Fault Tolerant= Use Multiple Availability zones and a complete replica of the application environment for failover. The application failure in one AZ should automatically lead to a recovery in a different AZ or region.

Real Time Data= It is concerned with real-time data processing such as Kinesis service.

Disaster= It means a service requires failover to a different region.

Long Term Storage= Includes Glacier or Deep Glacier storage.

Managed Service= It includes S3 for data storage, Lambda for computing, Aurora for RDS and DynamoDB for NoSQL database.

Use AWS Network= It relates to the communication between VPC instances and AWS services, and use VPC gateways and endpoints to connect to these services.

Take Practice Exams

You have probably heard the phrase “Practice makes a man perfect”. Frequently or repeatedly doing something makes one become an expert or skilful at it. Thus, you can practice the knowledge or skills of the real exam-based questions in the practice exam. It provides simulated questions that are very similar to the actual exam. Each question carries detailed explanations that will help to gain a better understanding of the AWS services. It not just explains what the correct answer is, but also explains why other answers are wrong. It helps to recognize the difference between similar services. Another thing is that the exam pattern is close to the actual exam. It simulates very well because the questions are like the actual exam.

Note that the practise exams are tough as they represent the most challenging questions you would face in the exam. As a rule of thumb, passing these practise exams, one must be able to clear the real exam with extra room. Fail fast in practice exam to avoid failure in real exam. For instance, attempt the mock exam multiple times to practice attempting the exam. Make sure to score above 75% baseline (the passing mark is 73%).

Attempting all the practice questions at the end of this book prepares you well for the real exam. Set of practice exams are also available on the AWS website. Purchase them and work through them.

Cheatsheets

Cheat sheets are useful as they provide a quick summary of key AWS services that are necessary to clear the AWS certification exam. They are short bullet points which are concise and easy to remember. It saves us from going through the whole AWS documentation. Comparison of similar services is the most useful part of cheat sheets. Often, there are two similar answers in the exam that seems to resolve the problem in the given scenario. However, a minor difference between the services points to the right answer, but not the other.

Get to learn the comparison table/chart for all of the services that will allow you to clear the exam.

Hands-On Experience

All certifications are worthy as they are the add- ons for a professional career. However, the only certification is not enough. One must have hands-on experience to get a job. It is good to have practical experience after learning related concepts. It helps to strengthen the knowledge you learnt by getting your hands dirty. The Hands-on labs provide immediate access to AWS services that help put theory into practice. Every organization demands practical experience. You have to gain practical experience if you want to work with different technology and are passionate to work with AWS.

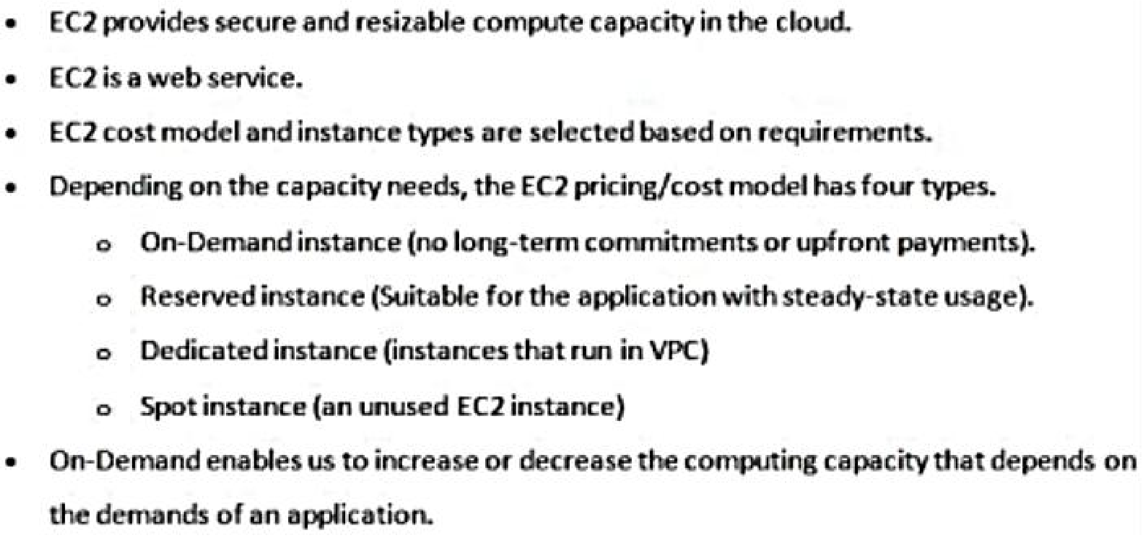
Since AWS offers cloud services from development to deployment, you can create and launch resources into AWS Platform. You can also build applications using AWS account. You can always use AWS free- tier account for experiments. It allows an individual to create and learn at the same time. You can exercise important concepts on AWS free account (it is free as long as you know what you are doing). For example, you can use your domain using Route 53 and launch some personal applications in EC2 instance to increase your understanding. While building an application with basic AWS services, you are getting hands-on experience along with studying. Both are crucial for a better understanding of AWS and boosted confidence with the technology.

Conclusion

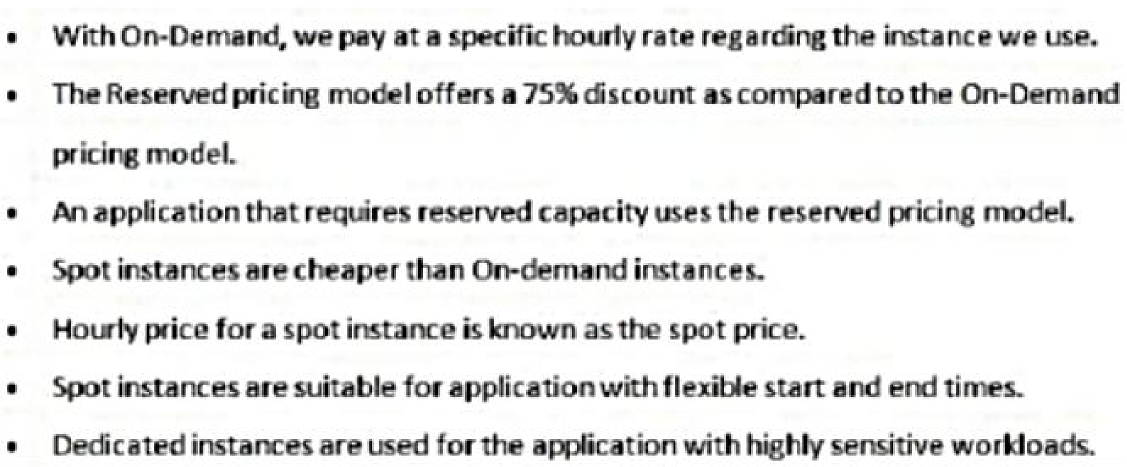
Like any other exam, studying ensures success. Same is the case for AWS SAA exam. Although it can take up some of your time, the benefits are great. Besides validating your technical skills, this certification will improve the expertise making you more appealing to a potential recruiter, which will continue to enhance your career in the long term.

Cheat Sheets

ELASTIC CLOUD COMPUTE (EC2)

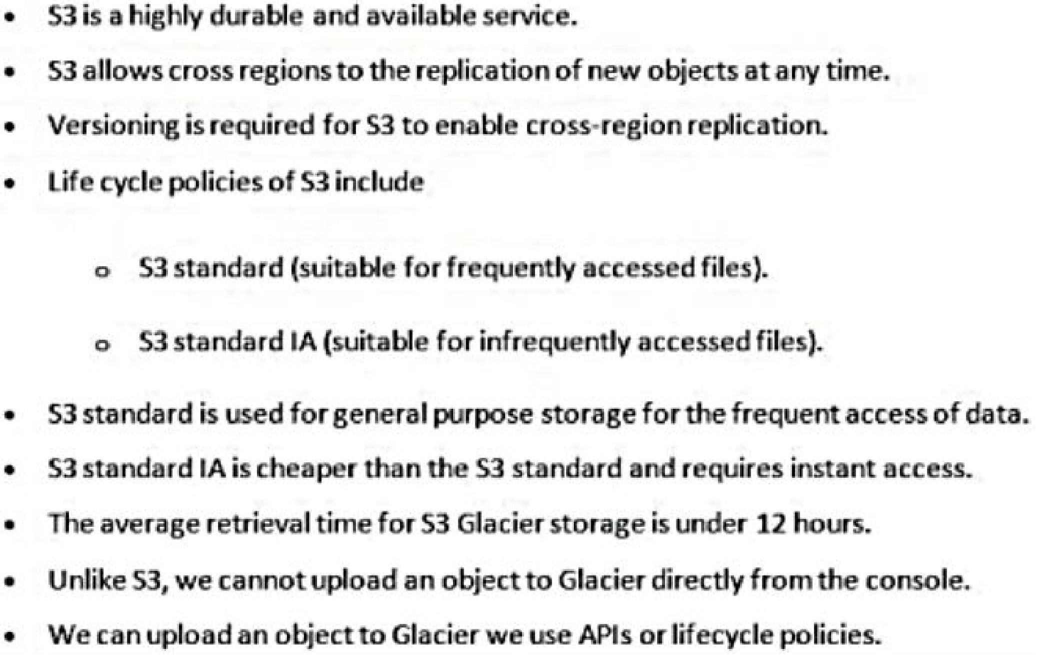


- EC2 provides secure and resizable compute capacity in the cloud.
- EC2 is a web service.
- EC2 cost model and instance types are selected based on requirements.
- Depending on the capacity needs, the EC2 pricing/cost model has four types.
  - On-Demand instance (no long-term commitments or upfront payments).
  - Reserved instance (Suitable for the application with steady-state usage).
  - Dedicated instance (instances that run in VPC)
  - Spot instance (an unused EC2 instance)
- On-Demand enables us to increase or decrease the computing capacity that depends on the demands of an application.



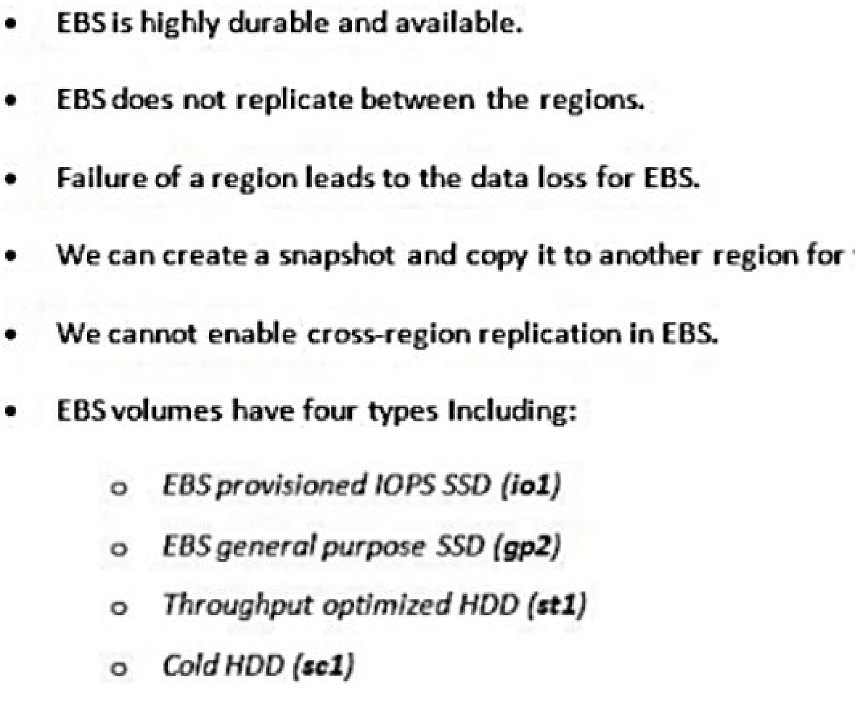
- With On-Demand, we pay at a specific hourly rate regarding the instance we use.
- The Reserved pricing model offers a 75% discount as compared to the On-Demand pricing model.
- An application that requires reserved capacity uses the reserved pricing model.
- Spot instances are cheaper than On-demand instances.
- Hourly price for a spot instance is known as the spot price.
- Spot instances are suitable for application with flexible start and end times.
- Dedicated instances are used for the application with highly sensitive workloads.

AMAZON SIMPLE STORAGE SERVICE (S3)

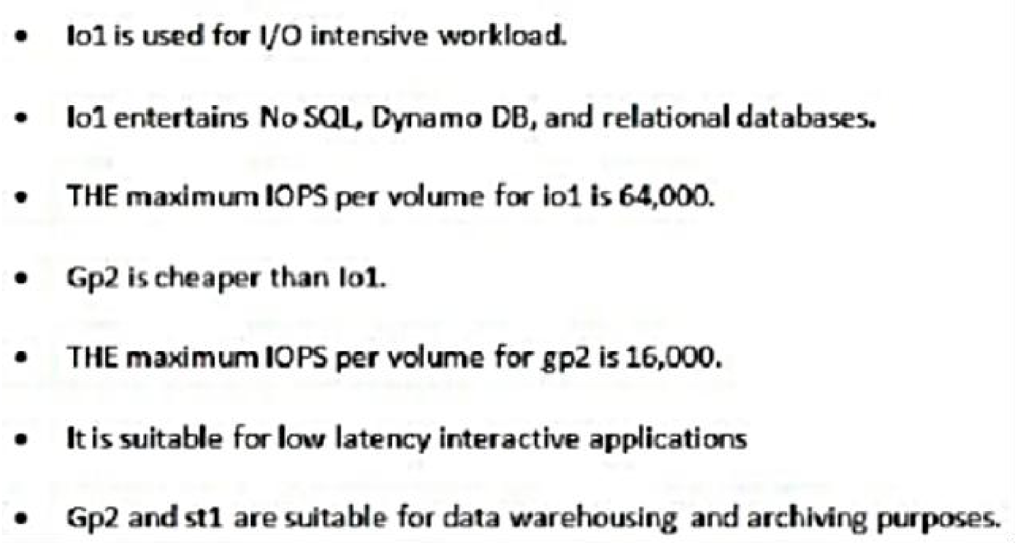


- S3 is a highly durable and available service.
- S3 allows cross regions to the replication of new objects at any time.
- Versioning is required for S3 to enable cross-region replication.
- Life cycle policies of S3 include
    - S3 standard (suitable for frequently accessed files).
    - S3 standard IA (suitable for infrequently accessed files).
- S3 standard is used for general purpose storage for the frequent access of data.
- S3 standard IA is cheaper than the S3 standard and requires instant access.
- The average retrieval time for S3 Glacier storage is under 12 hours.
- Unlike S3, we cannot upload an object to Glacier directly from the console.
- We can upload an object to Glacier we use APIs or lifecycle policies.

ELASTIC BLOCK STORE (EBS)

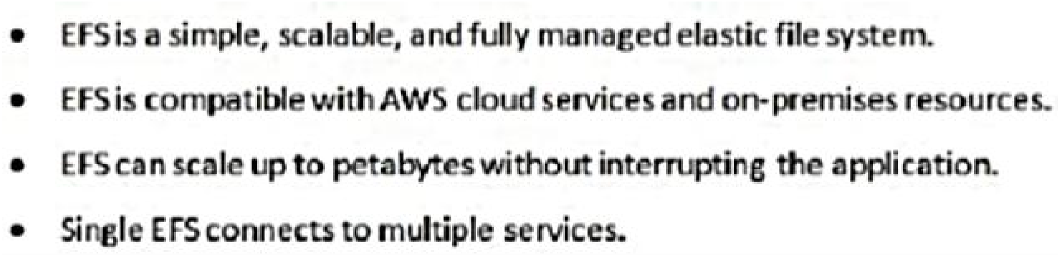


- EBS is highly durable and available.
- EBS does not replicate between the regions.
- Failure of a region leads to the data loss for EBS.
- We can create a snapshot and copy it to another region for
- We cannot enable cross-region replication in EBS.
- EBS volumes have four types Including:
    - *EBS provisioned IOPS SSD (**io1**)*
    - *EBS general purpose SSD (**gp2**)*
    - *Throughput optimized HDD (**st1**)*
    - *Cold HDD (**sc1**)*



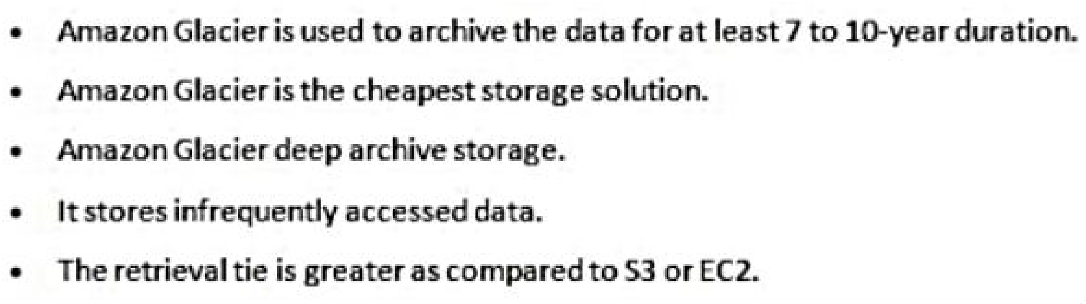
- Io1 is used for I/O intensive workload.
- Io1 entertains No SQL, Dynamo DB, and relational databases.
- THE maximum IOPS per volume for io1 is 64,000.
- Gp2 is cheaper than Io1.
- THE maximum IOPS per volume for gp2 is 16,000.
- It is suitable for low latency interactive applications
- Gp2 and st1 are suitable for data warehousing and archiving purposes.

ELASTIC FILE SYSTEM



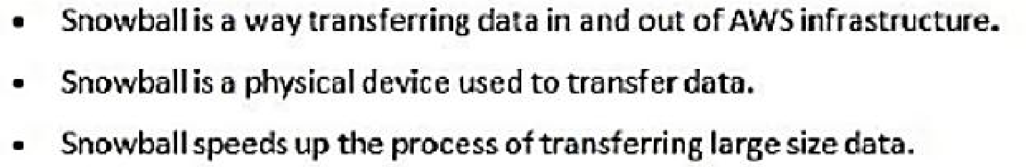
- EFS is a simple, scalable, and fully managed elastic file system.
- EFS is compatible with AWS cloud services and on-premises resources.
- EFS can scale up to petabytes without interrupting the application.
- Single EFS connects to multiple services.

Amazon GLACIER



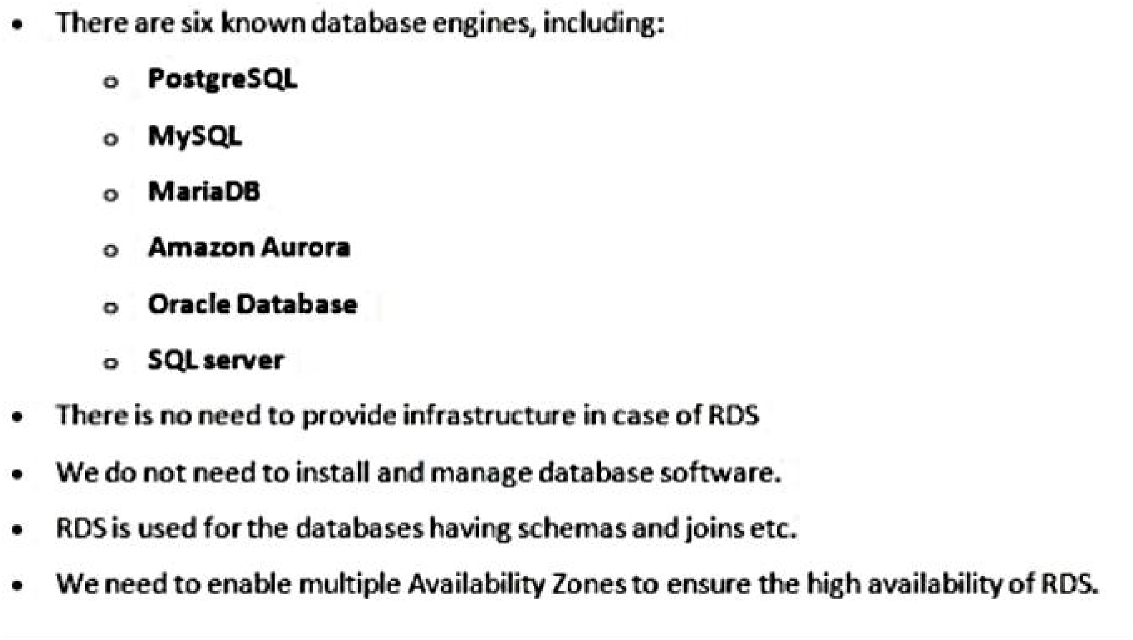
- Amazon Glacier is used to archive the data for at least 7 to 10-year duration.
- Amazon Glacier is the cheapest storage solution.
- Amazon Glacier deep archive storage.
- It stores infrequently accessed data.
- The retrieval tie is greater as compared to S3 or EC2.

SNOWBALL



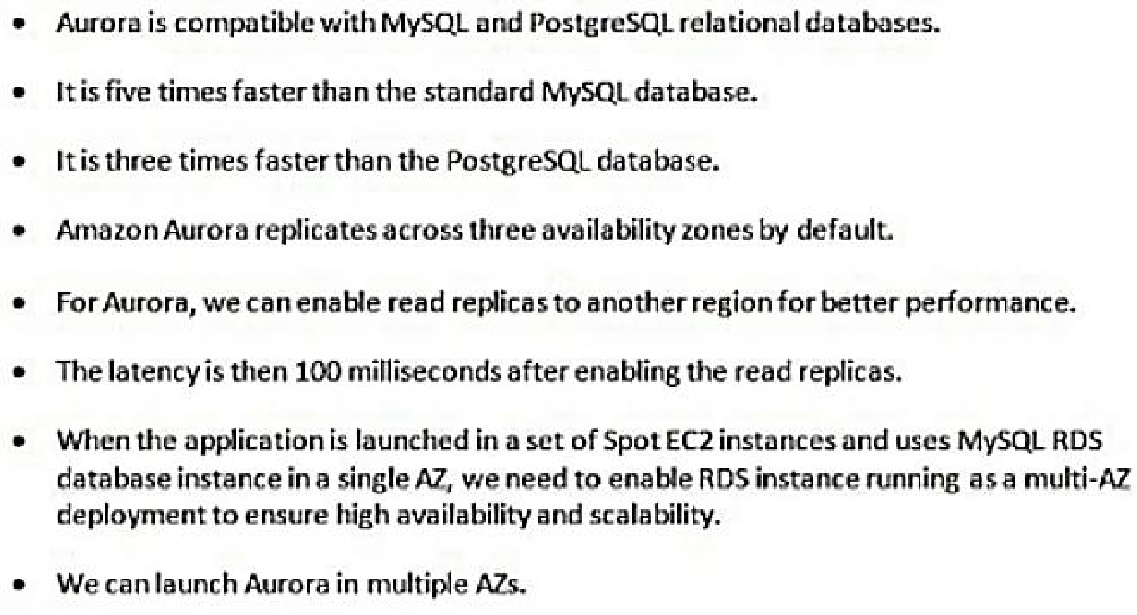
- Snowball is a way transferring data in and out of AWS infrastructure.
- Snowball is a physical device used to transfer data.
- Snowball speeds up the process of transferring large size data.

RELATIONAL DATABASE SERVICE (RDS)



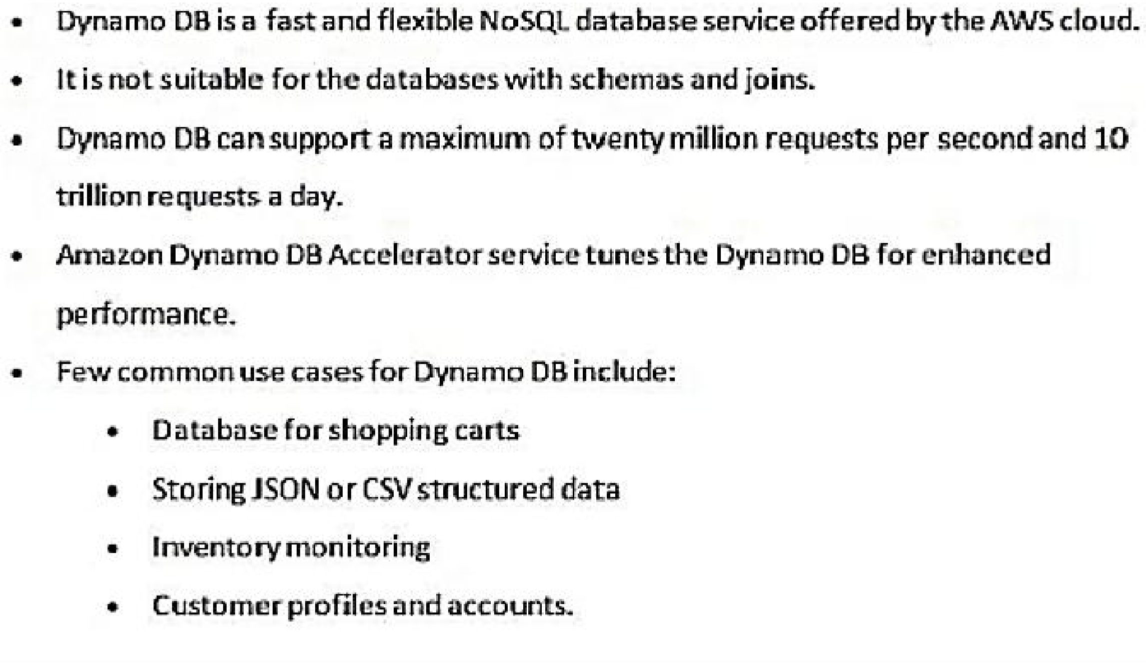
- There are six known database engines, including:
  - **PostgreSQL**
  - **MySQL**
  - **MariaDB**
  - **Amazon Aurora**
  - **Oracle Database**
  - **SQL server**
- There is no need to provide infrastructure in case of RDS
- We do not need to install and manage database software.
- RDS is used for the databases having schemas and joins etc.
- We need to enable multiple Availability Zones to ensure the high availability of RDS.

AMAZON AURORA



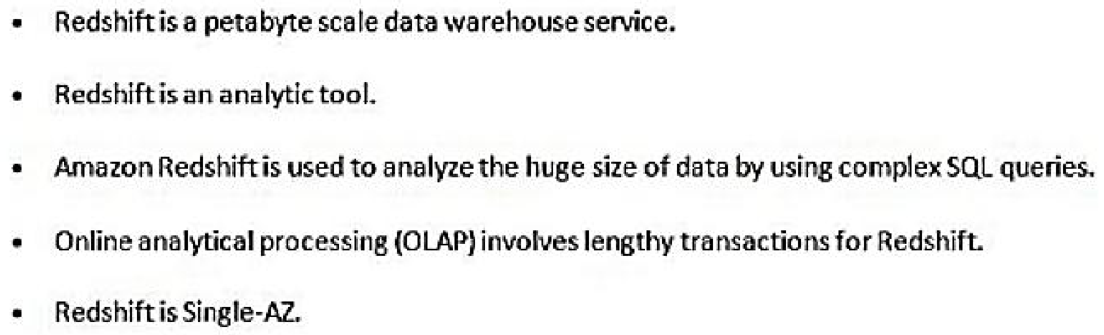
- Aurora is compatible with MySQL and PostgreSQL relational databases.
- It is five times faster than the standard MySQL database.
- It is three times faster than the PostgreSQL database.
- Amazon Aurora replicates across three availability zones by default.
- For Aurora, we can enable read replicas to another region for better performance.
- The latency is then 100 milliseconds after enabling the read replicas.
- When the application is launched in a set of Spot EC2 instances and uses MySQL RDS database instance in a single AZ, we need to enable RDS instance running as a multi-AZ deployment to ensure high availability and scalability.
- We can launch Aurora in multiple AZs.

DYNAMO DB



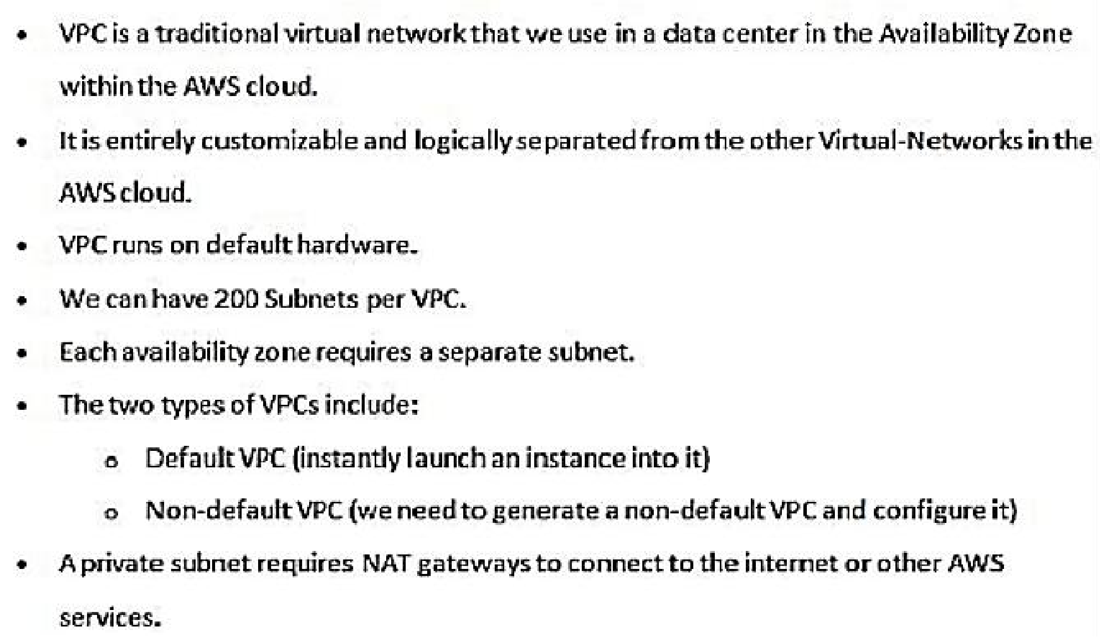
- Dynamo DB is a fast and flexible NoSQL database service offered by the AWS cloud.
- It is not suitable for the databases with schemas and joins.
- Dynamo DB can support a maximum of twenty million requests per second and 10 trillion requests a day.
- Amazon Dynamo DB Accelerator service tunes the Dynamo DB for enhanced performance.
- Few common use cases for Dynamo DB include:
  - Database for shopping carts
  - Storing JSON or CSV structured data
  - Inventory monitoring
  - Customer profiles and accounts.

REDSHIFT

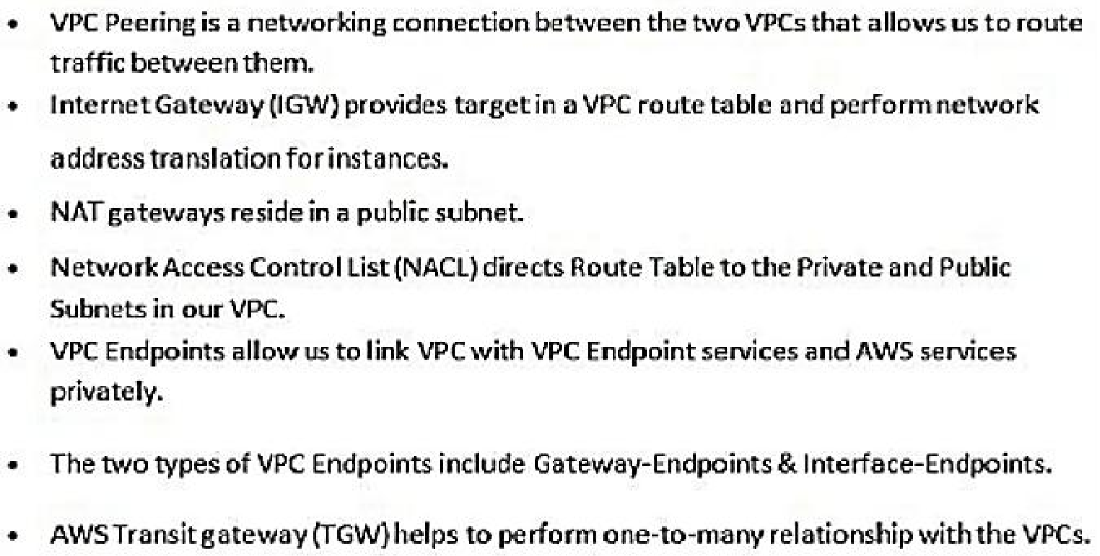


- Redshift is a petabyte scale data warehouse service.
- Redshift is an analytic tool.
- Amazon Redshift is used to analyze the huge size of data by using complex SQL queries.
- Online analytical processing (OLAP) involves lengthy transactions for Redshift.
- Redshift is Single-AZ.

VPC



- VPC is a traditional virtual network that we use in a data center in the Availability Zone within the AWS cloud.
- It is entirely customizable and logically separated from the other Virtual-Networks in the AWS cloud.
- VPC runs on default hardware.
- We can have 200 Subnets per VPC.
- Each availability zone requires a separate subnet.
- The two types of VPCs include:
    - Default VPC (instantly launch an instance into it)
    - Non-default VPC (we need to generate a non-default VPC and configure it)
- A private subnet requires NAT gateways to connect to the internet or other AWS services.



- VPC Peering is a networking connection between the two VPCs that allows us to route traffic between them.
- Internet Gateway (IGW) provides target in a VPC route table and perform network address translation for instances.
- NAT gateways reside in a public subnet.
- Network Access Control List (NACL) directs Route Table to the Private and Public Subnets in our VPC.
- VPC Endpoints allow us to link VPC with VPC Endpoint services and AWS services privately.
- The two types of VPC Endpoints include Gateway-Endpoints & Interface-Endpoints.
- AWS Transit gateway (TGW) helps to perform one-to-many relationship with the VPCs.

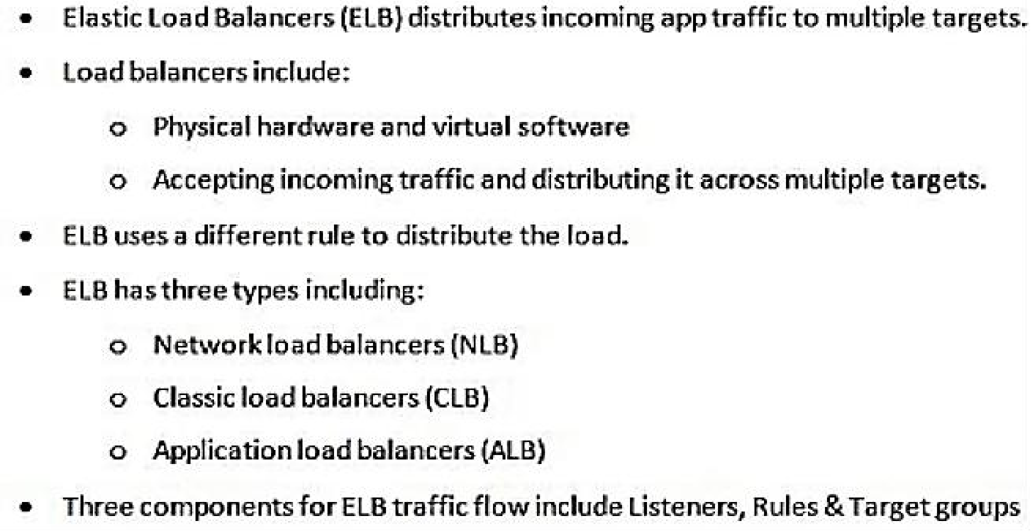
ROUTE 53

- Route 53 is a domain name system
- It registers and manages domains
- Route 53 Implement complicated traffic flows
- Route53 can manage the name servers
- We can create the link for Route53 to point certain resources by creating Record Sets.
- Continuously monitor the records through health checks and resolve VPCs outside of AWS.

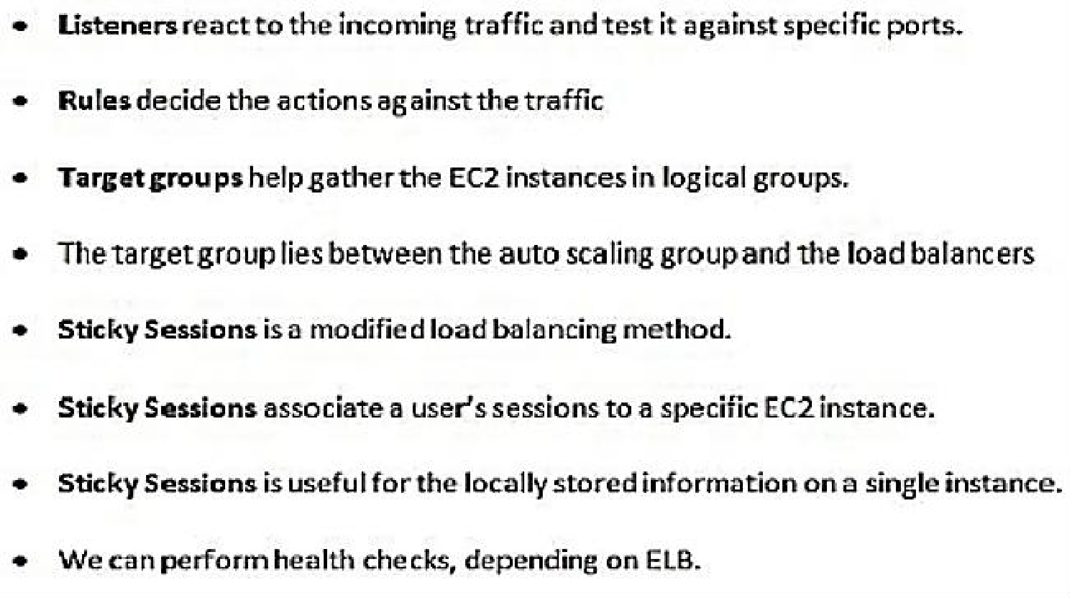
PRIVATE LINK

- Private Link establishes a private connection between AWS VPCs, an on-premises application, and AWS services.
- Private Link secures the traffic from the VPC environments of the users residing in AWS.

ELASTIC LOAD BALANCER (ELB)

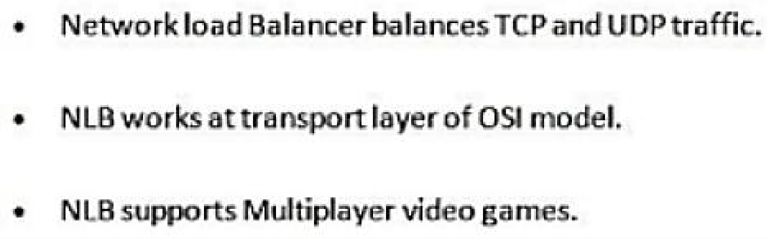


- Elastic Load Balancers (ELB) distributes incoming app traffic to multiple targets.
- Load balancers include:
  - Physical hardware and virtual software
  - Accepting incoming traffic and distributing it across multiple targets.
- ELB uses a different rule to distribute the load.
- ELB has three types including:
  - Network load balancers (NLB)
  - Classic load balancers (CLB)
  - Application load balancers (ALB)
- Three components for ELB traffic flow include Listeners, Rules & Target groups



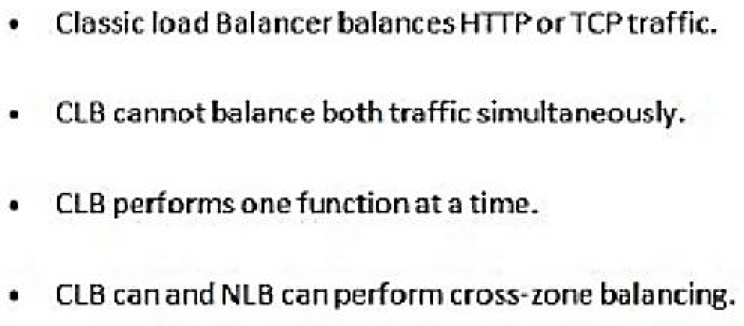
- **Listeners** react to the incoming traffic and test it against specific ports.
- **Rules** decide the actions against the traffic
- **Target groups** help gather the EC2 instances in logical groups.
- The target group lies between the auto scaling group and the load balancers
- **Sticky Sessions** is a modified load balancing method.
- **Sticky Sessions** associate a user's sessions to a specific EC2 instance.
- **Sticky Sessions** is useful for the locally stored information on a single instance.
- We can perform health checks, depending on ELB.

NETWORK LOAD BALANCER (NLB)



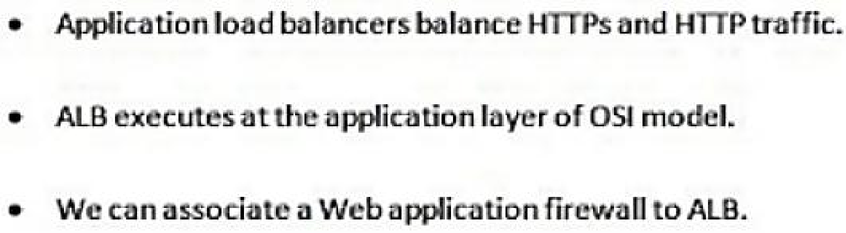
- Network load Balancer balances TCP and UDP traffic.
- NLB works at transport layer of OSI model.
- NLB supports Multiplayer video games.

CLASSIC LOAD BALANCER (CLB)



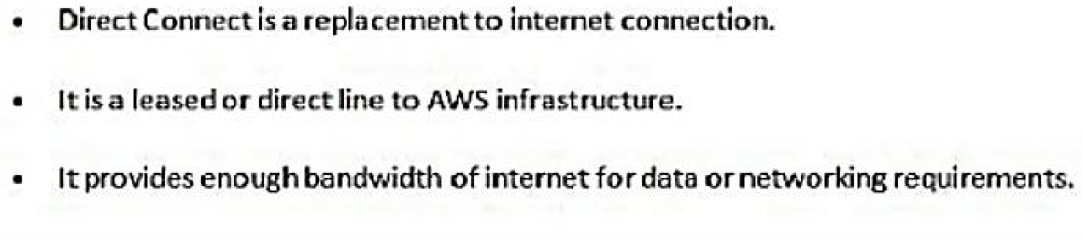
- Classic load Balancer balances HTTP or TCP traffic.
- CLB cannot balance both traffic simultaneously.
- CLB performs one function at a time.
- CLB can and NLB can perform cross-zone balancing.

APPLICATION LOAD BALANCERS



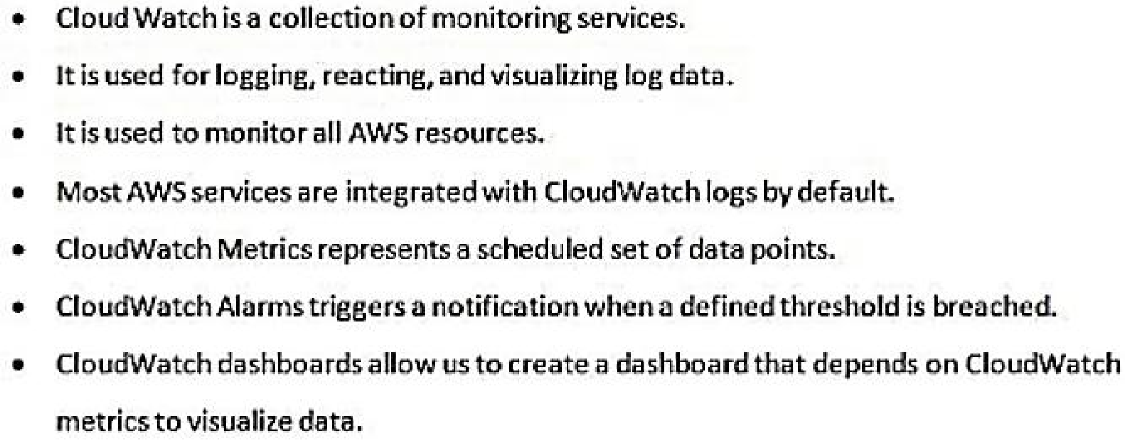
- Application load balancers balance HTTPs and HTTP traffic.
- ALB executes at the application layer of OSI model.
- We can associate a Web application firewall to ALB.

DIRECT CONNECT



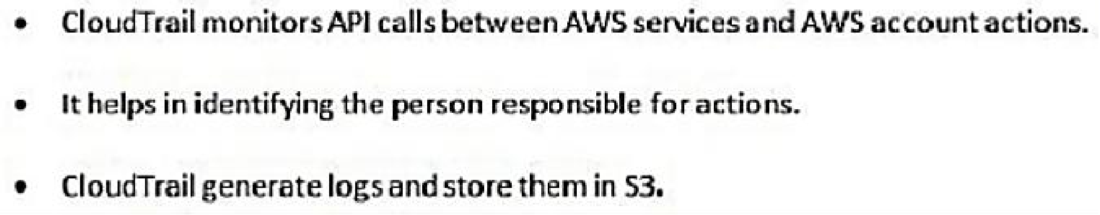
- Direct Connect is a replacement to internet connection.
- It is a leased or direct line to AWS infrastructure.
- It provides enough bandwidth of internet for data or networking requirements.

CLOUDWATCH



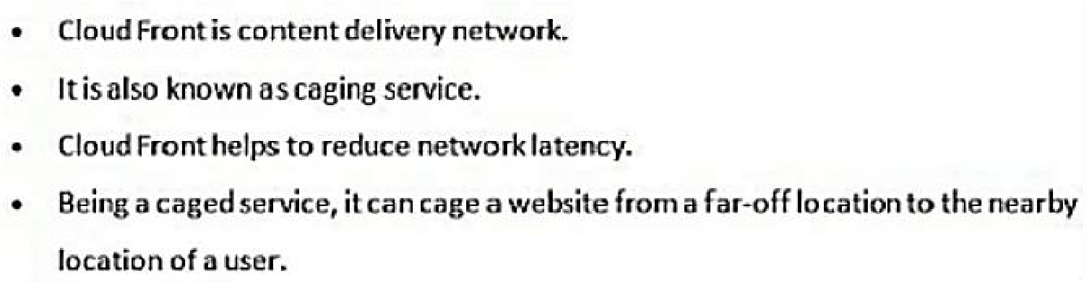
- Cloud Watch is a collection of monitoring services.
- It is used for logging, reacting, and visualizing log data.
- It is used to monitor all AWS resources.
- Most AWS services are integrated with CloudWatch logs by default.
- CloudWatch Metrics represents a scheduled set of data points.
- CloudWatch Alarms triggers a notification when a defined threshold is breached.
- CloudWatch dashboards allow us to create a dashboard that depends on CloudWatch metrics to visualize data.

CLOUDTRAIL



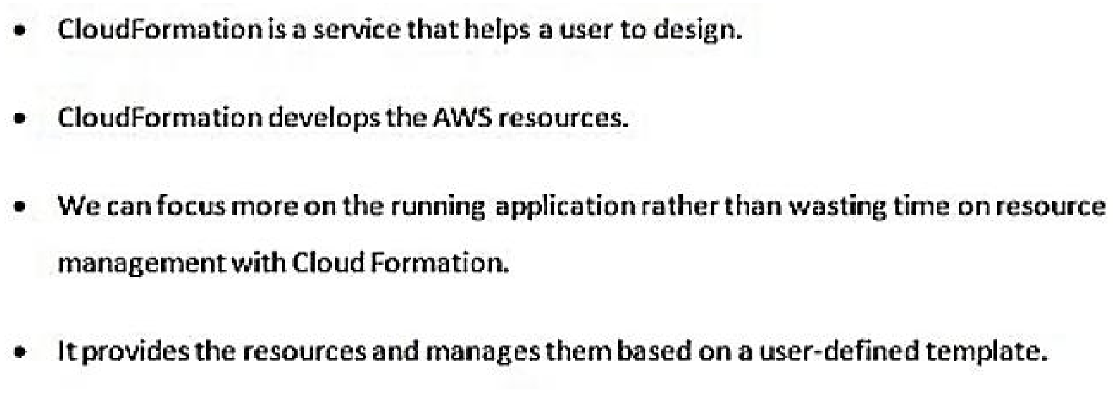
- CloudTrail monitors API calls between AWS services and AWS account actions.
- It helps in identifying the person responsible for actions.
- CloudTrail generate logs and store them in S3.

CLOUDFRONT



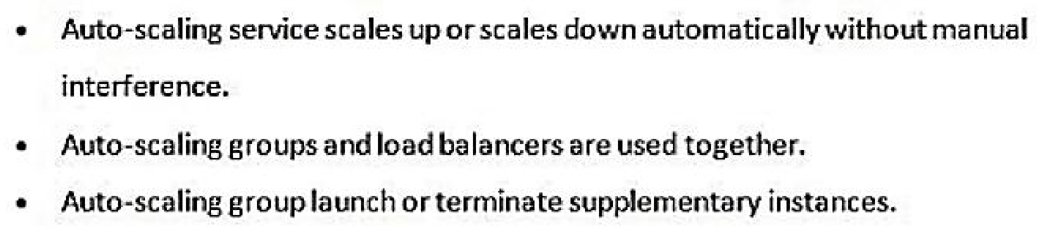
- Cloud Front is content delivery network.
- It is also known as caging service.
- Cloud Front helps to reduce network latency.
- Being a caged service, it can cage a website from a far-off location to the nearby location of a user.

CLOUDFORMATION



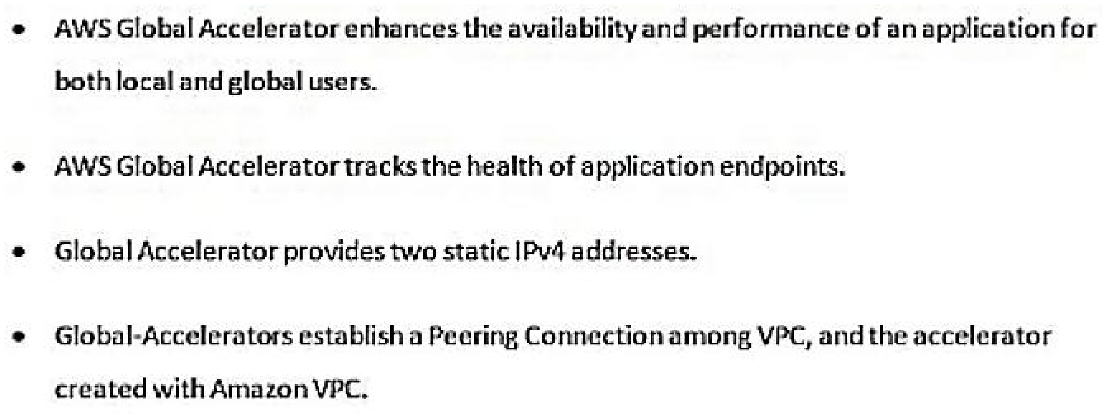
- CloudFormation is a service that helps a user to design.
- CloudFormation develops the AWS resources.
- We can focus more on the running application rather than wasting time on resource management with Cloud Formation.
- It provides the resources and manages them based on a user-defined template.

AUTO-SCALING GROUPS



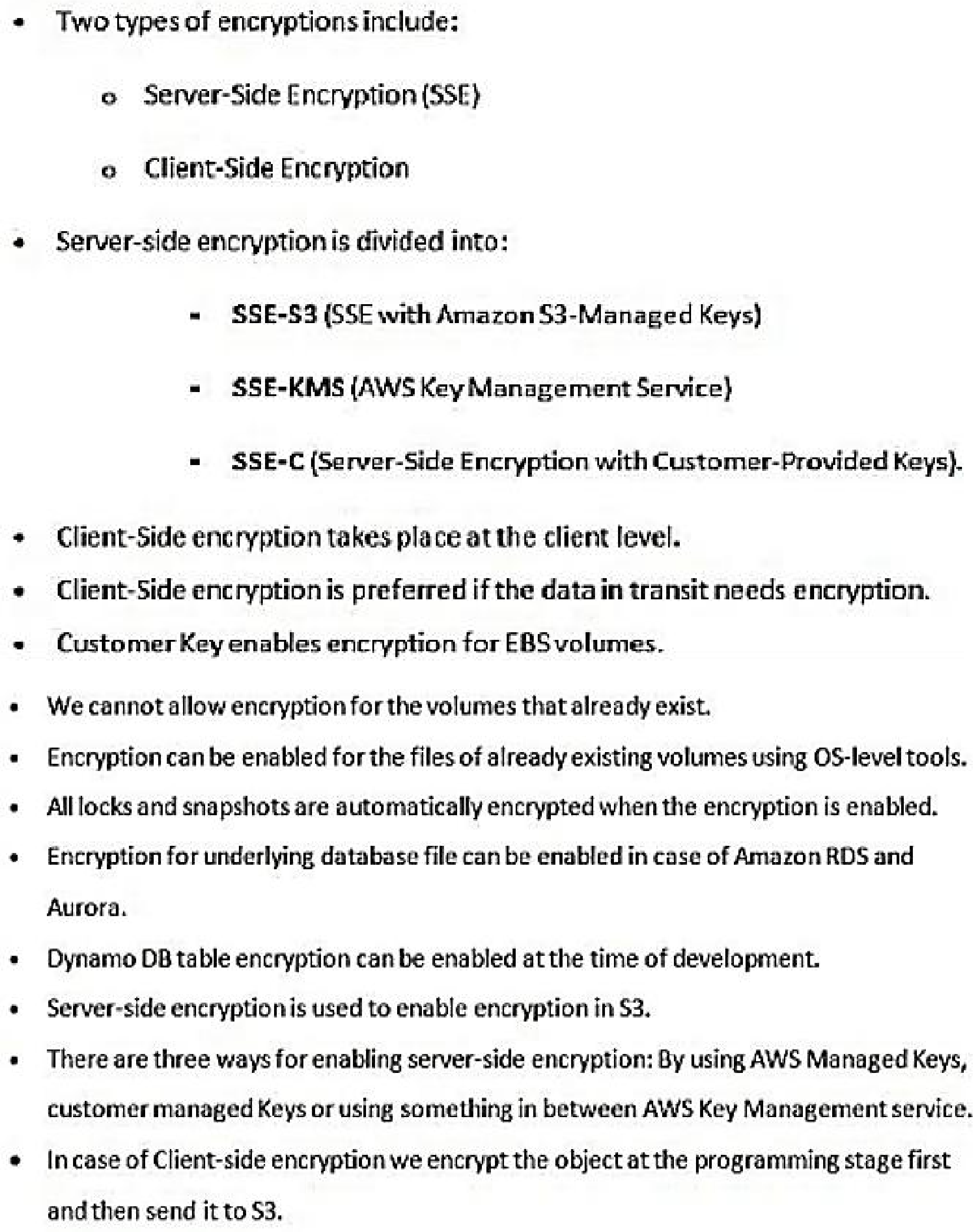
- Auto-scaling service scales up or scales down automatically without manual interference.
- Auto-scaling groups and load balancers are used together.
- Auto-scaling group launch or terminate supplementary instances.

AWS GLOBAL ACCELERATOR



- AWS Global Accelerator enhances the availability and performance of an application for both local and global users.
- AWS Global Accelerator tracks the health of application endpoints.
- Global Accelerator provides two static IPv4 addresses.
- Global-Accelerators establish a Peering Connection among VPC, and the accelerator created with Amazon VPC.

ENCRYPTION

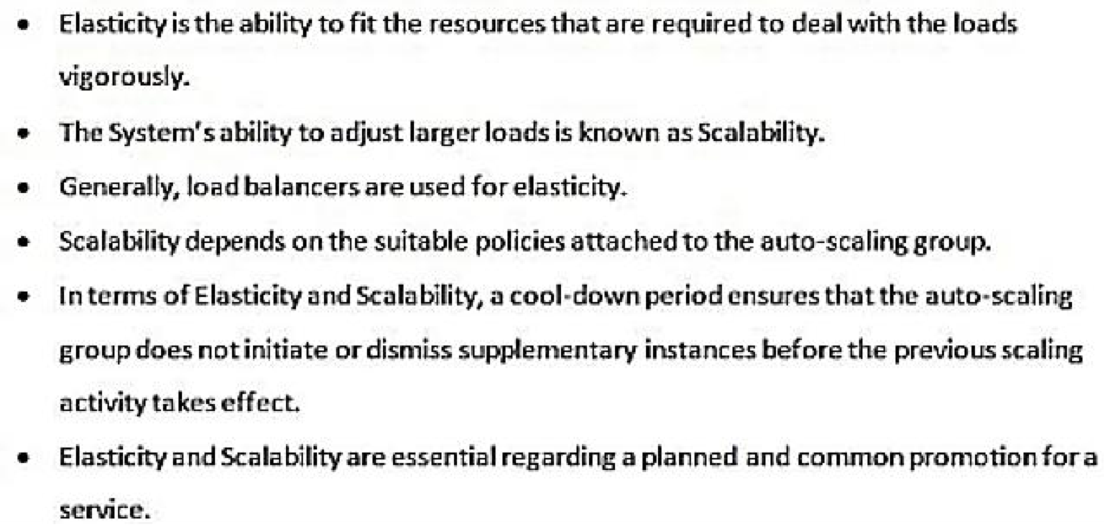


- Two types of encryptions include:
  - Server-Side Encryption (SSE)
  - Client-Side Encryption
- Server-side encryption is divided into:
  - **SSE-S3** (SSE with Amazon S3-Managed Keys)
  - **SSE-KMS** (AWS Key Management Service)
  - **SSE-C** (Server-Side Encryption with Customer-Provided Keys).
- Client-Side encryption takes place at the client level.
- Client-Side encryption is preferred if the data in transit needs encryption.
- Customer Key enables encryption for EBS volumes.
- We cannot allow encryption for the volumes that already exist.
- Encryption can be enabled for the files of already existing volumes using OS-level tools.
- All locks and snapshots are automatically encrypted when the encryption is enabled.
- Encryption for underlying database file can be enabled in case of Amazon RDS and Aurora.
- Dynamo DB table encryption can be enabled at the time of development.
- Server-side encryption is used to enable encryption in S3.
- There are three ways for enabling server-side encryption: By using AWS Managed Keys, customer managed Keys or using something in between AWS Key Management service.
- In case of Client-side encryption we encrypt the object at the programming stage first and then send it to S3.

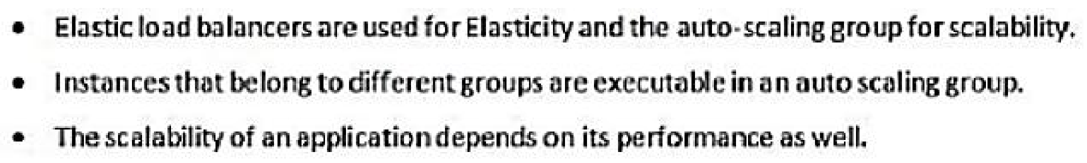
KMS & IAM

- KMS stands for Key Management service.
- KMS create and manage encryptions keys to encrypt the data.
- The encryption keys include:
  - Public Key
  - Private Key
- IAM is Identification and Authentication management tool.
- IAM enable us to provide different permissions.

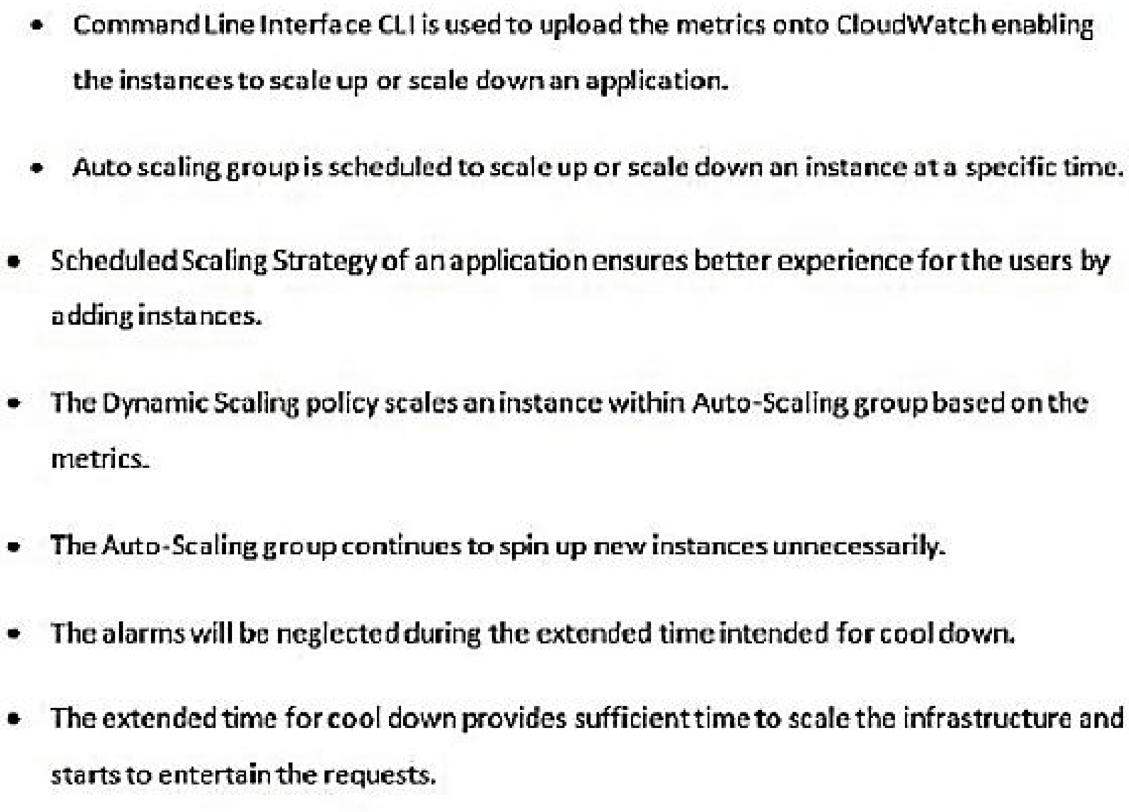
ELASTICITY & SCALABILITY



- Elasticity is the ability to fit the resources that are required to deal with the loads vigorously.
- The System's ability to adjust larger loads is known as Scalability.
- Generally, load balancers are used for elasticity.
- Scalability depends on the suitable policies attached to the auto-scaling group.
- In terms of Elasticity and Scalability, a cool-down period ensures that the auto-scaling group does not initiate or dismiss supplementary instances before the previous scaling activity takes effect.
- Elasticity and Scalability are essential regarding a planned and common promotion for a service.

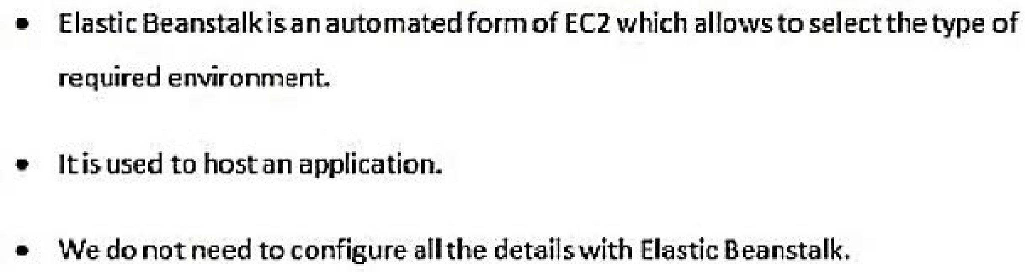


- Elastic load balancers are used for Elasticity and the auto-scaling group for scalability.
- Instances that belong to different groups are executable in an auto scaling group.
- The scalability of an application depends on its performance as well.



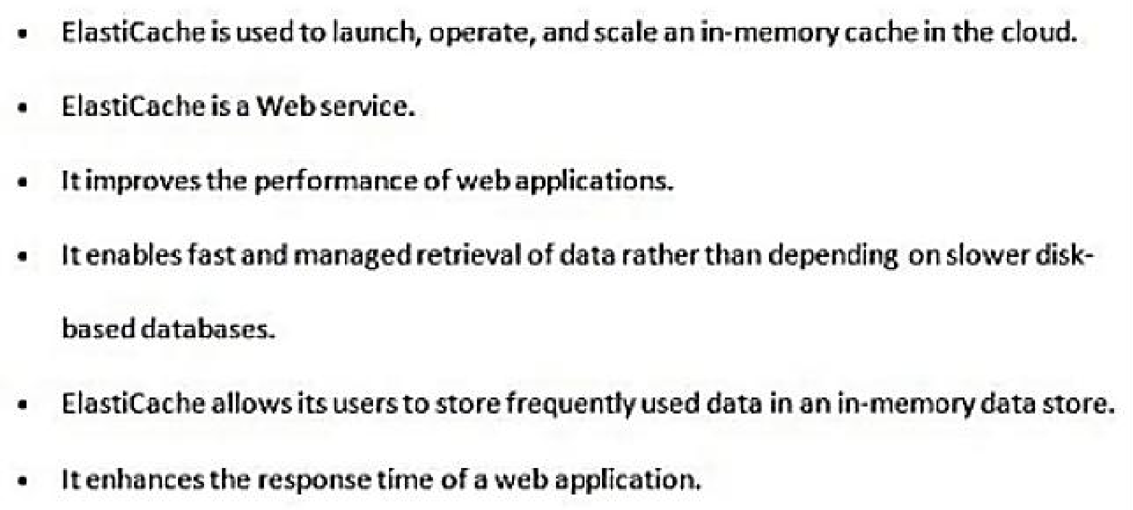
  - Command Line Interface CLI is used to upload the metrics onto CloudWatch enabling the instances to scale up or scale down an application.
  - Auto scaling group is scheduled to scale up or scale down an instance at a specific time.
- Scheduled Scaling Strategy of an application ensures better experience for the users by adding instances.
- The Dynamic Scaling policy scales an instance within Auto-Scaling group based on the metrics.
- The Auto-Scaling group continues to spin up new instances unnecessarily.
- The alarms will be neglected during the extended time intended for cool down.
- The extended time for cool down provides sufficient time to scale the infrastructure and starts to entertain the requests.

ELASTIC BEANSTALK



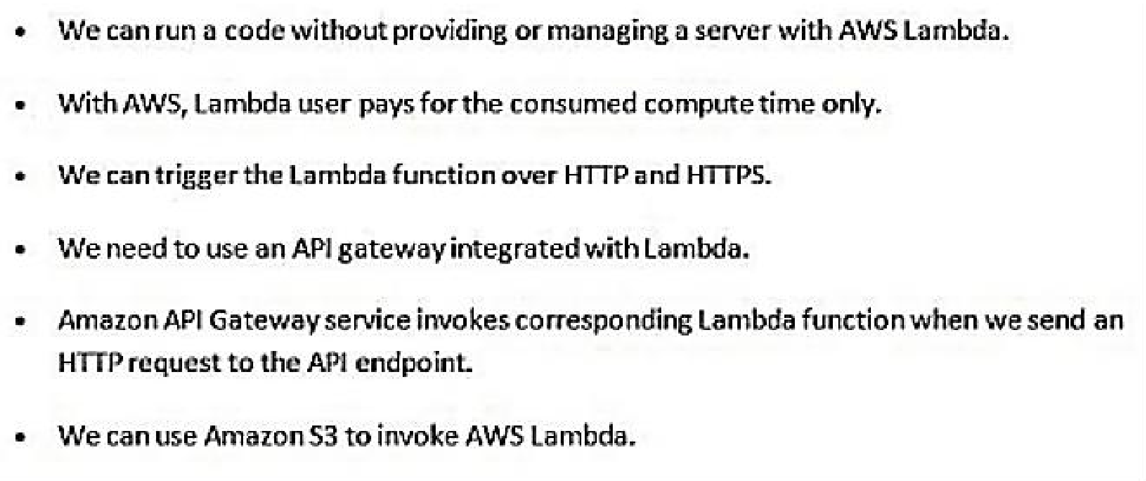
- Elastic Beanstalk is an automated form of EC2 which allows to select the type of required environment.
- It is used to host an application.
- We do not need to configure all the details with Elastic Beanstalk.

AMAZON ELASTICACHE



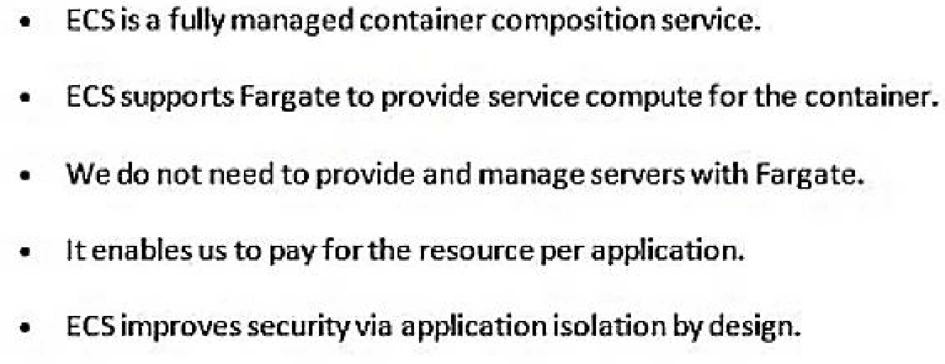
- ElastiCache is used to launch, operate, and scale an in-memory cache in the cloud.
- ElastiCache is a Web service.
- It improves the performance of web applications.
- It enables fast and managed retrieval of data rather than depending on slower disk-based databases.
- ElastiCache allows its users to store frequently used data in an in-memory data store.
- It enhances the response time of a web application.

LAMBDA



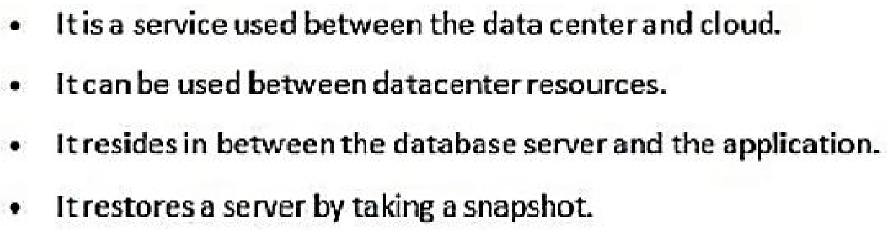
- We can run a code without providing or managing a server with AWS Lambda.
- With AWS, Lambda user pays for the consumed compute time only.
- We can trigger the Lambda function over HTTP and HTTPS.
- We need to use an API gateway integrated with Lambda.
- Amazon API Gateway service invokes corresponding Lambda function when we send an HTTP request to the API endpoint.
- We can use Amazon S3 to invoke AWS Lambda.

ELASTIC CONTAINER SERVICE (ECS)



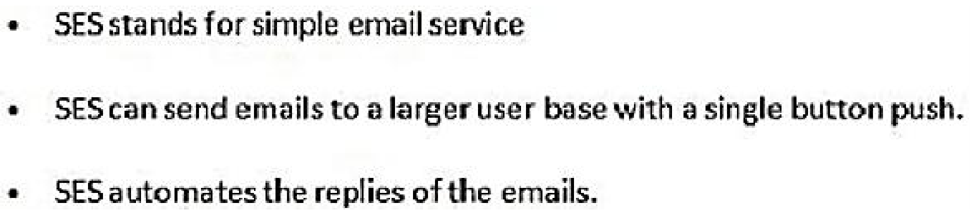
- ECS is a fully managed container composition service.
- ECS supports Fargate to provide service compute for the container.
- We do not need to provide and manage servers with Fargate.
- It enables us to pay for the resource per application.
- ECS improves security via application isolation by design.

STORAGE GATEWAY



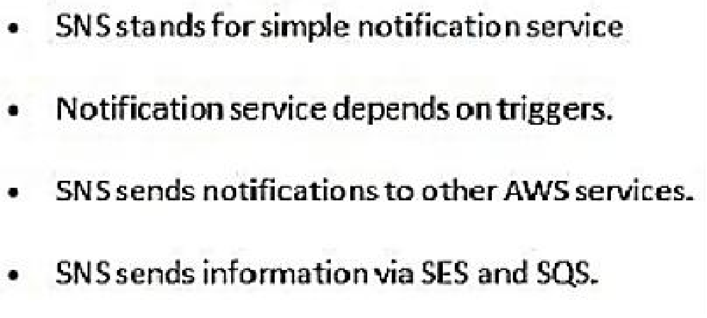
- It is a service used between the data center and cloud.
- It can be used between datacenter resources.
- It resides in between the database server and the application.
- It restores a server by taking a snapshot.

SES



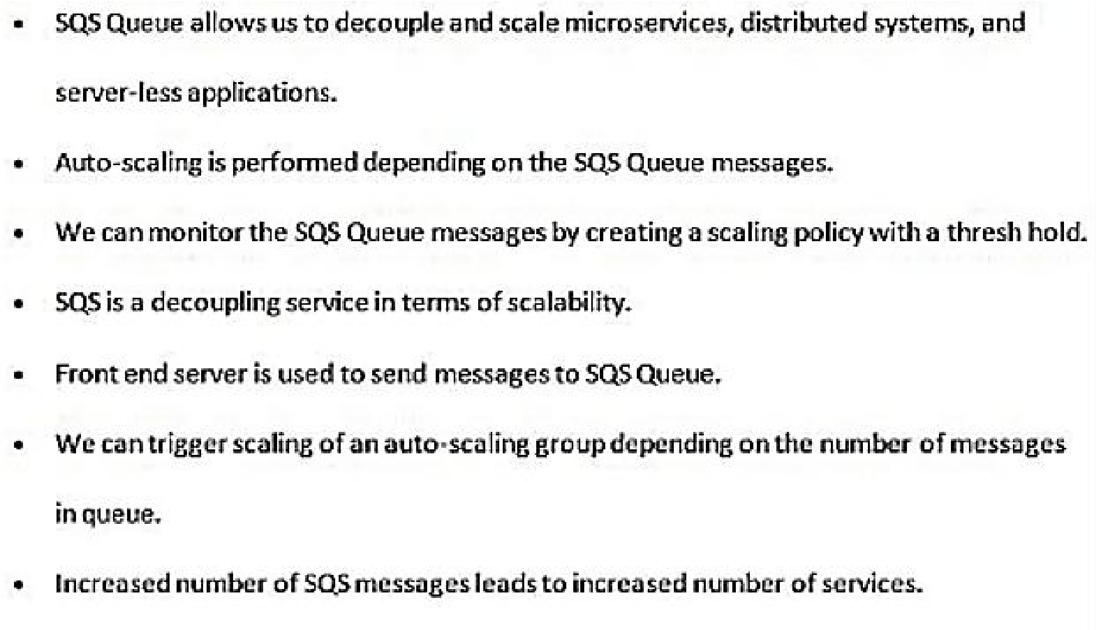
- SES stands for simple email service
- SES can send emails to a larger user base with a single button push.
- SES automates the replies of the emails.

SNS



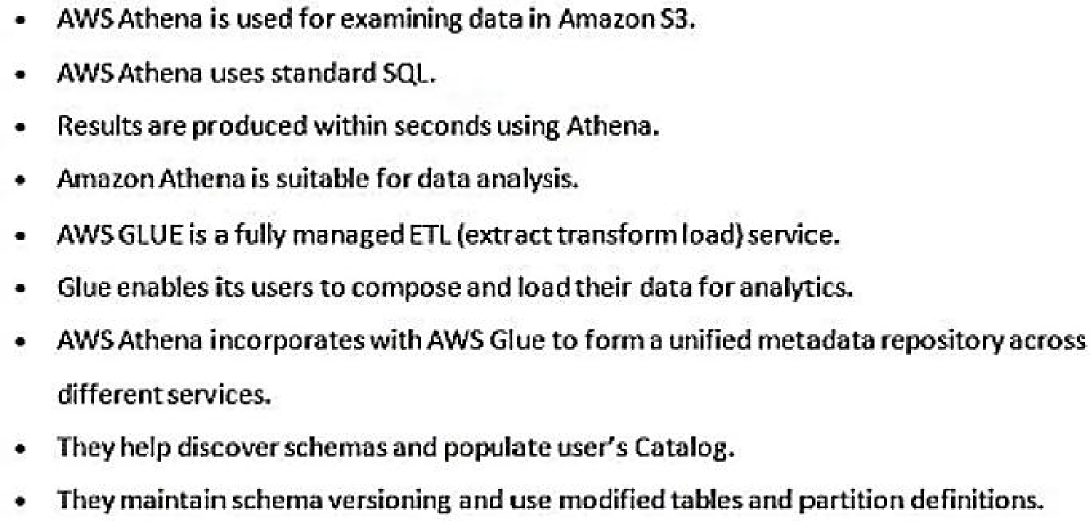
- SNS stands for simple notification service
- Notification service depends on triggers.
- SNS sends notifications to other AWS services.
- SNS sends information via SES and SQS.

SQS QUEUE



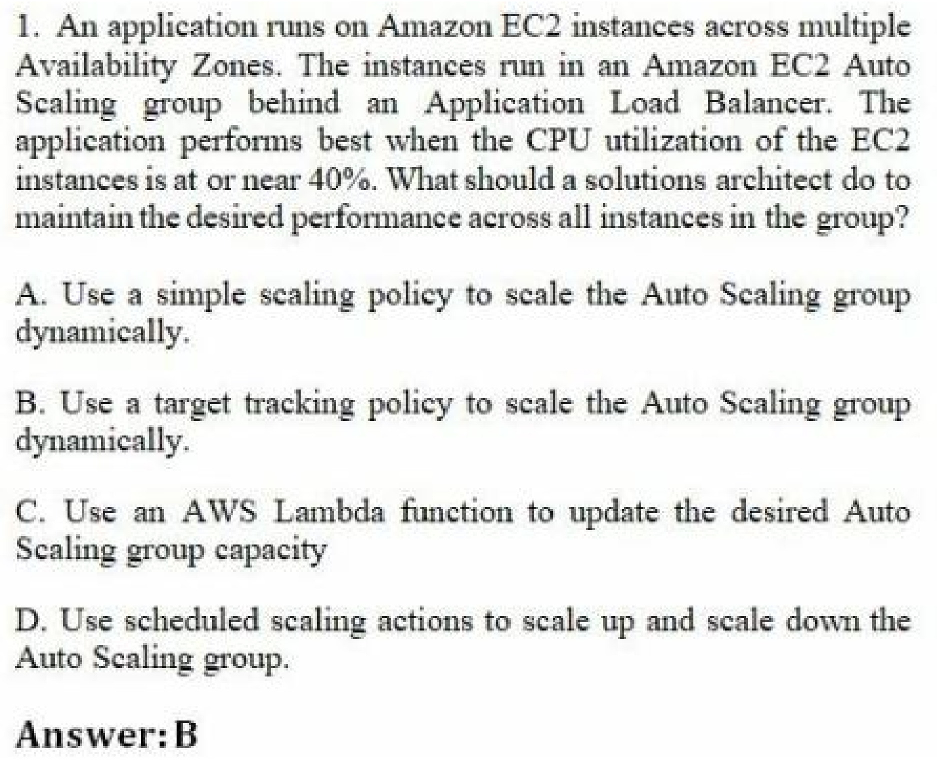
- SQS Queue allows us to decouple and scale microservices, distributed systems, and server-less applications.
- Auto-scaling is performed depending on the SQS Queue messages.
- We can monitor the SQS Queue messages by creating a scaling policy with a thresh hold.
- SQS is a decoupling service in terms of scalability.
- Front end server is used to send messages to SQS Queue.
- We can trigger scaling of an auto-scaling group depending on the number of messages in queue.
- Increased number of SQS messages leads to increased number of services.

AWS ATHENA & AWS GLUE



- AWS Athena is used for examining data in Amazon S3.
- AWS Athena uses standard SQL.
- Results are produced within seconds using Athena.
- Amazon Athena is suitable for data analysis.
- AWS GLUE is a fully managed ETL (extract transform load) service.
- Glue enables its users to compose and load their data for analytics.
- AWS Athena incorporates with AWS Glue to form a unified metadata repository across different services.
- They help discover schemas and populate user's Catalog.
- They maintain schema versioning and use modified tables and partition definitions.

Exam Practice Questions!



1. An application runs on Amazon EC2 instances across multiple Availability Zones. The instances run in an Amazon EC2 Auto Scaling group behind an Application Load Balancer. The application performs best when the CPU utilization of the EC2 instances is at or near 40%. What should a solutions architect do to maintain the desired performance across all instances in the group?

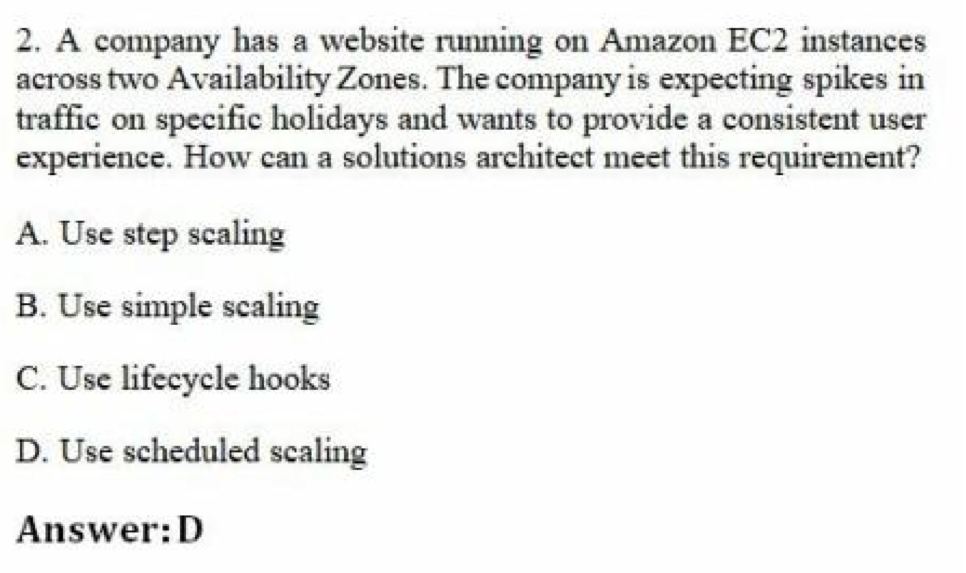
A. Use a simple scaling policy to scale the Auto Scaling group dynamically.

B. Use a target tracking policy to scale the Auto Scaling group dynamically.

C. Use an AWS Lambda function to update the desired Auto Scaling group capacity

D. Use scheduled scaling actions to scale up and scale down the Auto Scaling group.

Answer: B



2. A company has a website running on Amazon EC2 instances across two Availability Zones. The company is expecting spikes in traffic on specific holidays and wants to provide a consistent user experience. How can a solutions architect meet this requirement?

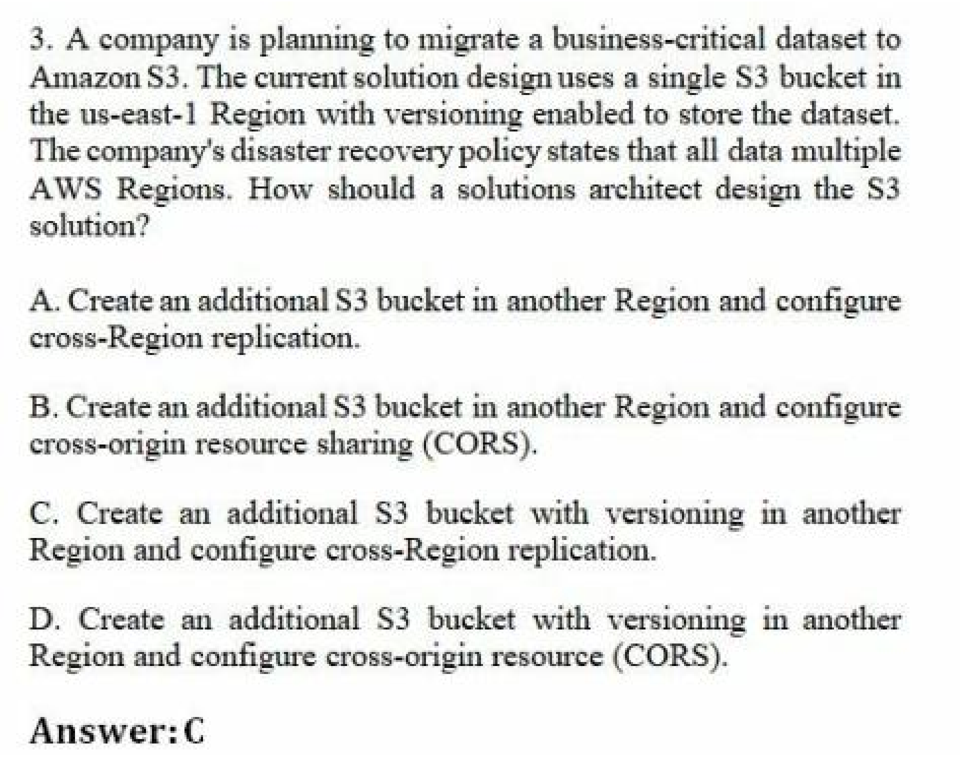
A. Use step scaling

B. Use simple scaling

C. Use lifecycle hooks

D. Use scheduled scaling

Answer: D



3. A company is planning to migrate a business-critical dataset to Amazon S3. The current solution design uses a single S3 bucket in the us-east-1 Region with versioning enabled to store the dataset. The company's disaster recovery policy states that all data multiple AWS Regions. How should a solutions architect design the S3 solution?

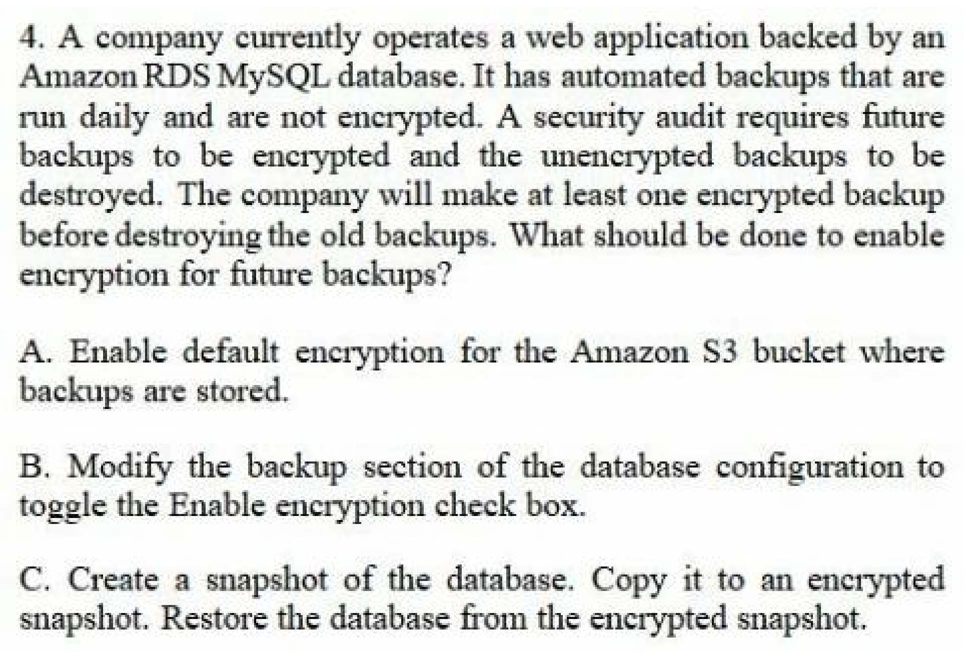
A. Create an additional S3 bucket in another Region and configure cross-Region replication.

B. Create an additional S3 bucket in another Region and configure cross-origin resource sharing (CORS).

C. Create an additional S3 bucket with versioning in another Region and configure cross-Region replication.

D. Create an additional S3 bucket with versioning in another Region and configure cross-origin resource (CORS).

Answer: C

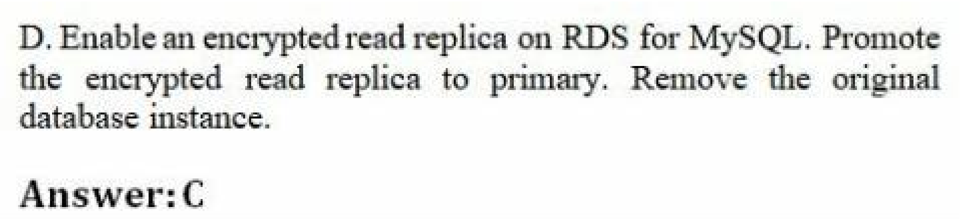


4. A company currently operates a web application backed by an Amazon RDS MySQL database. It has automated backups that are run daily and are not encrypted. A security audit requires future backups to be encrypted and the unencrypted backups to be destroyed. The company will make at least one encrypted backup before destroying the old backups. What should be done to enable encryption for future backups?

A. Enable default encryption for the Amazon S3 bucket where backups are stored.

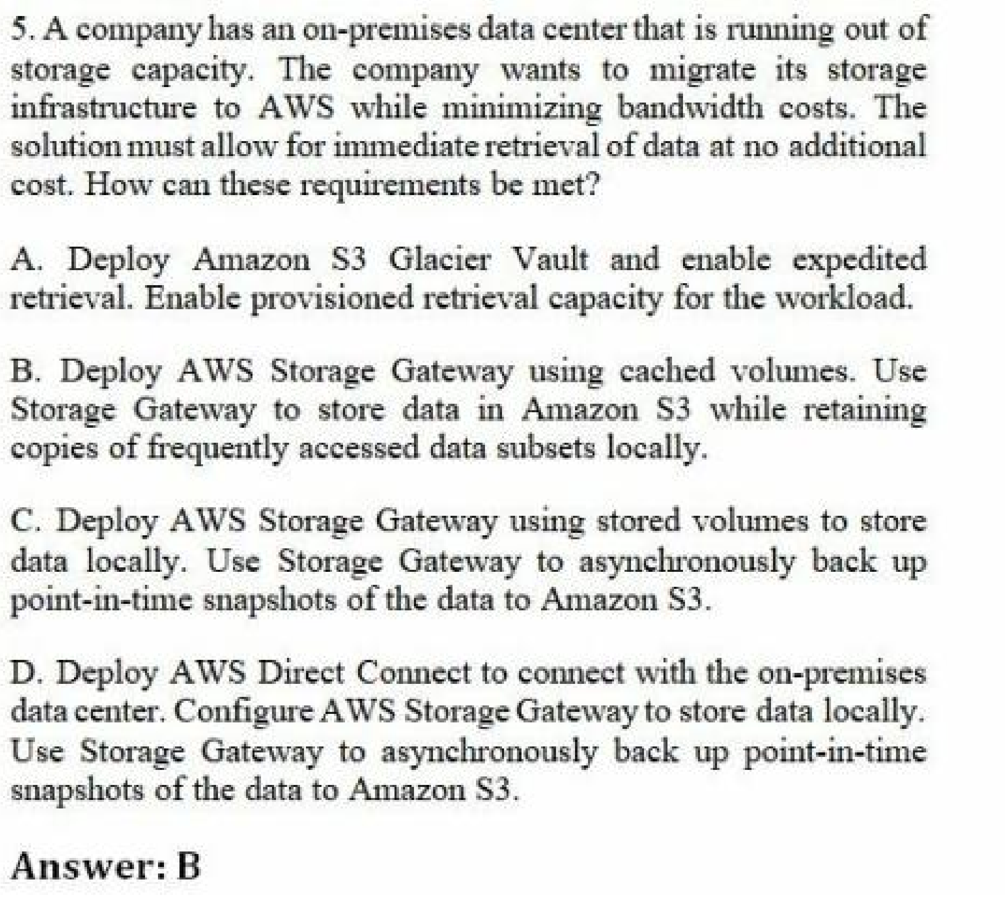
B. Modify the backup section of the database configuration to toggle the Enable encryption check box.

C. Create a snapshot of the database. Copy it to an encrypted snapshot. Restore the database from the encrypted snapshot.



D. Enable an encrypted read replica on RDS for MySQL. Promote the encrypted read replica to primary. Remove the original database instance.

Answer: C



5. A company has an on-premises data center that is running out of storage capacity. The company wants to migrate its storage infrastructure to AWS while minimizing bandwidth costs. The solution must allow for immediate retrieval of data at no additional cost. How can these requirements be met?

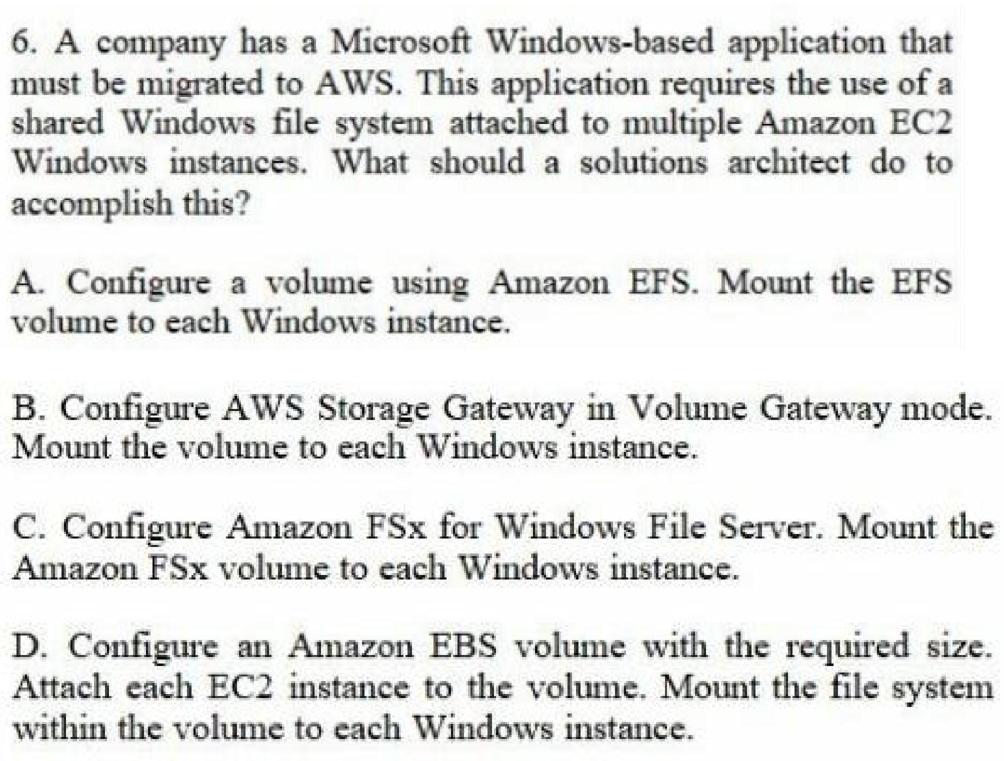
A. Deploy Amazon S3 Glacier Vault and enable expedited retrieval. Enable provisioned retrieval capacity for the workload.

B. Deploy AWS Storage Gateway using cached volumes. Use Storage Gateway to store data in Amazon S3 while retaining copies of frequently accessed data subsets locally.

C. Deploy AWS Storage Gateway using stored volumes to store data locally. Use Storage Gateway to asynchronously back up point-in-time snapshots of the data to Amazon S3.

D. Deploy AWS Direct Connect to connect with the on-premises data center. Configure AWS Storage Gateway to store data locally. Use Storage Gateway to asynchronously back up point-in-time snapshots of the data to Amazon S3.

Answer: B



6. A company has a Microsoft Windows-based application that must be migrated to AWS. This application requires the use of a shared Windows file system attached to multiple Amazon EC2 Windows instances. What should a solutions architect do to accomplish this?

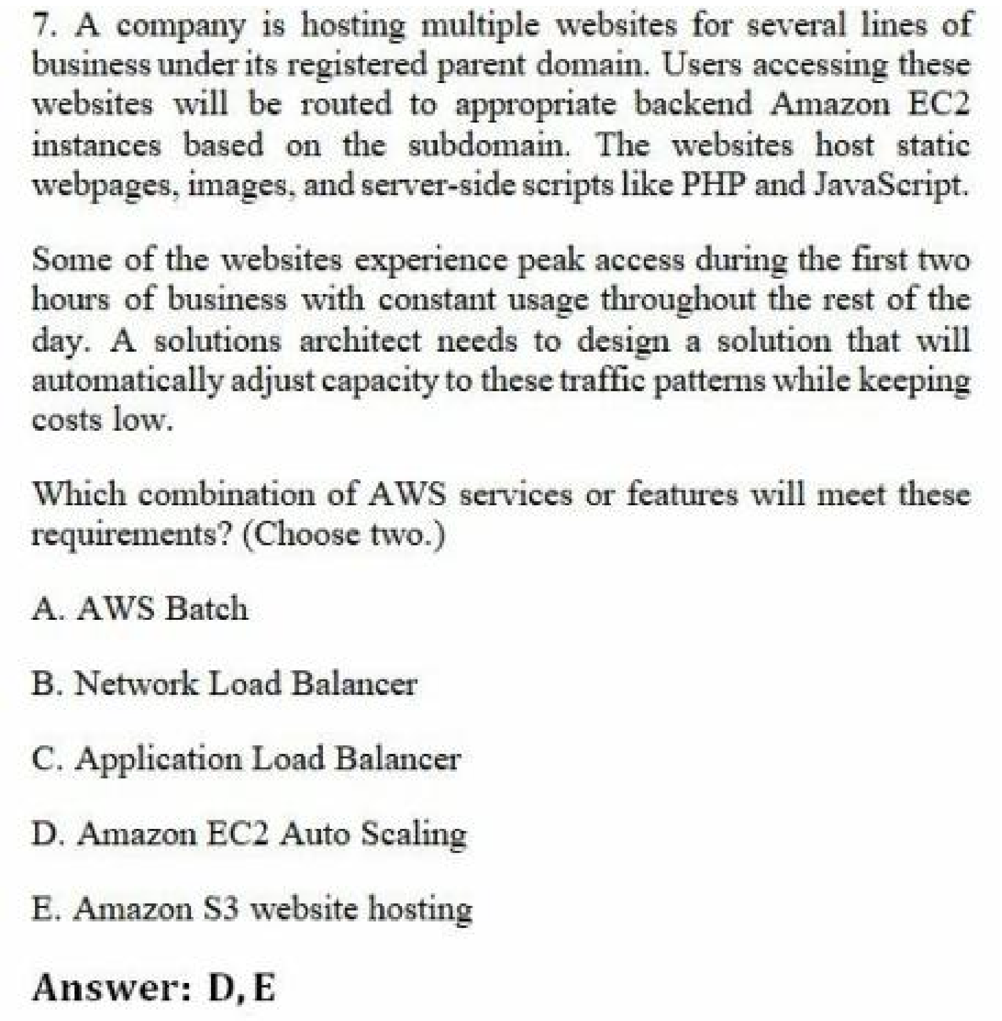
A. Configure a volume using Amazon EFS. Mount the EFS volume to each Windows instance.

B. Configure AWS Storage Gateway in Volume Gateway mode. Mount the volume to each Windows instance.

C. Configure Amazon FSx for Windows File Server. Mount the Amazon FSx volume to each Windows instance.

D. Configure an Amazon EBS volume with the required size. Attach each EC2 instance to the volume. Mount the file system within the volume to each Windows instance.

Answer: C



7. A company is hosting multiple websites for several lines of business under its registered parent domain. Users accessing these websites will be routed to appropriate backend Amazon EC2 instances based on the subdomain. The websites host static webpages, images, and server-side scripts like PHP and JavaScript.

Some of the websites experience peak access during the first two hours of business with constant usage throughout the rest of the day. A solutions architect needs to design a solution that will automatically adjust capacity to these traffic patterns while keeping costs low.

Which combination of AWS services or features will meet these requirements? (Choose two.)

A. AWS Batch

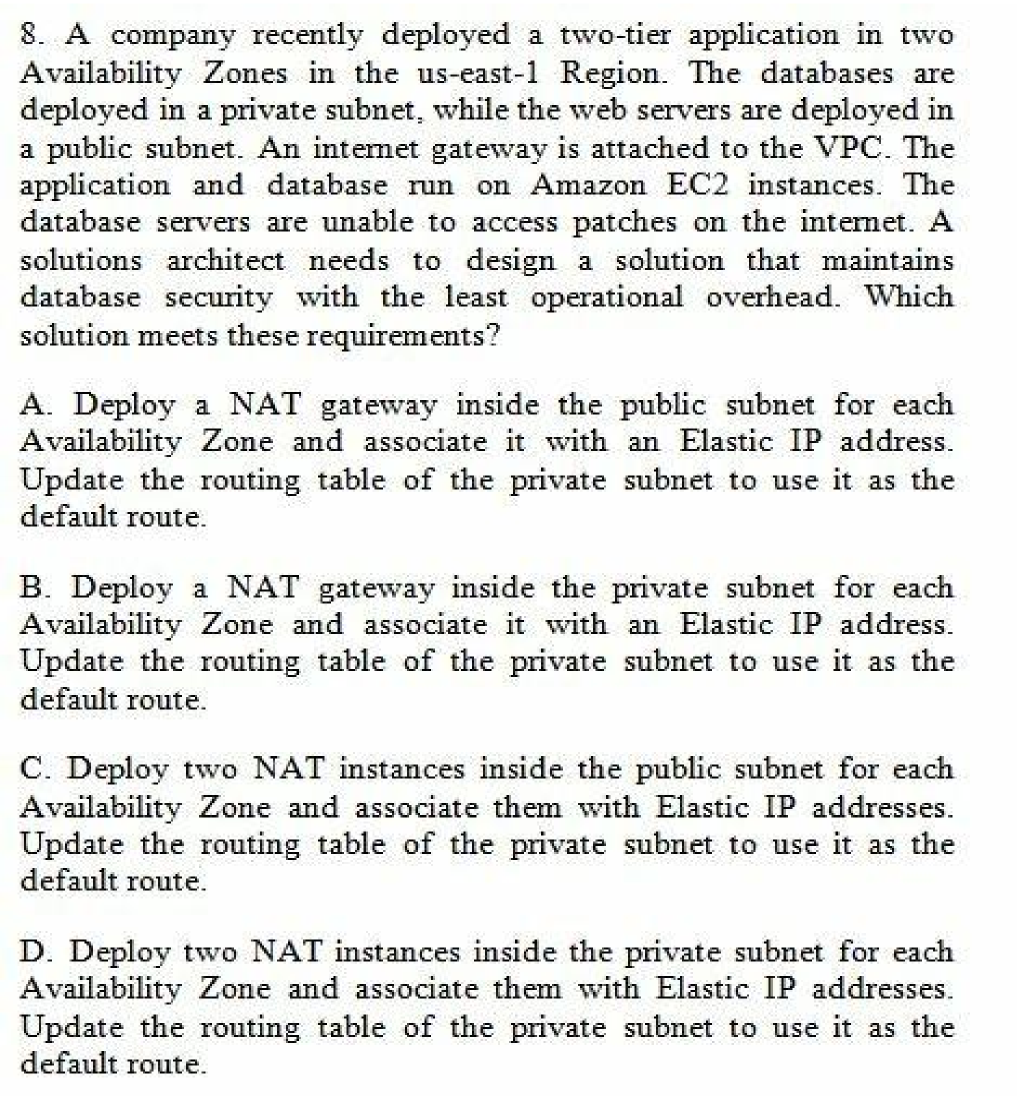
B. Network Load Balancer

C. Application Load Balancer

D. Amazon EC2 Auto Scaling

E. Amazon S3 website hosting

Answer: D, E



8. A company recently deployed a two-tier application in two Availability Zones in the us-east-1 Region. The databases are deployed in a private subnet, while the web servers are deployed in a public subnet. An internet gateway is attached to the VPC. The application and database run on Amazon EC2 instances. The database servers are unable to access patches on the internet. A solutions architect needs to design a solution that maintains database security with the least operational overhead. Which solution meets these requirements?

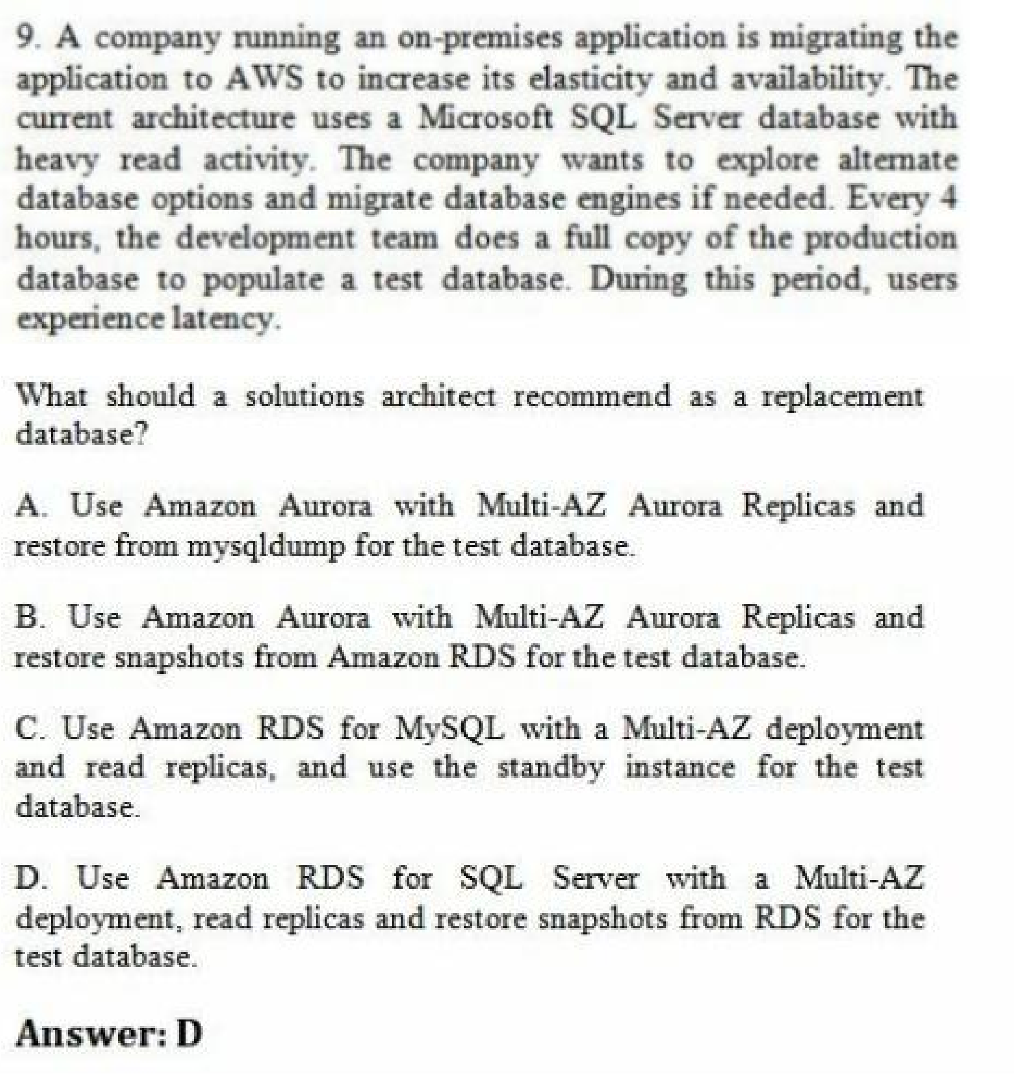
A. Deploy a NAT gateway inside the public subnet for each Availability Zone and associate it with an Elastic IP address. Update the routing table of the private subnet to use it as the default route.

B. Deploy a NAT gateway inside the private subnet for each Availability Zone and associate it with an Elastic IP address. Update the routing table of the private subnet to use it as the default route.

C. Deploy two NAT instances inside the public subnet for each Availability Zone and associate them with Elastic IP addresses. Update the routing table of the private subnet to use it as the default route.

D. Deploy two NAT instances inside the private subnet for each Availability Zone and associate them with Elastic IP addresses. Update the routing table of the private subnet to use it as the default route.

Answer: A



9. A company running an on-premises application is migrating the application to AWS to increase its elasticity and availability. The current architecture uses a Microsoft SQL Server database with heavy read activity. The company wants to explore alternate database options and migrate database engines if needed. Every 4 hours, the development team does a full copy of the production database to populate a test database. During this period, users experience latency.

What should a solutions architect recommend as a replacement database?

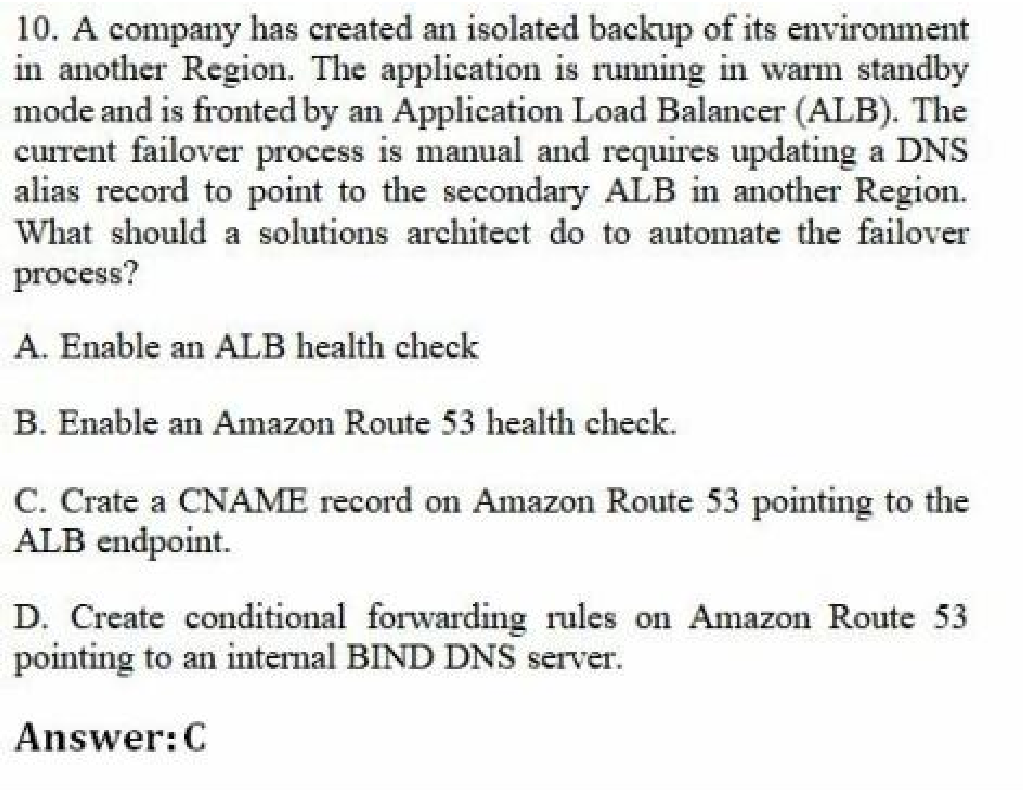
A. Use Amazon Aurora with Multi-AZ Aurora Replicas and restore from mysqldump for the test database.

B. Use Amazon Aurora with Multi-AZ Aurora Replicas and restore snapshots from Amazon RDS for the test database.

C. Use Amazon RDS for MySQL with a Multi-AZ deployment and read replicas, and use the standby instance for the test database.

D. Use Amazon RDS for SQL Server with a Multi-AZ deployment, read replicas and restore snapshots from RDS for the test database.

Answer: D



10. A company has created an isolated backup of its environment in another Region. The application is running in warm standby mode and is fronted by an Application Load Balancer (ALB). The current failover process is manual and requires updating a DNS alias record to point to the secondary ALB in another Region. What should a solutions architect do to automate the failover process?

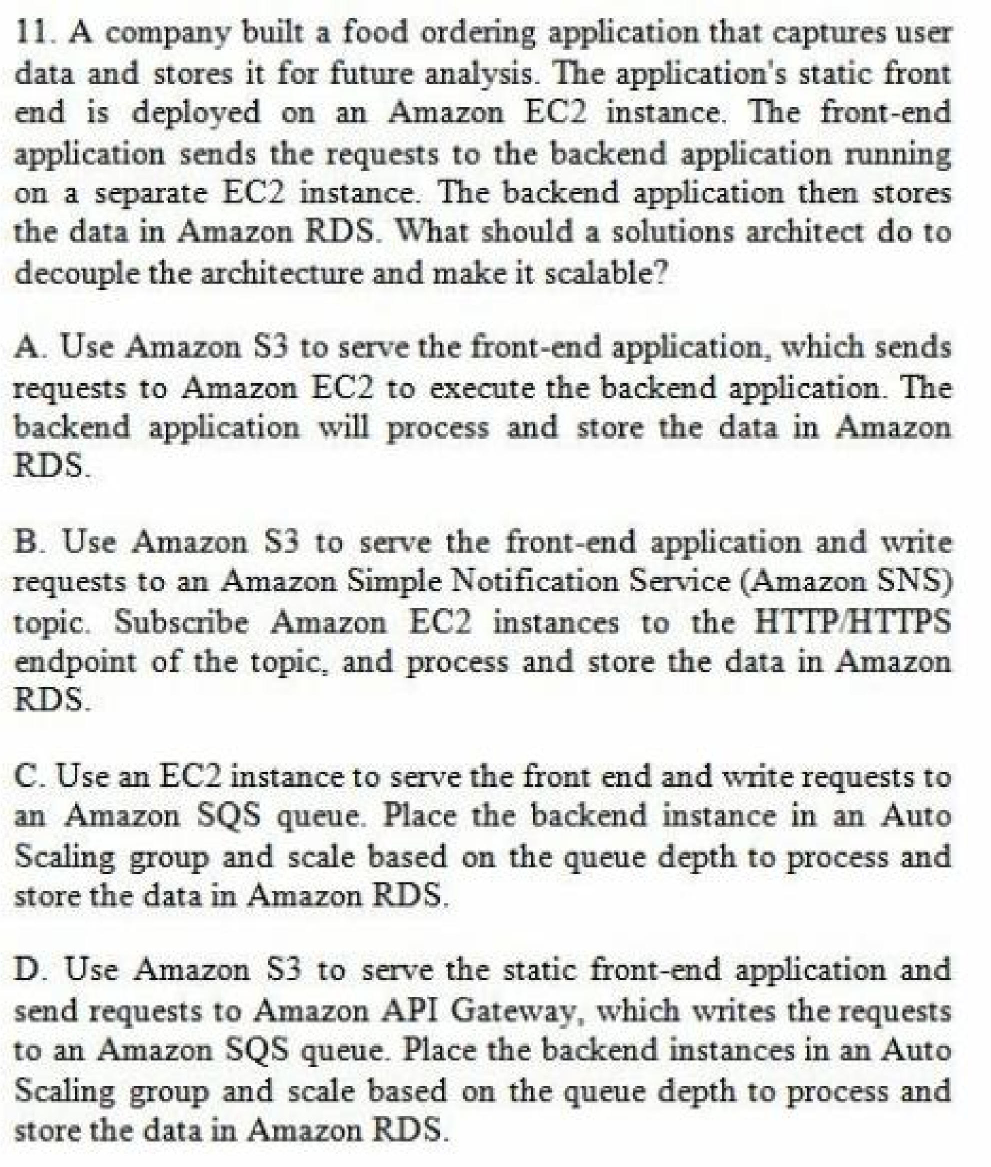
A. Enable an ALB health check

B. Enable an Amazon Route 53 health check.

C. Crate a CNAME record on Amazon Route 53 pointing to the ALB endpoint.

D. Create conditional forwarding rules on Amazon Route 53 pointing to an internal BIND DNS server.

Answer:C



11. A company built a food ordering application that captures user data and stores it for future analysis. The application's static front end is deployed on an Amazon EC2 instance. The front-end application sends the requests to the backend application running on a separate EC2 instance. The backend application then stores the data in Amazon RDS. What should a solutions architect do to decouple the architecture and make it scalable?

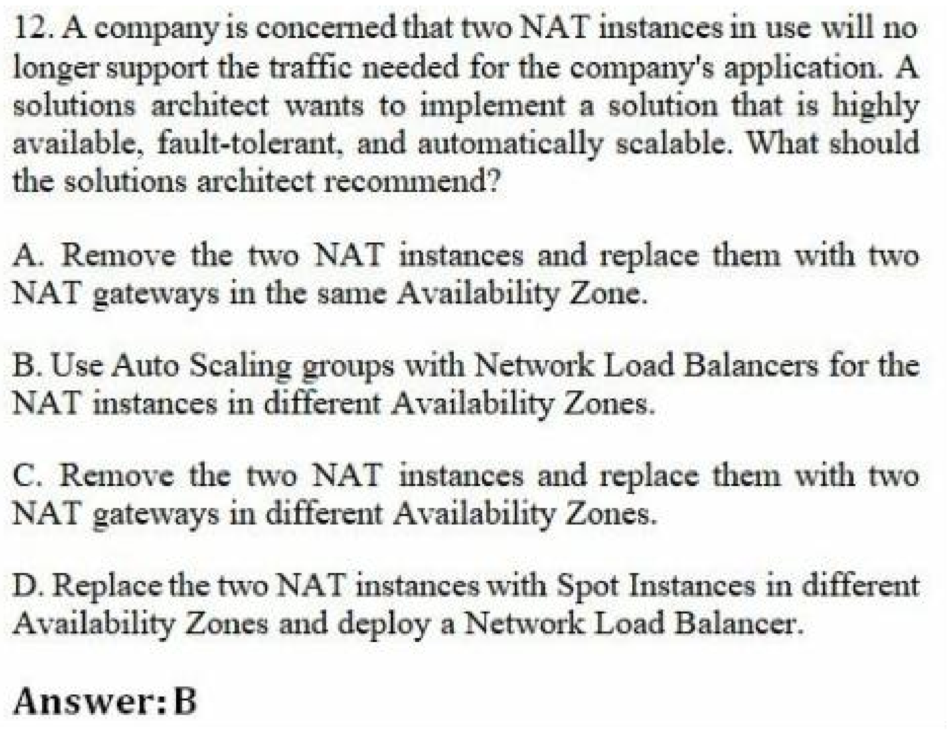
A. Use Amazon S3 to serve the front-end application, which sends requests to Amazon EC2 to execute the backend application. The backend application will process and store the data in Amazon RDS.

B. Use Amazon S3 to serve the front-end application and write requests to an Amazon Simple Notification Service (Amazon SNS) topic. Subscribe Amazon EC2 instances to the HTTP/HTTPS endpoint of the topic, and process and store the data in Amazon RDS.

C. Use an EC2 instance to serve the front end and write requests to an Amazon SQS queue. Place the backend instance in an Auto Scaling group and scale based on the queue depth to process and store the data in Amazon RDS.

D. Use Amazon S3 to serve the static front-end application and send requests to Amazon API Gateway, which writes the requests to an Amazon SQS queue. Place the backend instances in an Auto Scaling group and scale based on the queue depth to process and store the data in Amazon RDS.

Answer: D



12. A company is concerned that two NAT instances in use will no longer support the traffic needed for the company's application. A solutions architect wants to implement a solution that is highly available, fault-tolerant, and automatically scalable. What should the solutions architect recommend?

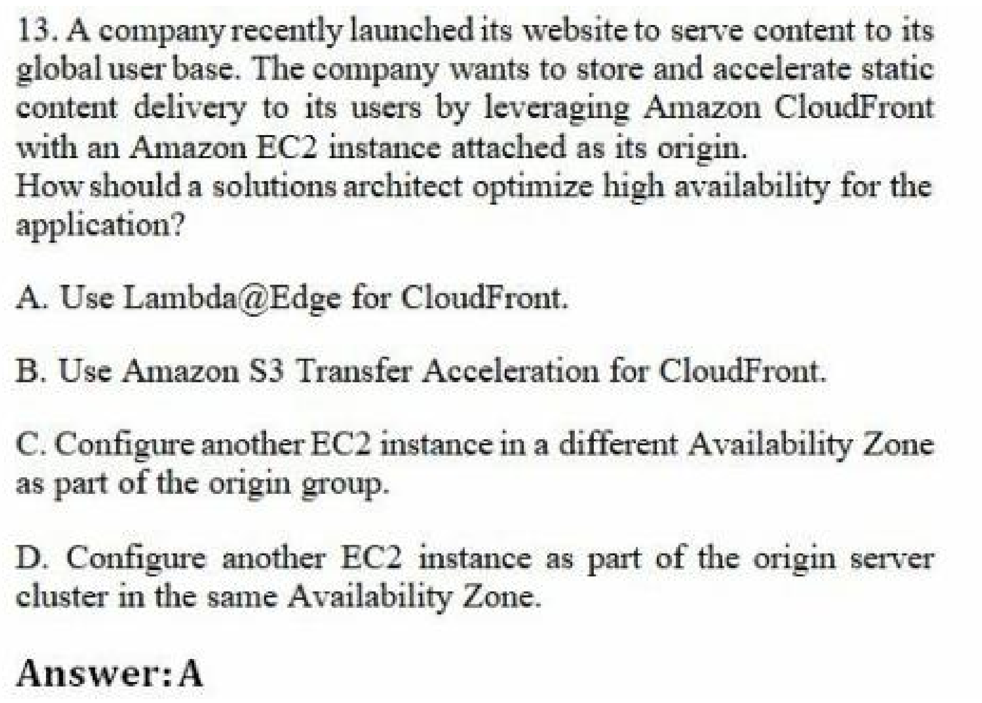
A. Remove the two NAT instances and replace them with two NAT gateways in the same Availability Zone.

B. Use Auto Scaling groups with Network Load Balancers for the NAT instances in different Availability Zones.

C. Remove the two NAT instances and replace them with two NAT gateways in different Availability Zones.

D. Replace the two NAT instances with Spot Instances in different Availability Zones and deploy a Network Load Balancer.

Answer: B



13. A company recently launched its website to serve content to its global user base. The company wants to store and accelerate static content delivery to its users by leveraging Amazon CloudFront with an Amazon EC2 instance attached as its origin.
How should a solutions architect optimize high availability for the application?

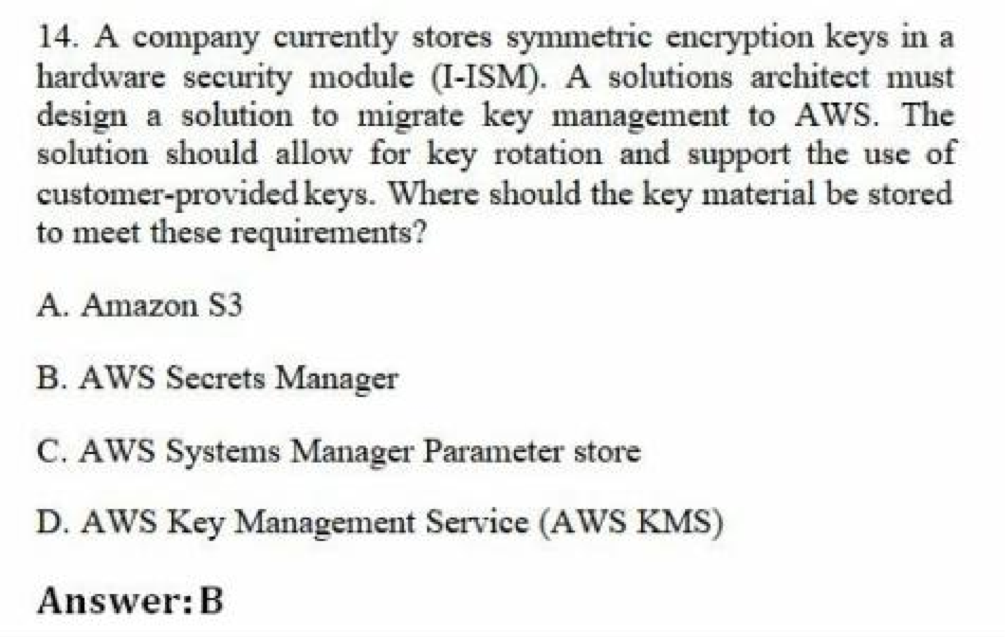
A. Use Lambda@Edge for CloudFront.

B. Use Amazon S3 Transfer Acceleration for CloudFront.

C. Configure another EC2 instance in a different Availability Zone as part of the origin group.

D. Configure another EC2 instance as part of the origin server cluster in the same Availability Zone.

Answer: A



14. A company currently stores symmetric encryption keys in a hardware security module (I-ISM). A solutions architect must design a solution to migrate key management to AWS. The solution should allow for key rotation and support the use of customer-provided keys. Where should the key material be stored to meet these requirements?

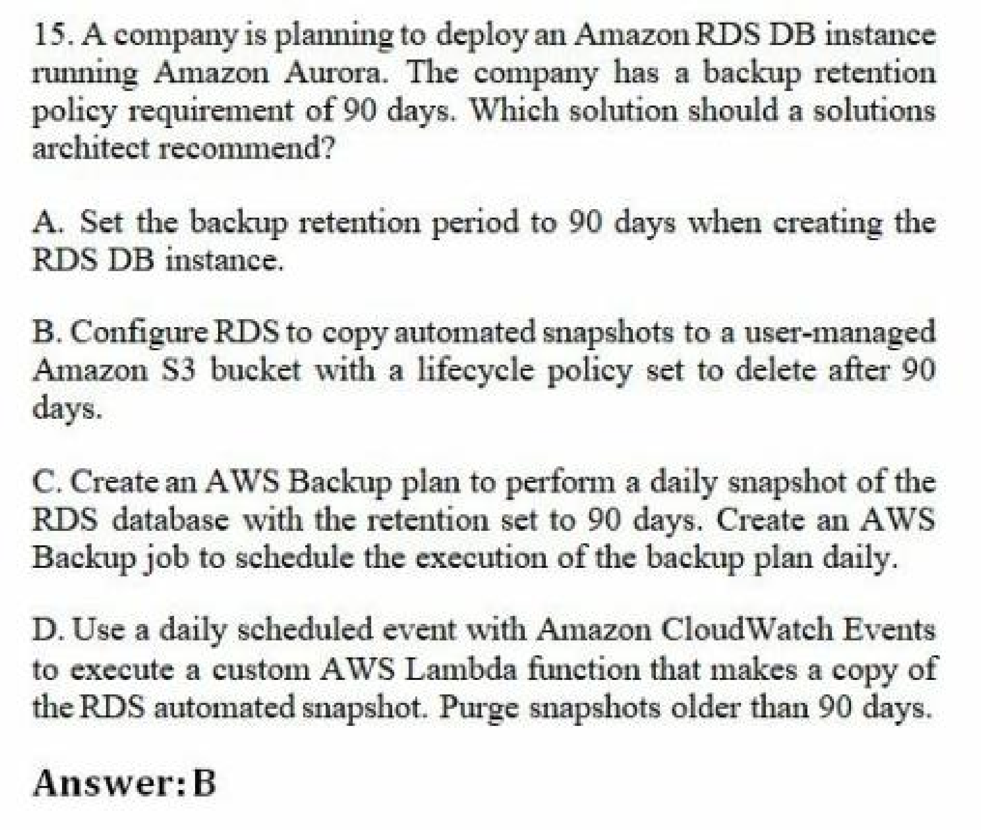
A. Amazon S3

B. AWS Secrets Manager

C. AWS Systems Manager Parameter store

D. AWS Key Management Service (AWS KMS)

Answer: B



15. A company is planning to deploy an Amazon RDS DB instance running Amazon Aurora. The company has a backup retention policy requirement of 90 days. Which solution should a solutions architect recommend?

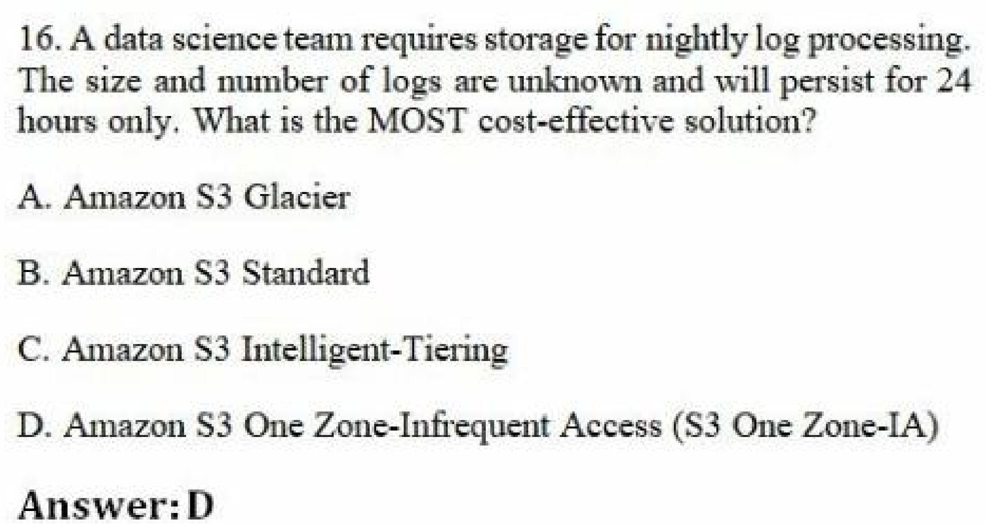
A. Set the backup retention period to 90 days when creating the RDS DB instance.

B. Configure RDS to copy automated snapshots to a user-managed Amazon S3 bucket with a lifecycle policy set to delete after 90 days.

C. Create an AWS Backup plan to perform a daily snapshot of the RDS database with the retention set to 90 days. Create an AWS Backup job to schedule the execution of the backup plan daily.

D. Use a daily scheduled event with Amazon CloudWatch Events to execute a custom AWS Lambda function that makes a copy of the RDS automated snapshot. Purge snapshots older than 90 days.

Answer: B



16. A data science team requires storage for nightly log processing. The size and number of logs are unknown and will persist for 24 hours only. What is the MOST cost-effective solution?

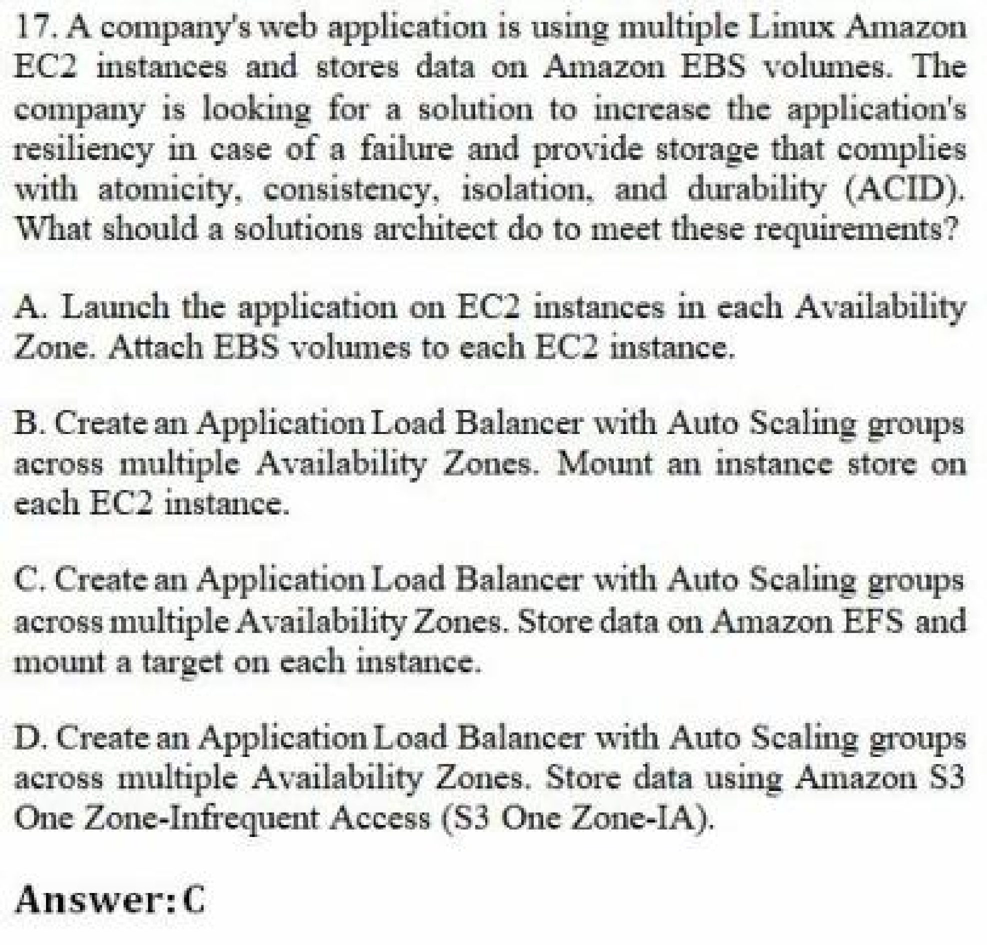
A. Amazon S3 Glacier

B. Amazon S3 Standard

C. Amazon S3 Intelligent-Tiering

D. Amazon S3 One Zone-Infrequent Access (S3 One Zone-IA)

Answer: D



17. A company's web application is using multiple Linux Amazon EC2 instances and stores data on Amazon EBS volumes. The company is looking for a solution to increase the application's resiliency in case of a failure and provide storage that complies with atomicity, consistency, isolation, and durability (ACID). What should a solutions architect do to meet these requirements?

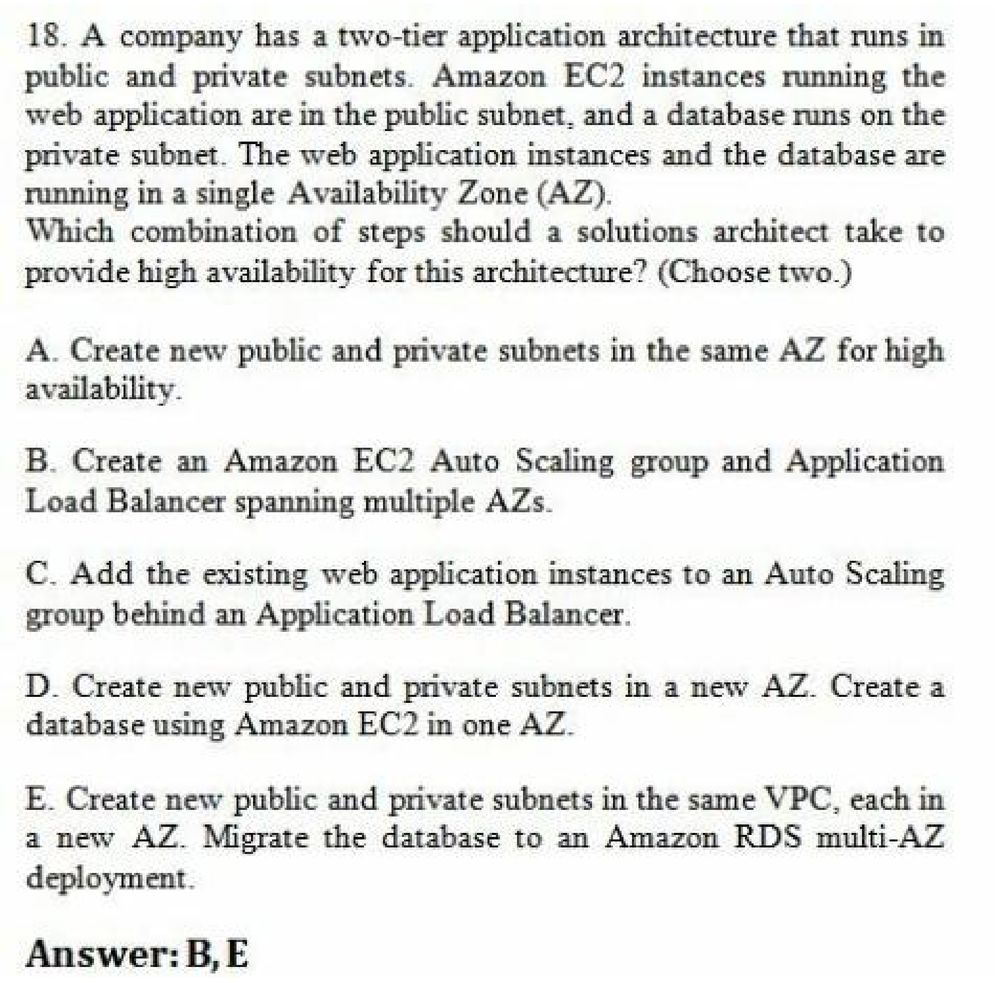
A. Launch the application on EC2 instances in each Availability Zone. Attach EBS volumes to each EC2 instance.

B. Create an Application Load Balancer with Auto Scaling groups across multiple Availability Zones. Mount an instance store on each EC2 instance.

C. Create an Application Load Balancer with Auto Scaling groups across multiple Availability Zones. Store data on Amazon EFS and mount a target on each instance.

D. Create an Application Load Balancer with Auto Scaling groups across multiple Availability Zones. Store data using Amazon S3 One Zone-Infrequent Access (S3 One Zone-IA).

Answer: C



18. A company has a two-tier application architecture that runs in public and private subnets. Amazon EC2 instances running the web application are in the public subnet, and a database runs on the private subnet. The web application instances and the database are running in a single Availability Zone (AZ).
Which combination of steps should a solutions architect take to provide high availability for this architecture? (Choose two.)

A. Create new public and private subnets in the same AZ for high availability.

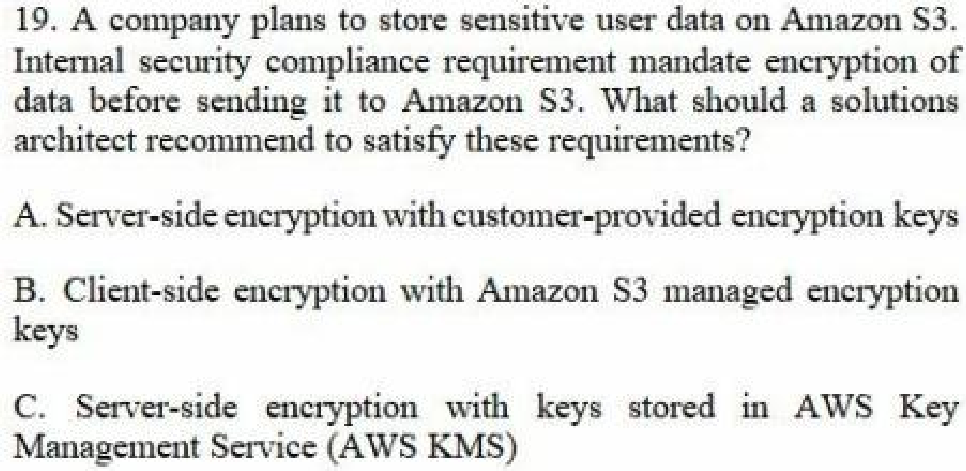
B. Create an Amazon EC2 Auto Scaling group and Application Load Balancer spanning multiple AZs.

C. Add the existing web application instances to an Auto Scaling group behind an Application Load Balancer.

D. Create new public and private subnets in a new AZ. Create a database using Amazon EC2 in one AZ.

E. Create new public and private subnets in the same VPC, each in a new AZ. Migrate the database to an Amazon RDS multi-AZ deployment.

Answer: B, E

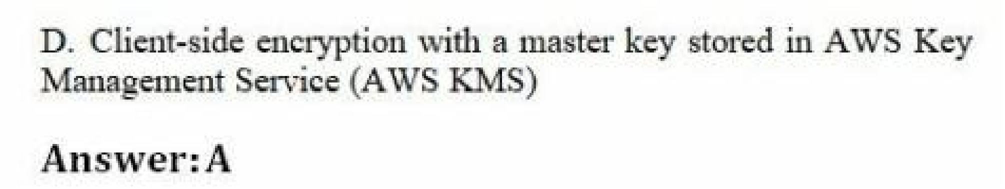


19. A company plans to store sensitive user data on Amazon S3. Internal security compliance requirement mandate encryption of data before sending it to Amazon S3. What should a solutions architect recommend to satisfy these requirements?

A. Server-side encryption with customer-provided encryption keys

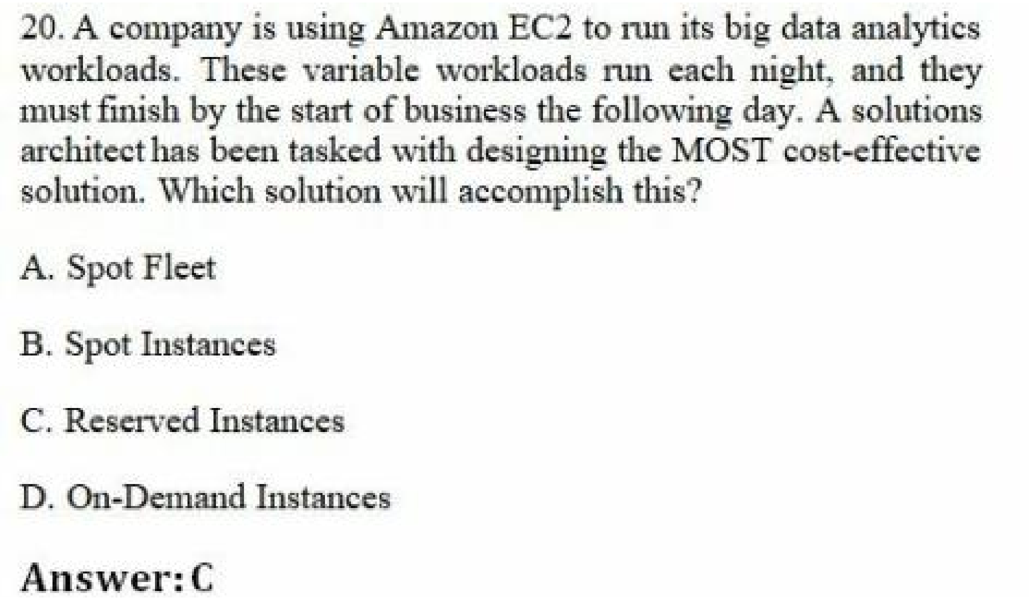
B. Client-side encryption with Amazon S3 managed encryption keys

C. Server-side encryption with keys stored in AWS Key Management Service (AWS KMS)



D. Client-side encryption with a master key stored in AWS Key Management Service (AWS KMS)

Answer:A



20. A company is using Amazon EC2 to run its big data analytics workloads. These variable workloads run each night, and they must finish by the start of business the following day. A solutions architect has been tasked with designing the MOST cost-effective solution. Which solution will accomplish this?

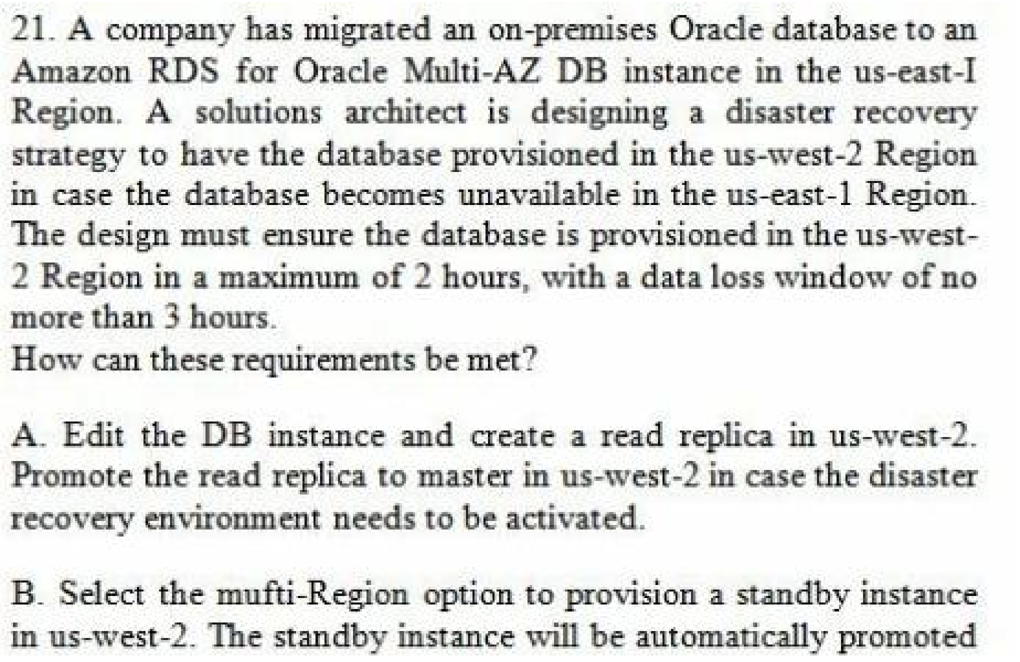
A. Spot Fleet

B. Spot Instances

C. Reserved Instances

D. On-Demand Instances

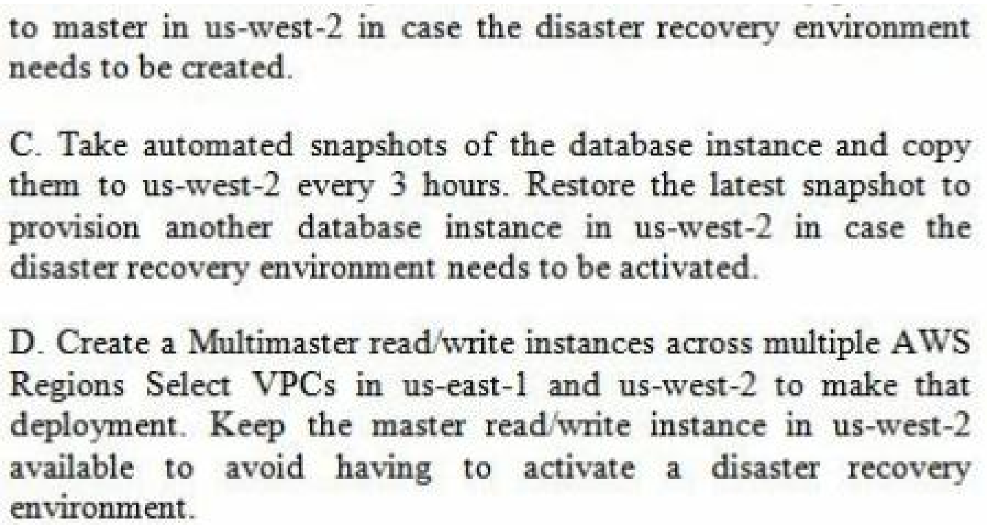
Answer:C



21. A company has migrated an on-premises Oracle database to an Amazon RDS for Oracle Multi-AZ DB instance in the us-east-I Region. A solutions architect is designing a disaster recovery strategy to have the database provisioned in the us-west-2 Region in case the database becomes unavailable in the us-east-1 Region. The design must ensure the database is provisioned in the us-west-2 Region in a maximum of 2 hours, with a data loss window of no more than 3 hours.
How can these requirements be met?

A. Edit the DB instance and create a read replica in us-west-2. Promote the read replica to master in us-west-2 in case the disaster recovery environment needs to be activated.

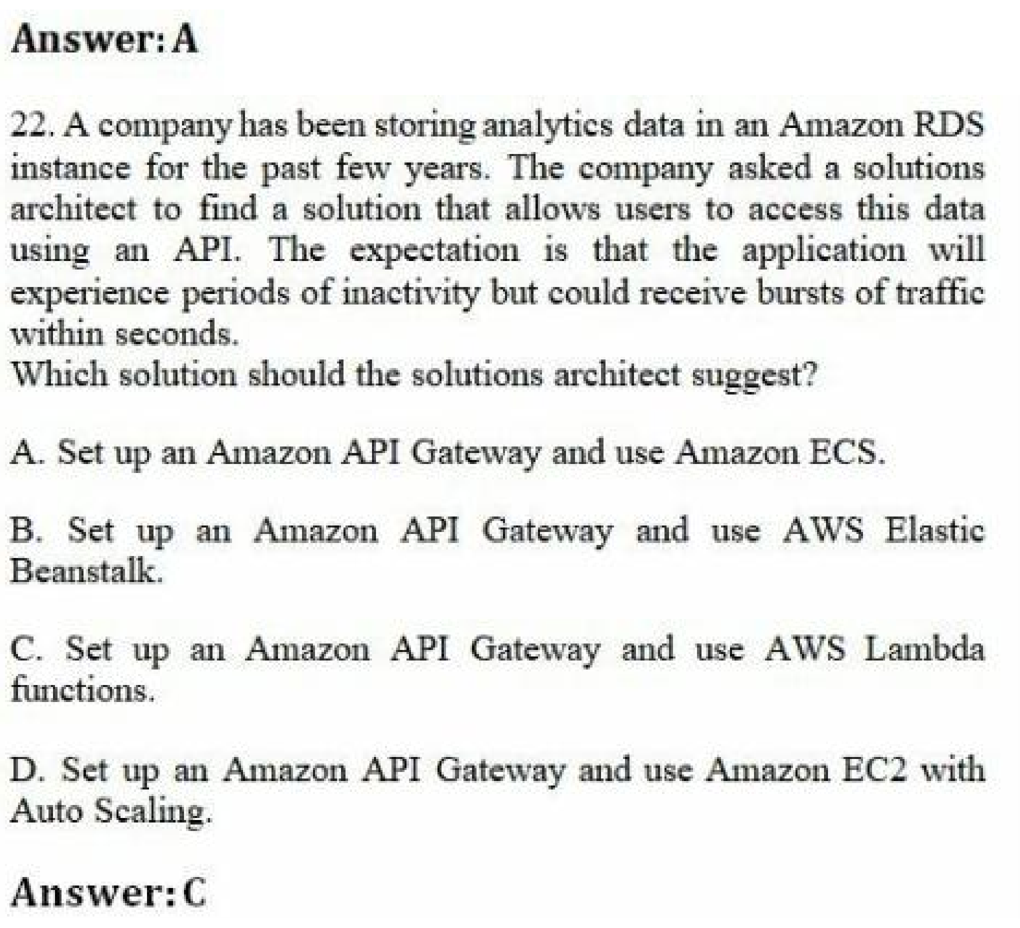
B. Select the mufti-Region option to provision a standby instance in us-west-2. The standby instance will be automatically promoted



to master in us-west-2 in case the disaster recovery environment needs to be created.

C. Take automated snapshots of the database instance and copy them to us-west-2 every 3 hours. Restore the latest snapshot to provision another database instance in us-west-2 in case the disaster recovery environment needs to be activated.

D. Create a Multimaster read/write instances across multiple AWS Regions Select VPCs in us-east-1 and us-west-2 to make that deployment. Keep the master read/write instance in us-west-2 available to avoid having to activate a disaster recovery environment.



Answer: A

22. A company has been storing analytics data in an Amazon RDS instance for the past few years. The company asked a solutions architect to find a solution that allows users to access this data using an API. The expectation is that the application will experience periods of inactivity but could receive bursts of traffic within seconds.
Which solution should the solutions architect suggest?

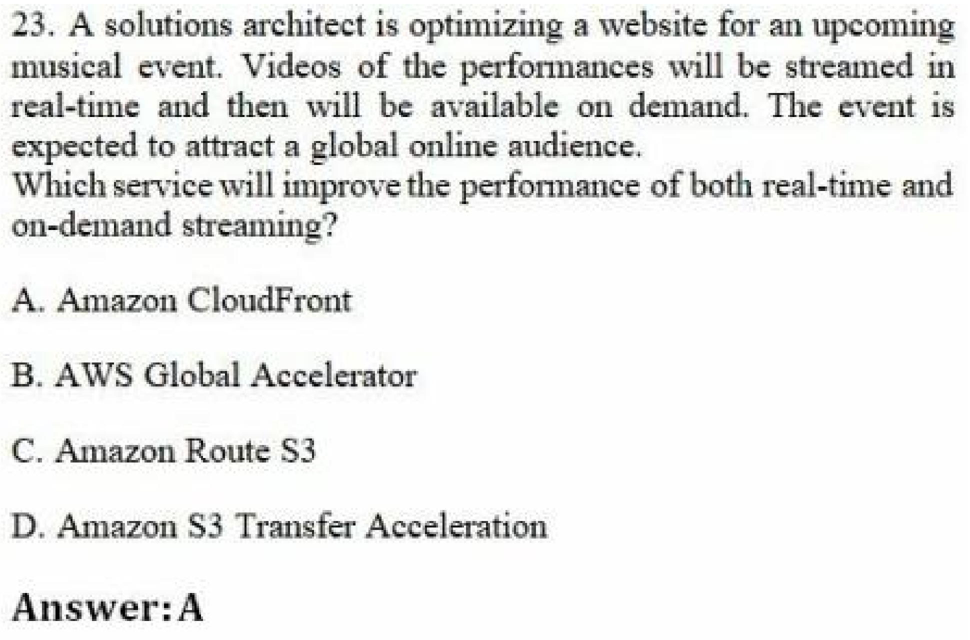
A. Set up an Amazon API Gateway and use Amazon ECS.

B. Set up an Amazon API Gateway and use AWS Elastic Beanstalk.

C. Set up an Amazon API Gateway and use AWS Lambda functions.

D. Set up an Amazon API Gateway and use Amazon EC2 with Auto Scaling.

Answer: C



23. A solutions architect is optimizing a website for an upcoming musical event. Videos of the performances will be streamed in real-time and then will be available on demand. The event is expected to attract a global online audience.
Which service will improve the performance of both real-time and on-demand streaming?

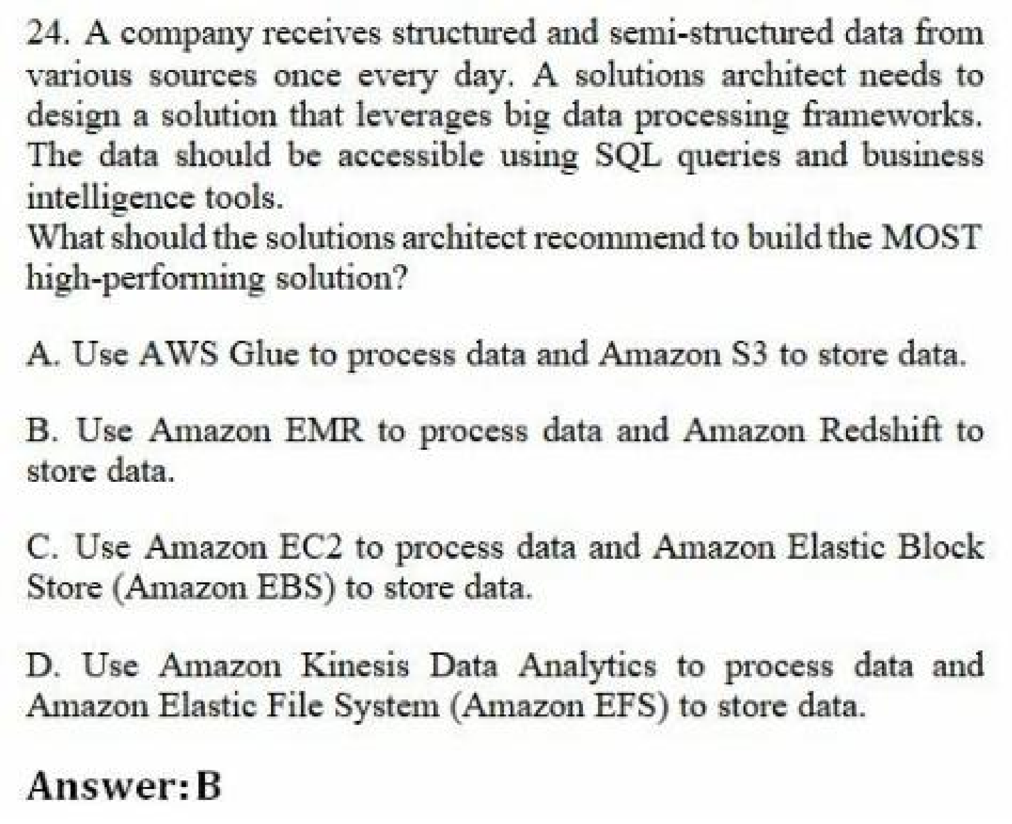
A. Amazon CloudFront

B. AWS Global Accelerator

C. Amazon Route S3

D. Amazon S3 Transfer Acceleration

Answer:A



24. A company receives structured and semi-structured data from various sources once every day. A solutions architect needs to design a solution that leverages big data processing frameworks. The data should be accessible using SQL queries and business intelligence tools.
What should the solutions architect recommend to build the MOST high-performing solution?

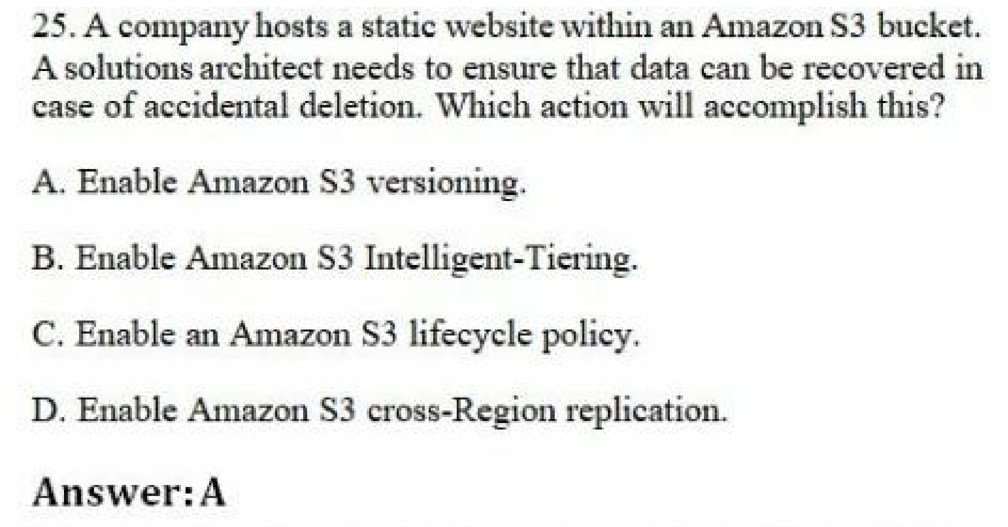
A. Use AWS Glue to process data and Amazon S3 to store data.

B. Use Amazon EMR to process data and Amazon Redshift to store data.

C. Use Amazon EC2 to process data and Amazon Elastic Block Store (Amazon EBS) to store data.

D. Use Amazon Kinesis Data Analytics to process data and Amazon Elastic File System (Amazon EFS) to store data.

Answer:B



25. A company hosts a static website within an Amazon S3 bucket. A solutions architect needs to ensure that data can be recovered in case of accidental deletion. Which action will accomplish this?

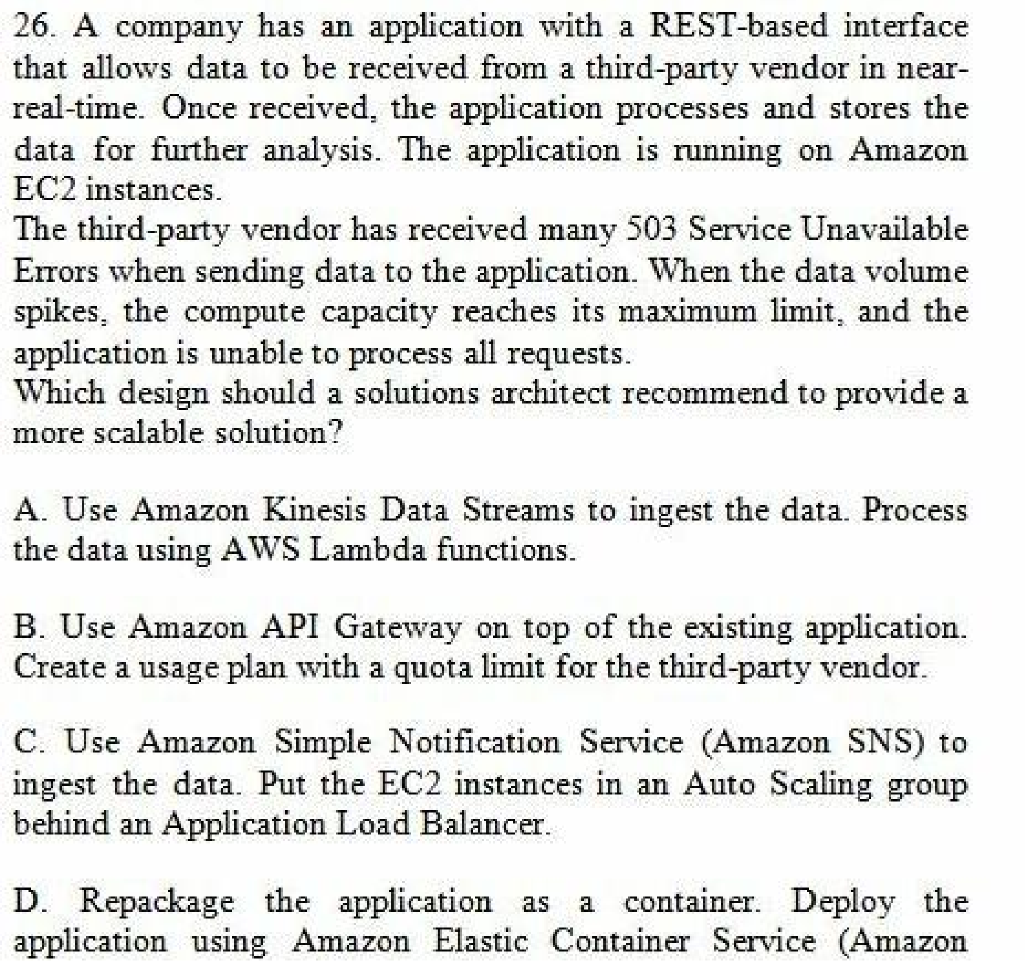
A. Enable Amazon S3 versioning.

B. Enable Amazon S3 Intelligent-Tiering.

C. Enable an Amazon S3 lifecycle policy.

D. Enable Amazon S3 cross-Region replication.

Answer:A



26. A company has an application with a REST-based interface that allows data to be received from a third-party vendor in near-real-time. Once received, the application processes and stores the data for further analysis. The application is running on Amazon EC2 instances.
The third-party vendor has received many 503 Service Unavailable Errors when sending data to the application. When the data volume spikes, the compute capacity reaches its maximum limit, and the application is unable to process all requests.
Which design should a solutions architect recommend to provide a more scalable solution?

A. Use Amazon Kinesis Data Streams to ingest the data. Process the data using AWS Lambda functions.

B. Use Amazon API Gateway on top of the existing application. Create a usage plan with a quota limit for the third-party vendor.

C. Use Amazon Simple Notification Service (Amazon SNS) to ingest the data. Put the EC2 instances in an Auto Scaling group behind an Application Load Balancer.

D. Repackage the application as a container. Deploy the application using Amazon Elastic Container Service (Amazon

ECS) using the EC2 launch type with an Auto Scaling group.

Answer: A

27. A company is using a VPC peering strategy to connect its VPCs in a single Region to allow for cross-communication. A recent increase in account creations and VPCs has made it difficult to maintain the VPC peering strategy, and the company expects to grow to hundreds of VPCs. There are also new requests to create site-to-site VPNs with some of the VPCs. A solutions architect has been tasked with creating a centrally managed networking setup for multiple accounts, VPCs, and VPNs.
Which networking solution meets these requirements?

A. Configure shared VPCs and VPNs and share with each other.

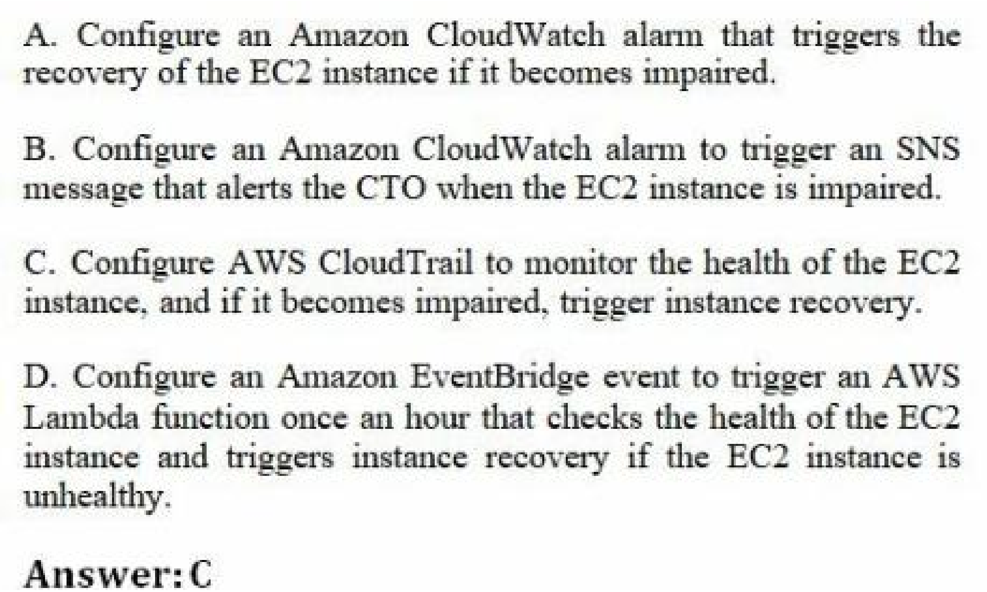
B. Configure a hub-and-spoke VPC and route all traffic through VPC peering.

C. Configure an AWS Direct Connect connection between all VPCs and VPNs.

D. Configure a transit gateway with AWS Transit Gateway and conned all VPCs and VPNs.

Answer: D

28. A monolithic application was recently migrated to AWS and is now running on a single Amazon EC2 instance. Due to application limitations, it is not possible to use automatic scaling to scale out the application. The chief technology officer (CTO) wants an automated solution to restore the EC2 instance in the unlikely event the underlying hardware fails.
What would allow for automatic recovery of the EC2 instance as quickly as possible?



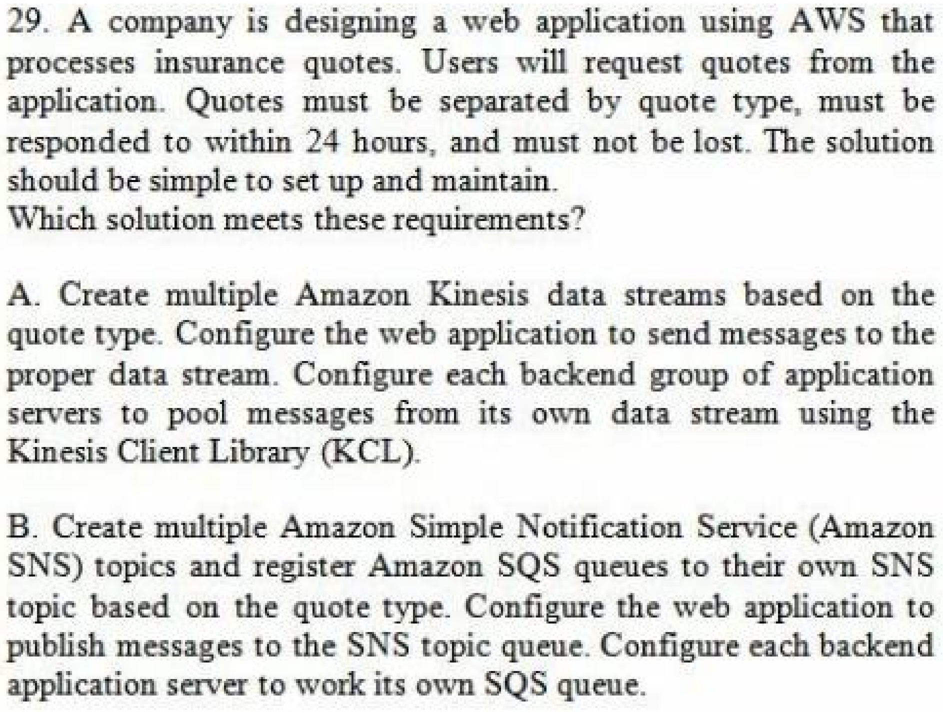
A. Configure an Amazon CloudWatch alarm that triggers the recovery of the EC2 instance if it becomes impaired.

B. Configure an Amazon CloudWatch alarm to trigger an SNS message that alerts the CTO when the EC2 instance is impaired.

C. Configure AWS CloudTrail to monitor the health of the EC2 instance, and if it becomes impaired, trigger instance recovery.

D. Configure an Amazon EventBridge event to trigger an AWS Lambda function once an hour that checks the health of the EC2 instance and triggers instance recovery if the EC2 instance is unhealthy.

Answer: C



29. A company is designing a web application using AWS that processes insurance quotes. Users will request quotes from the application. Quotes must be separated by quote type, must be responded to within 24 hours, and must not be lost. The solution should be simple to set up and maintain.
Which solution meets these requirements?

A. Create multiple Amazon Kinesis data streams based on the quote type. Configure the web application to send messages to the proper data stream. Configure each backend group of application servers to pool messages from its own data stream using the Kinesis Client Library (KCL).

B. Create multiple Amazon Simple Notification Service (Amazon SNS) topics and register Amazon SQS queues to their own SNS topic based on the quote type. Configure the web application to publish messages to the SNS topic queue. Configure each backend application server to work its own SQS queue.

C. Create a single Amazon Simple Notification Service (Amazon SNS) topic and subscribe the Amazon SQS queues to the SNS topic. Configure SNS message filtering to publish messages to the proper SQS queue based on the quote type. Configure each backend application server to work its own SQS queue.

D. Create multiple Amazon Kinesis Data Firehose delivery streams based on the quote type to deliver data streams to an Amazon Elastic Search Service (Amazon ES) cluster. Configure the web application to send messages to the proper delivery stream. Configure each backend group of application servers to search for the messages from Amazon ES and process them accordingly.

Answer: D

30. A three-tier web application processes orders from customers. The web tier consists of Amazon EC2 instances behind an Application Load Balancer, a middle tier of three EC2 instances decoupled from the web tier using Amazon SQS, and an Amazon DynamoDB backend. At peak times, customers who submit orders using the site have to wait much longer than normal to receive confirmations due to lengthy processing times. A solutions architect needs to reduce these processing times.
Which action will be MOST effective in accomplishing this?

A. Replace the SQS queue with Amazon Kinesis Data Firehose.

B. Use Amazon ElastiCache for Redis in front of the DynamoDB backend tier.

C. Add an Amazon CloudFront distribution to cache the responses for the web tier.

D. Use Amazon EC2 Auto Scaling to scale out the middle tier instances based on the SQS queue depth.

31. A company hosts its website on AWS. To address the highly variable demand, the company has implemented Amazon EC2 Auto Scaling. Management is concerned that the company is over-provisioning its infrastructure, especially at the front end of the three-tier application. A solutions architect needs to ensure costs are optimized without impacting performance. What should the solutions architect do to accomplish this?

A. Use Auto Scaling with Reserved Instances.

B. Use Auto Scaling with a scheduled scaling policy.

C. Use Auto Scaling with the suspend-resume feature.

D. Use Auto Scaling with a target tracking scaling policy.

Answer: B

32. A company uses an Amazon S3 bucket to store static images for its website. The company configured permissions to allow access to Amazon S3 objects by privileged users only. What should a solutions architect do to protect against data loss? (Choose two.)

A. Enable versioning on the S3 bucket.

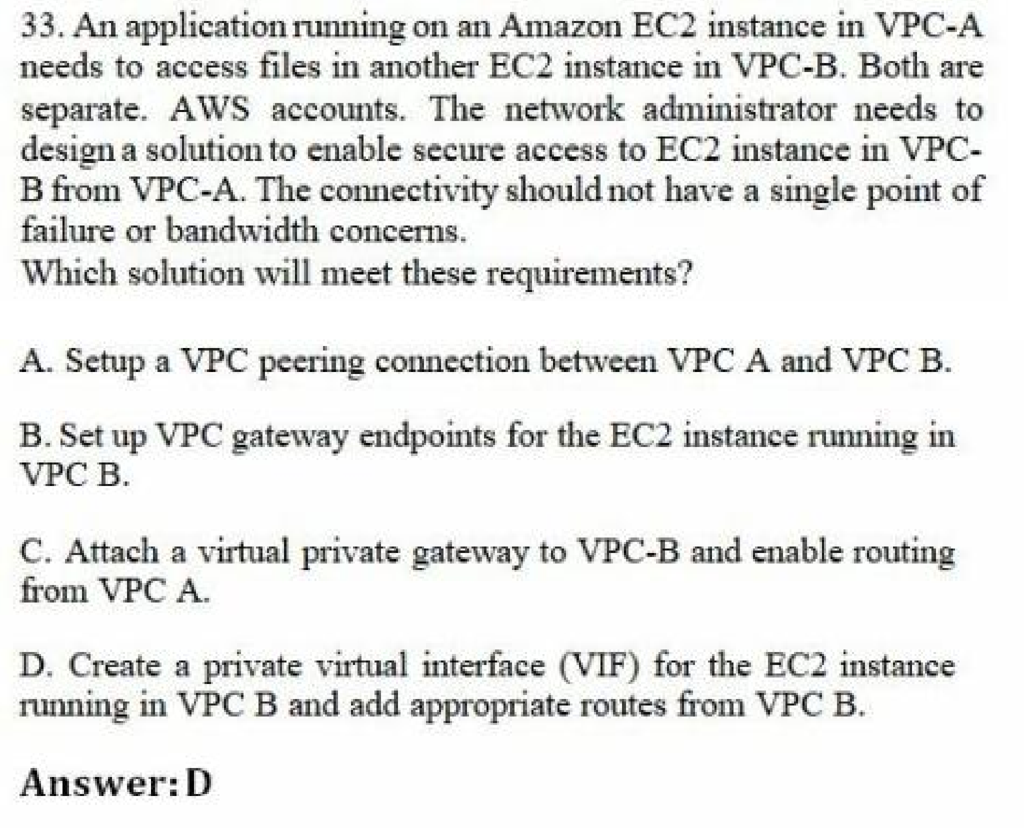
B. Enable access logging on the S3 bucket.

C. Enable server-side encryption on the S3 bucket.

D. Configure an S3 lifecycle rule to transition objects to Amazon S3 Glacier.

E. Use MFA Delete to require multifactor authentication to delete an object.

Answer: A, E



33. An application running on an Amazon EC2 instance in VPC-A needs to access files in another EC2 instance in VPC-B. Both are separate. AWS accounts. The network administrator needs to design a solution to enable secure access to EC2 instance in VPC-B from VPC-A. The connectivity should not have a single point of failure or bandwidth concerns.
Which solution will meet these requirements?

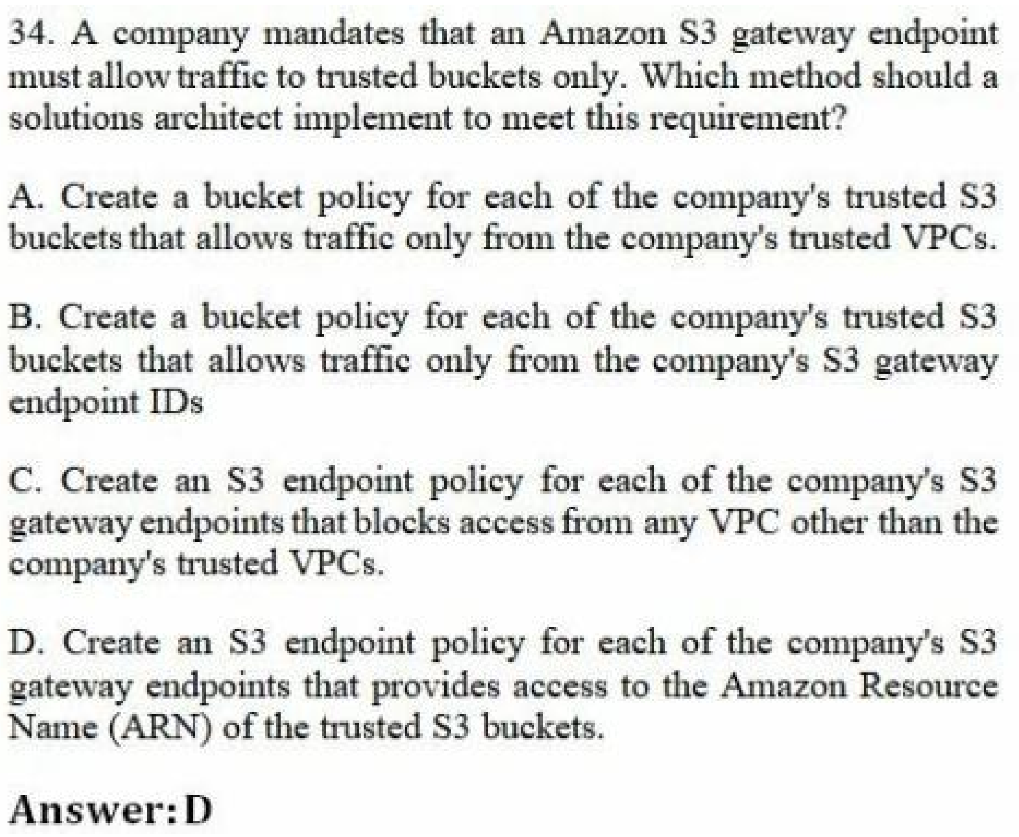
A. Setup a VPC peering connection between VPC A and VPC B.

B. Set up VPC gateway endpoints for the EC2 instance running in VPC B.

C. Attach a virtual private gateway to VPC-B and enable routing from VPC A.

D. Create a private virtual interface (VIF) for the EC2 instance running in VPC B and add appropriate routes from VPC B.

Answer: D



34. A company mandates that an Amazon S3 gateway endpoint must allow traffic to trusted buckets only. Which method should a solutions architect implement to meet this requirement?

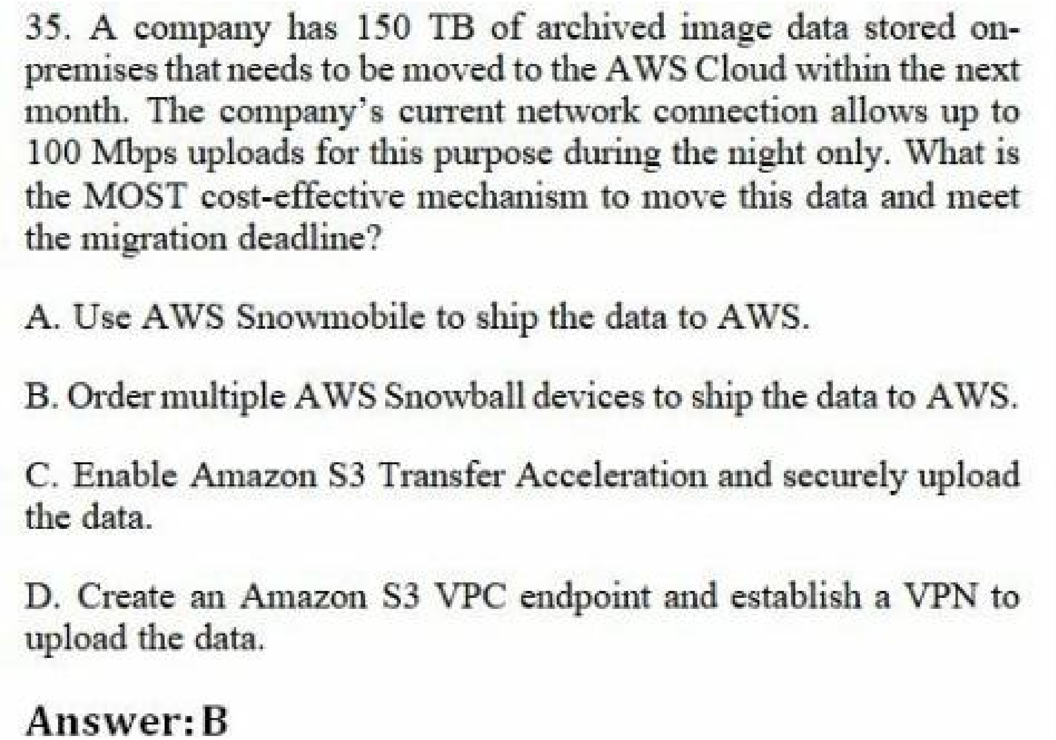
A. Create a bucket policy for each of the company's trusted S3 buckets that allows traffic only from the company's trusted VPCs.

B. Create a bucket policy for each of the company's trusted S3 buckets that allows traffic only from the company's S3 gateway endpoint IDs

C. Create an S3 endpoint policy for each of the company's S3 gateway endpoints that blocks access from any VPC other than the company's trusted VPCs.

D. Create an S3 endpoint policy for each of the company's S3 gateway endpoints that provides access to the Amazon Resource Name (ARN) of the trusted S3 buckets.

Answer: D



35. A company has 150 TB of archived image data stored on-premises that needs to be moved to the AWS Cloud within the next month. The company's current network connection allows up to 100 Mbps uploads for this purpose during the night only. What is the MOST cost-effective mechanism to move this data and meet the migration deadline?

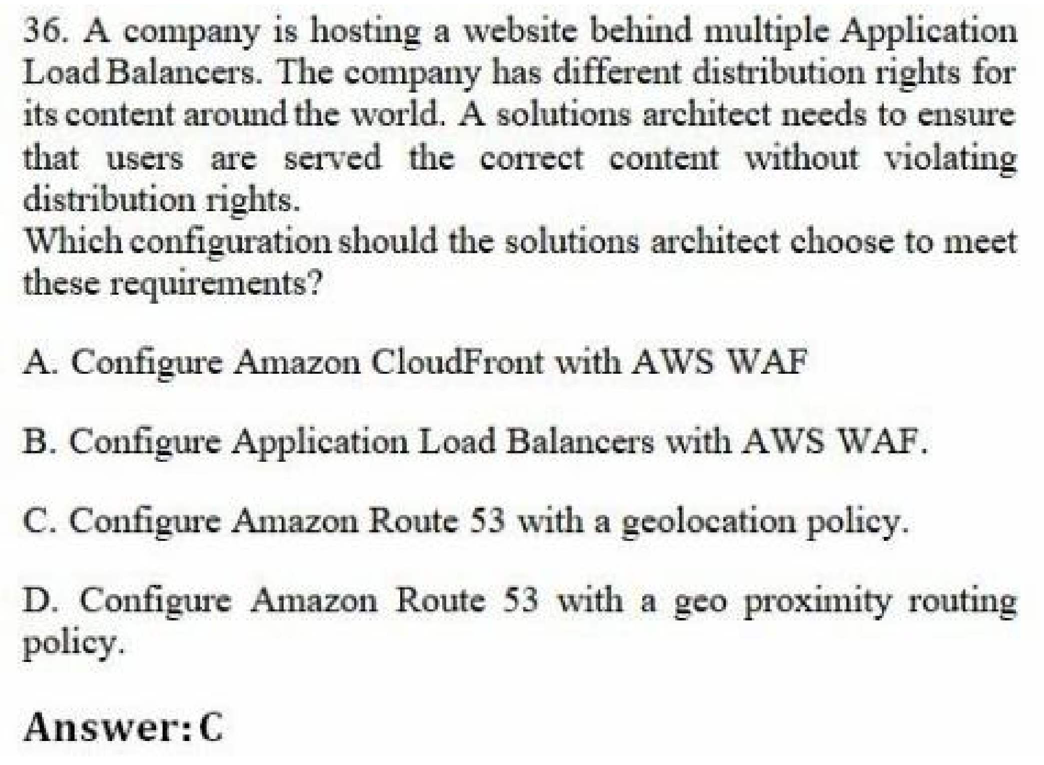
A. Use AWS Snowmobile to ship the data to AWS.

B. Order multiple AWS Snowball devices to ship the data to AWS.

C. Enable Amazon S3 Transfer Acceleration and securely upload the data.

D. Create an Amazon S3 VPC endpoint and establish a VPN to upload the data.

Answer: B



36. A company is hosting a website behind multiple Application Load Balancers. The company has different distribution rights for its content around the world. A solutions architect needs to ensure that users are served the correct content without violating distribution rights.
Which configuration should the solutions architect choose to meet these requirements?

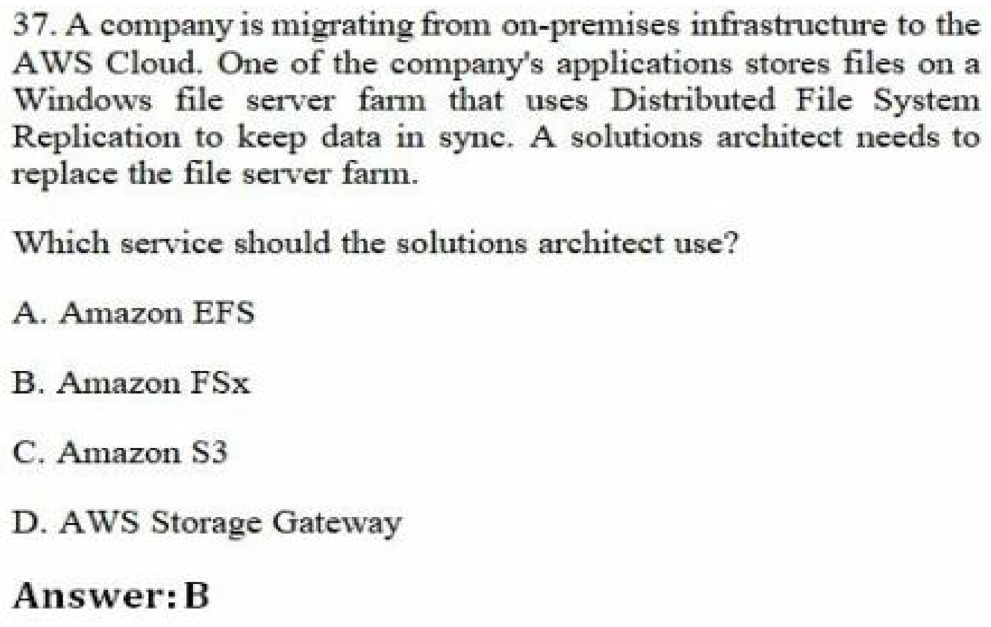
A. Configure Amazon CloudFront with AWS WAF

B. Configure Application Load Balancers with AWS WAF.

C. Configure Amazon Route 53 with a geolocation policy.

D. Configure Amazon Route 53 with a geo proximity routing policy.

Answer: C



37. A company is migrating from on-premises infrastructure to the AWS Cloud. One of the company's applications stores files on a Windows file server farm that uses Distributed File System Replication to keep data in sync. A solutions architect needs to replace the file server farm.

Which service should the solutions architect use?

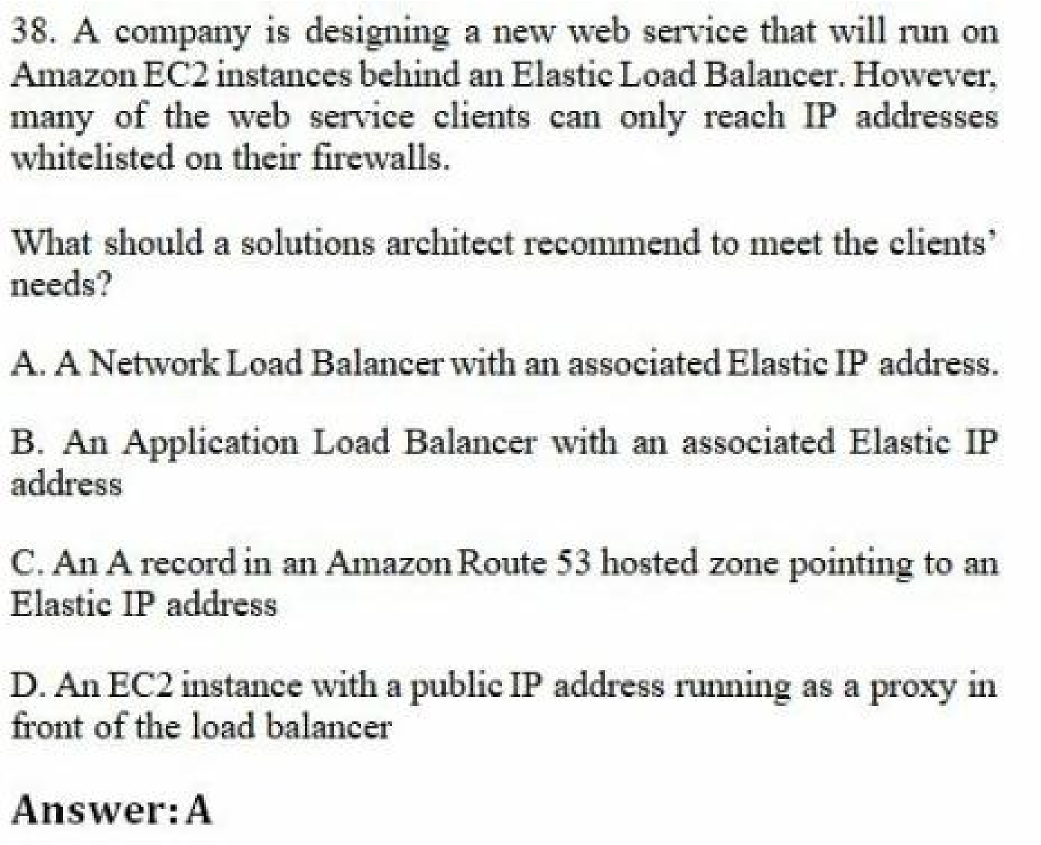
A. Amazon EFS

B. Amazon FSx

C. Amazon S3

D. AWS Storage Gateway

Answer: B



38. A company is designing a new web service that will run on Amazon EC2 instances behind an Elastic Load Balancer. However, many of the web service clients can only reach IP addresses whitelisted on their firewalls.

What should a solutions architect recommend to meet the clients' needs?

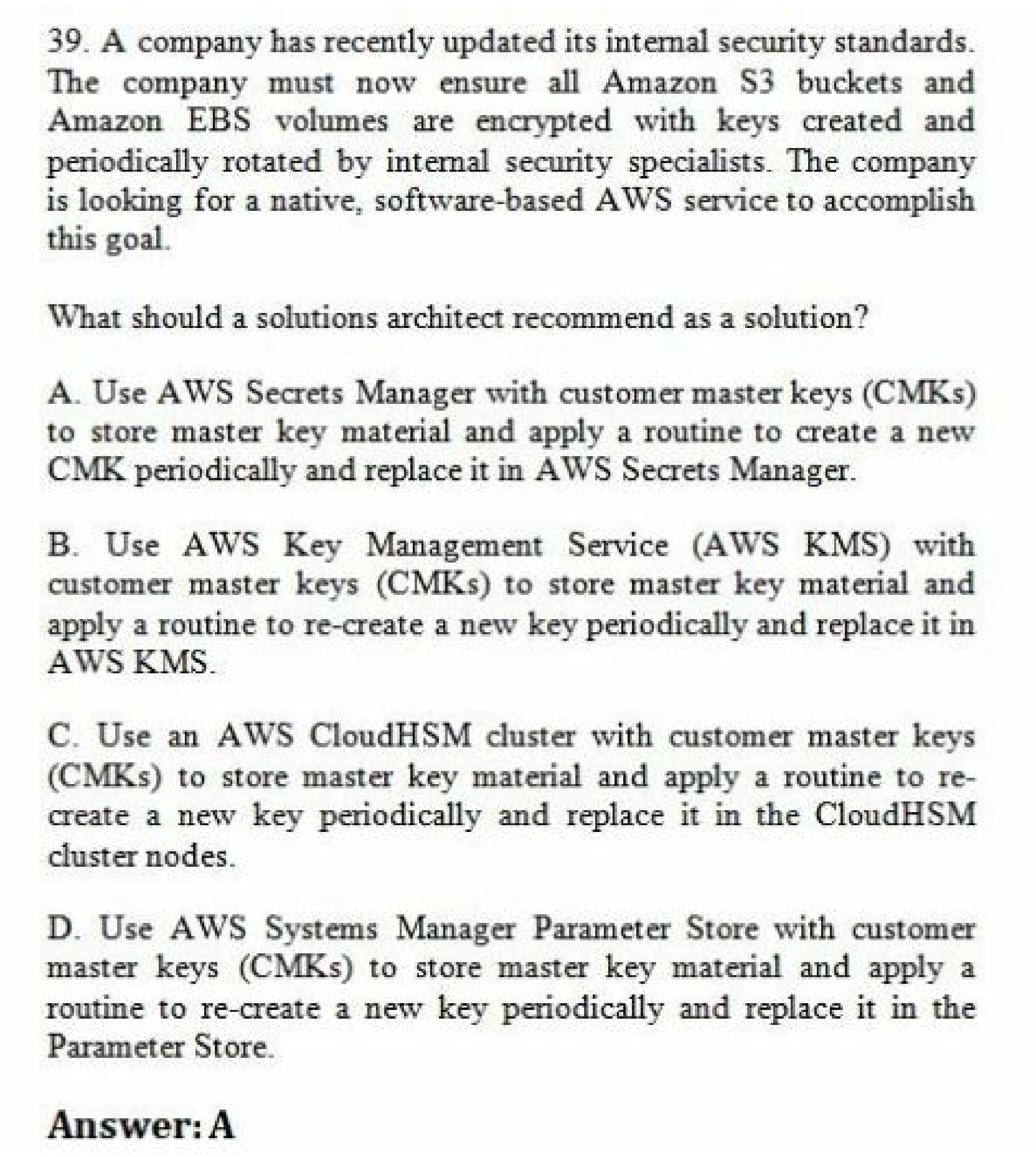
A. A Network Load Balancer with an associated Elastic IP address.

B. An Application Load Balancer with an associated Elastic IP address

C. An A record in an Amazon Route 53 hosted zone pointing to an Elastic IP address

D. An EC2 instance with a public IP address running as a proxy in front of the load balancer

Answer: A



39. A company has recently updated its internal security standards. The company must now ensure all Amazon S3 buckets and Amazon EBS volumes are encrypted with keys created and periodically rotated by internal security specialists. The company is looking for a native, software-based AWS service to accomplish this goal.

What should a solutions architect recommend as a solution?

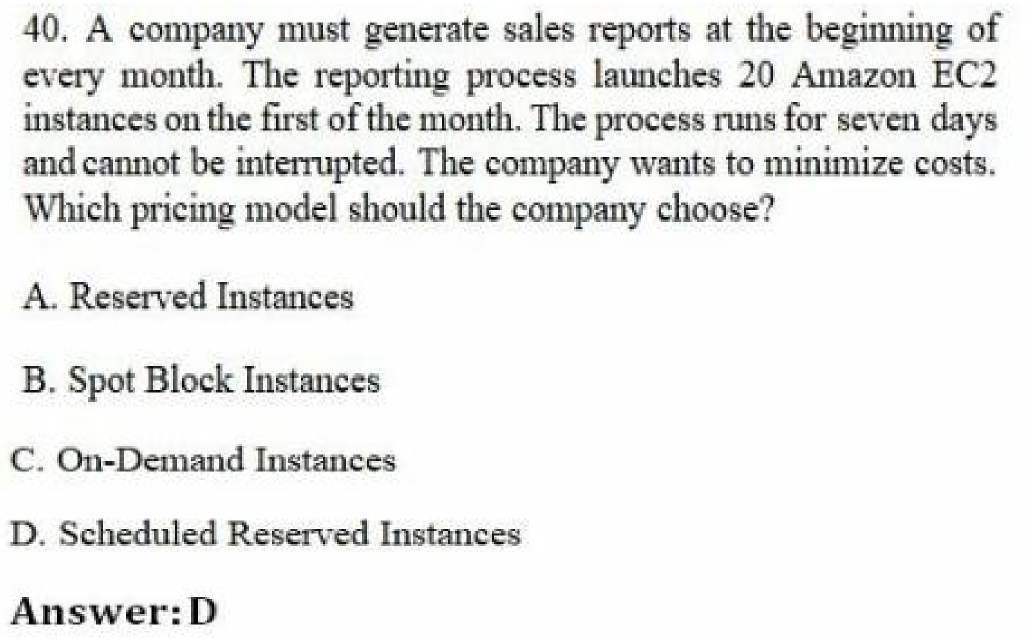
A. Use AWS Secrets Manager with customer master keys (CMKs) to store master key material and apply a routine to create a new CMK periodically and replace it in AWS Secrets Manager.

B. Use AWS Key Management Service (AWS KMS) with customer master keys (CMKs) to store master key material and apply a routine to re-create a new key periodically and replace it in AWS KMS.

C. Use an AWS CloudHSM cluster with customer master keys (CMKs) to store master key material and apply a routine to re-create a new key periodically and replace it in the CloudHSM cluster nodes.

D. Use AWS Systems Manager Parameter Store with customer master keys (CMKs) to store master key material and apply a routine to re-create a new key periodically and replace it in the Parameter Store.

Answer: A



40. A company must generate sales reports at the beginning of every month. The reporting process launches 20 Amazon EC2 instances on the first of the month. The process runs for seven days and cannot be interrupted. The company wants to minimize costs. Which pricing model should the company choose?

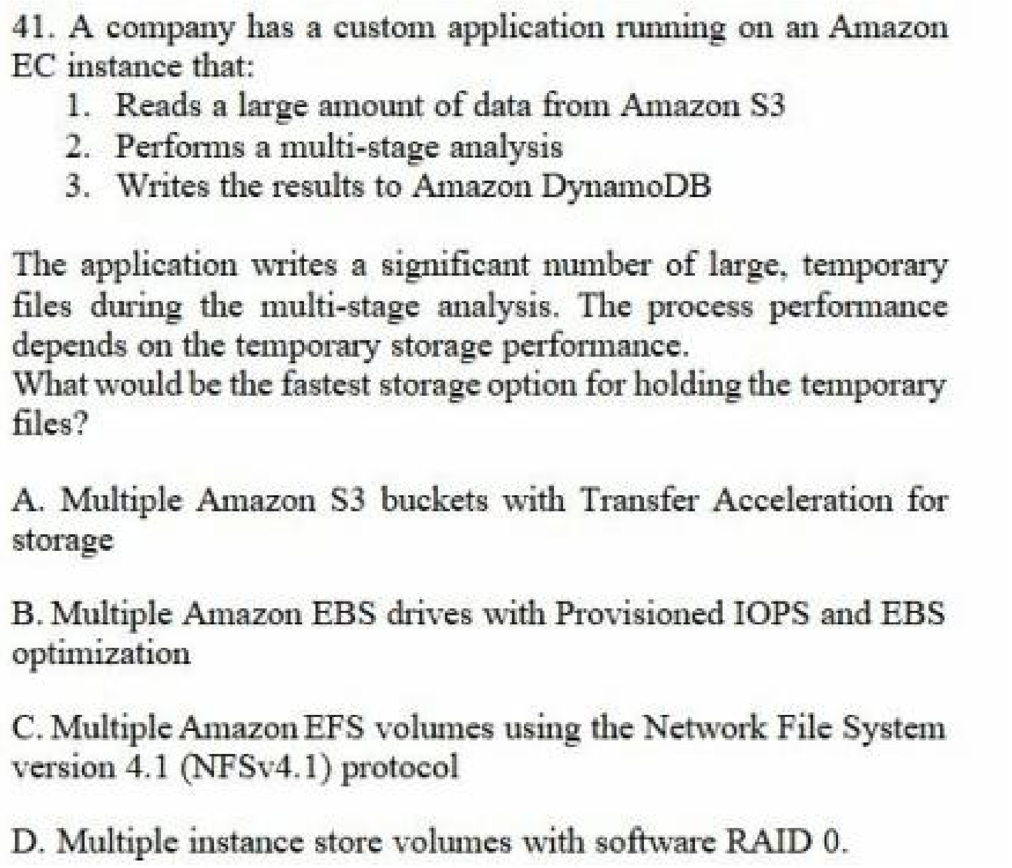
A. Reserved Instances

B. Spot Block Instances

C. On-Demand Instances

D. Scheduled Reserved Instances

Answer: D



41. A company has a custom application running on an Amazon EC instance that:

1. Reads a large amount of data from Amazon S3
2. Performs a multi-stage analysis
3. Writes the results to Amazon DynamoDB

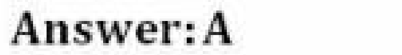
The application writes a significant number of large, temporary files during the multi-stage analysis. The process performance depends on the temporary storage performance.
What would be the fastest storage option for holding the temporary files?

A. Multiple Amazon S3 buckets with Transfer Acceleration for storage

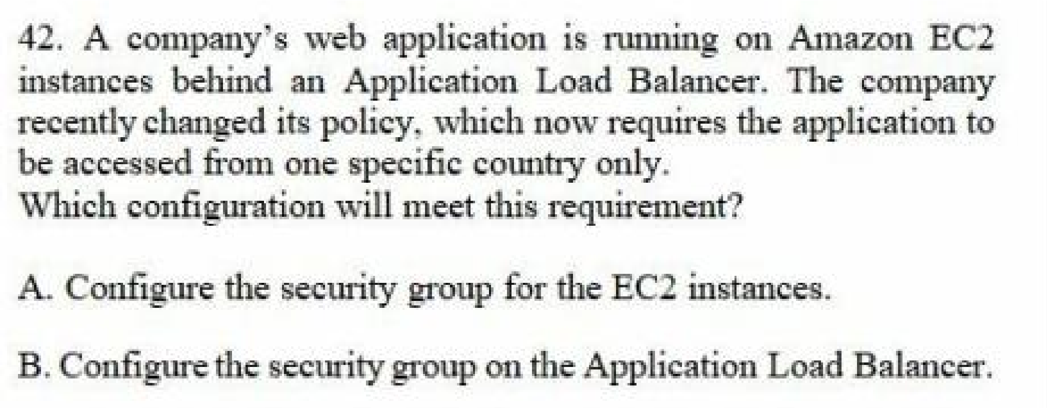
B. Multiple Amazon EBS drives with Provisioned IOPS and EBS optimization

C. Multiple Amazon EFS volumes using the Network File System version 4.1 (NFSv4.1) protocol

D. Multiple instance store volumes with software RAID 0.



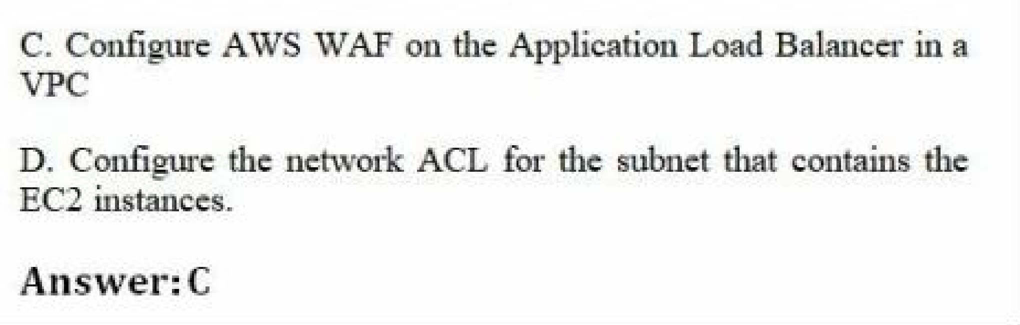
Answer: A



42. A company's web application is running on Amazon EC2 instances behind an Application Load Balancer. The company recently changed its policy, which now requires the application to be accessed from one specific country only.
Which configuration will meet this requirement?

A. Configure the security group for the EC2 instances.

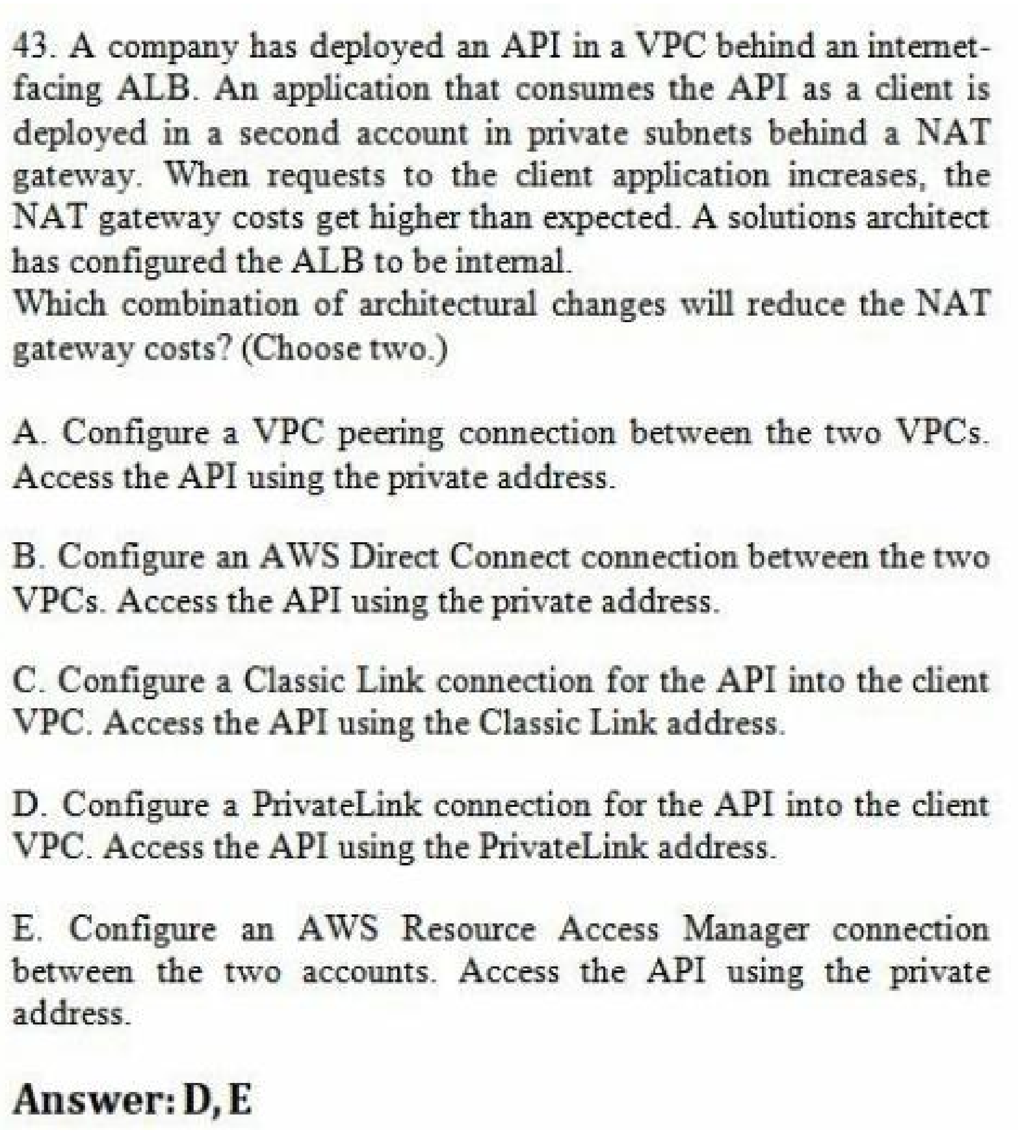
B. Configure the security group on the Application Load Balancer.



C. Configure AWS WAF on the Application Load Balancer in a VPC

D. Configure the network ACL for the subnet that contains the EC2 instances.

Answer: C



43. A company has deployed an API in a VPC behind an internet-facing ALB. An application that consumes the API as a client is deployed in a second account in private subnets behind a NAT gateway. When requests to the client application increases, the NAT gateway costs get higher than expected. A solutions architect has configured the ALB to be internal.
Which combination of architectural changes will reduce the NAT gateway costs? (Choose two.)

A. Configure a VPC peering connection between the two VPCs. Access the API using the private address.

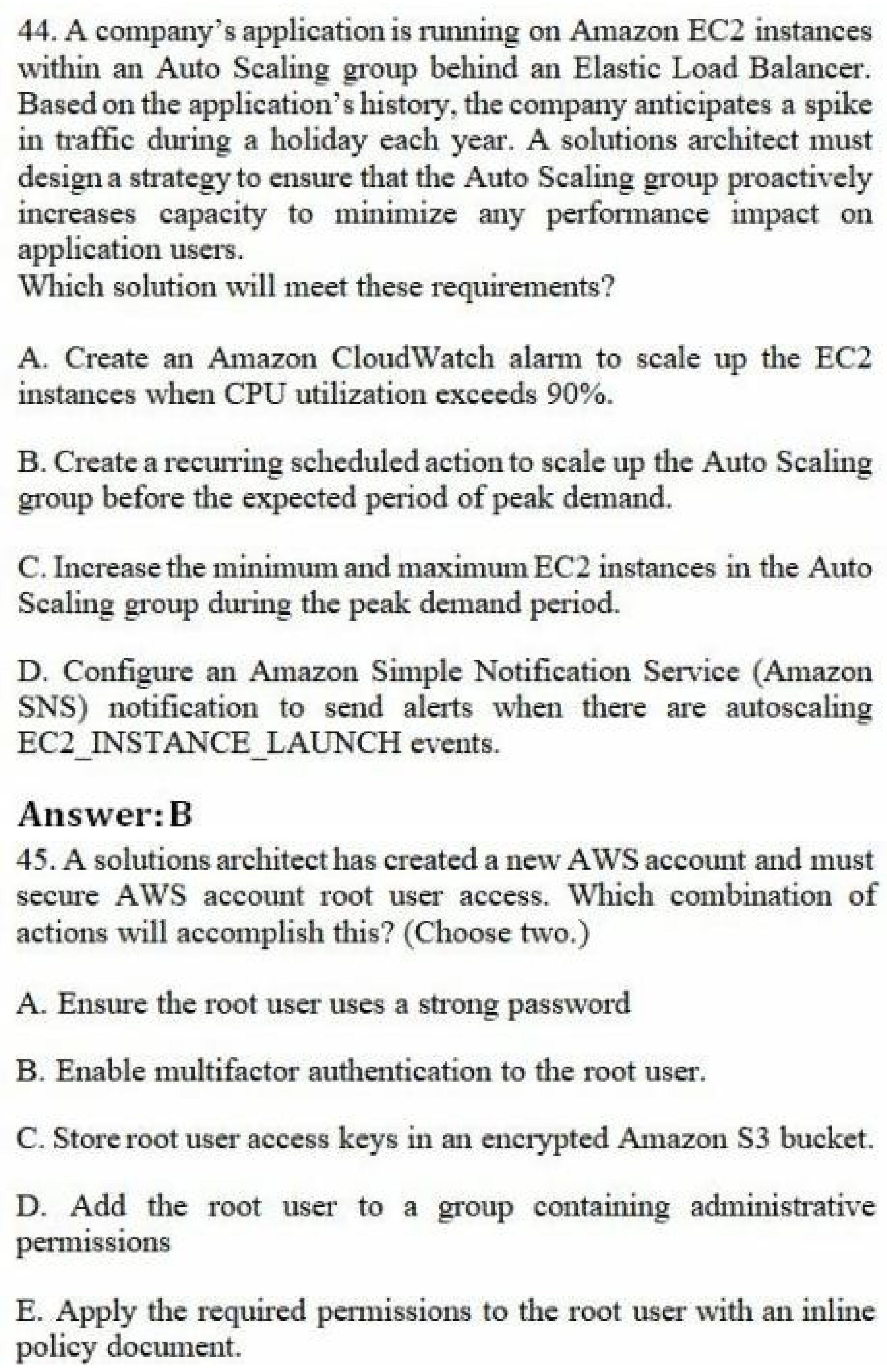
B. Configure an AWS Direct Connect connection between the two VPCs. Access the API using the private address.

C. Configure a Classic Link connection for the API into the client VPC. Access the API using the Classic Link address.

D. Configure a PrivateLink connection for the API into the client VPC. Access the API using the PrivateLink address.

E. Configure an AWS Resource Access Manager connection between the two accounts. Access the API using the private address.

Answer: D, E



44. A company's application is running on Amazon EC2 instances within an Auto Scaling group behind an Elastic Load Balancer. Based on the application's history, the company anticipates a spike in traffic during a holiday each year. A solutions architect must design a strategy to ensure that the Auto Scaling group proactively increases capacity to minimize any performance impact on application users.
Which solution will meet these requirements?

A. Create an Amazon CloudWatch alarm to scale up the EC2 instances when CPU utilization exceeds 90%.

B. Create a recurring scheduled action to scale up the Auto Scaling group before the expected period of peak demand.

C. Increase the minimum and maximum EC2 instances in the Auto Scaling group during the peak demand period.

D. Configure an Amazon Simple Notification Service (Amazon SNS) notification to send alerts when there are autoscaling EC2_INSTANCE_LAUNCH events.

Answer: B

45. A solutions architect has created a new AWS account and must secure AWS account root user access. Which combination of actions will accomplish this? (Choose two.)

A. Ensure the root user uses a strong password

B. Enable multifactor authentication to the root user.

C. Store root user access keys in an encrypted Amazon S3 bucket.

D. Add the root user to a group containing administrative permissions

E. Apply the required permissions to the root user with an inline policy document.

Answer: A, B

46. A company that develops web applications has launched hundreds of Application Load Balancers (ALBs) in multiple Regions. The company wants to create an allow list (or the IPs of all the load balancers on its firewall device. A solutions architect is looking for a one-time, highly available solution to address this request, which will also help reduce the number of IPs that need to be allowed by the firewall.
What should the solutions architect recommend to meet these requirements?

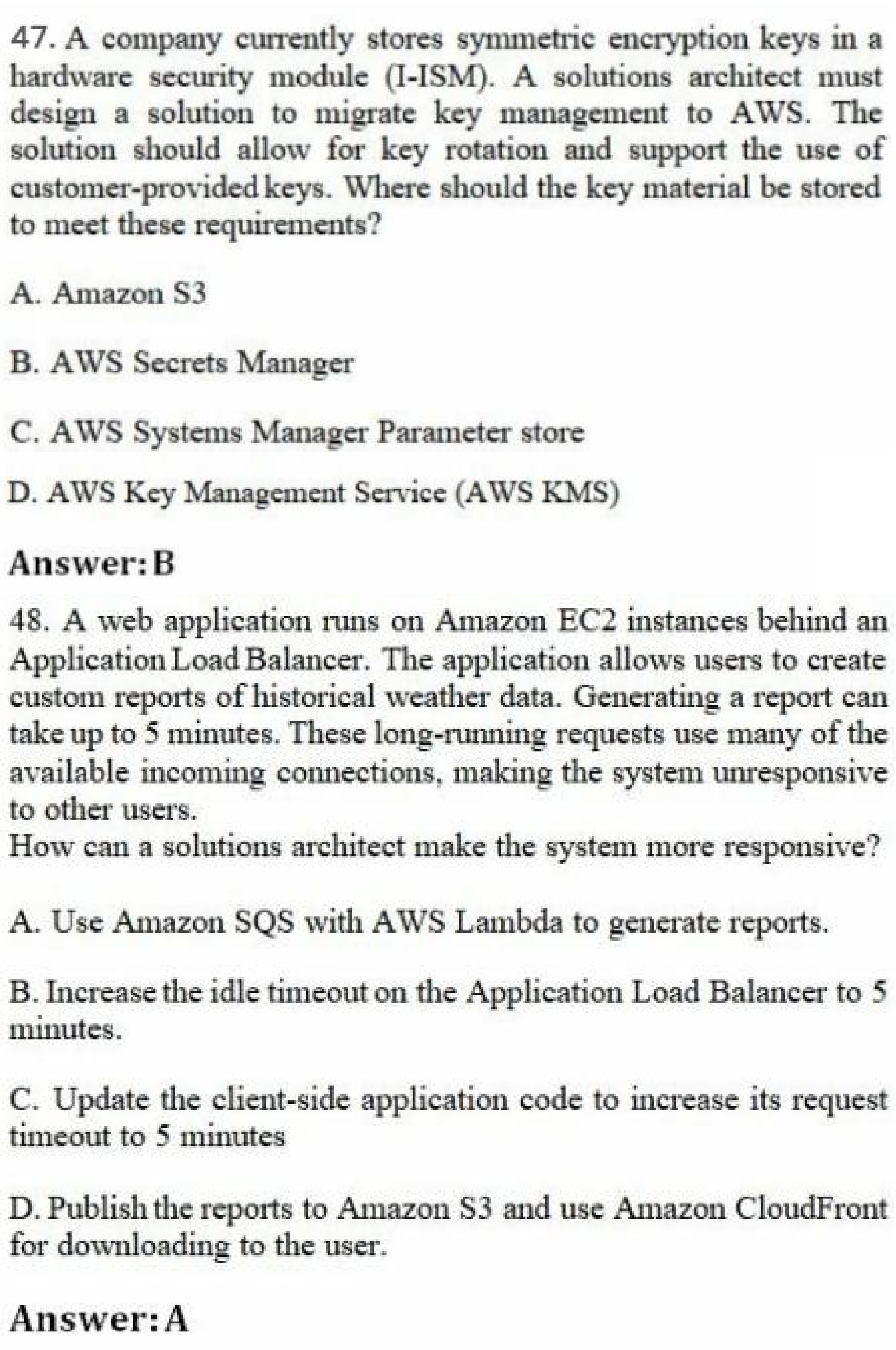
A. Create an AWS Lambda function to keep track of the IPs for all the ALBs in different Regions. Keep refreshing this list.

D. Setup a Network Load Balancer (NLB) with Elastic IPs. Register the private IPs of all the ALBs as targets to this NLB

C. Launch AWS Global Accelerator and create endpoints for all the Regions. Register all the ALBs in different Regions to the corresponding endpoints

D. Setup an Amazon EC2 instance, assign an Elastic IP to this EC2 instance and configure the instance as a proxy to forward traffic to all the ALBs.

Answer: C



47. A company currently stores symmetric encryption keys in a hardware security module (I-ISM). A solutions architect must design a solution to migrate key management to AWS. The solution should allow for key rotation and support the use of customer-provided keys. Where should the key material be stored to meet these requirements?

A. Amazon S3

B. AWS Secrets Manager

C. AWS Systems Manager Parameter store

D. AWS Key Management Service (AWS KMS)

Answer:B

48. A web application runs on Amazon EC2 instances behind an Application Load Balancer. The application allows users to create custom reports of historical weather data. Generating a report can take up to 5 minutes. These long-running requests use many of the available incoming connections, making the system unresponsive to other users.
How can a solutions architect make the system more responsive?

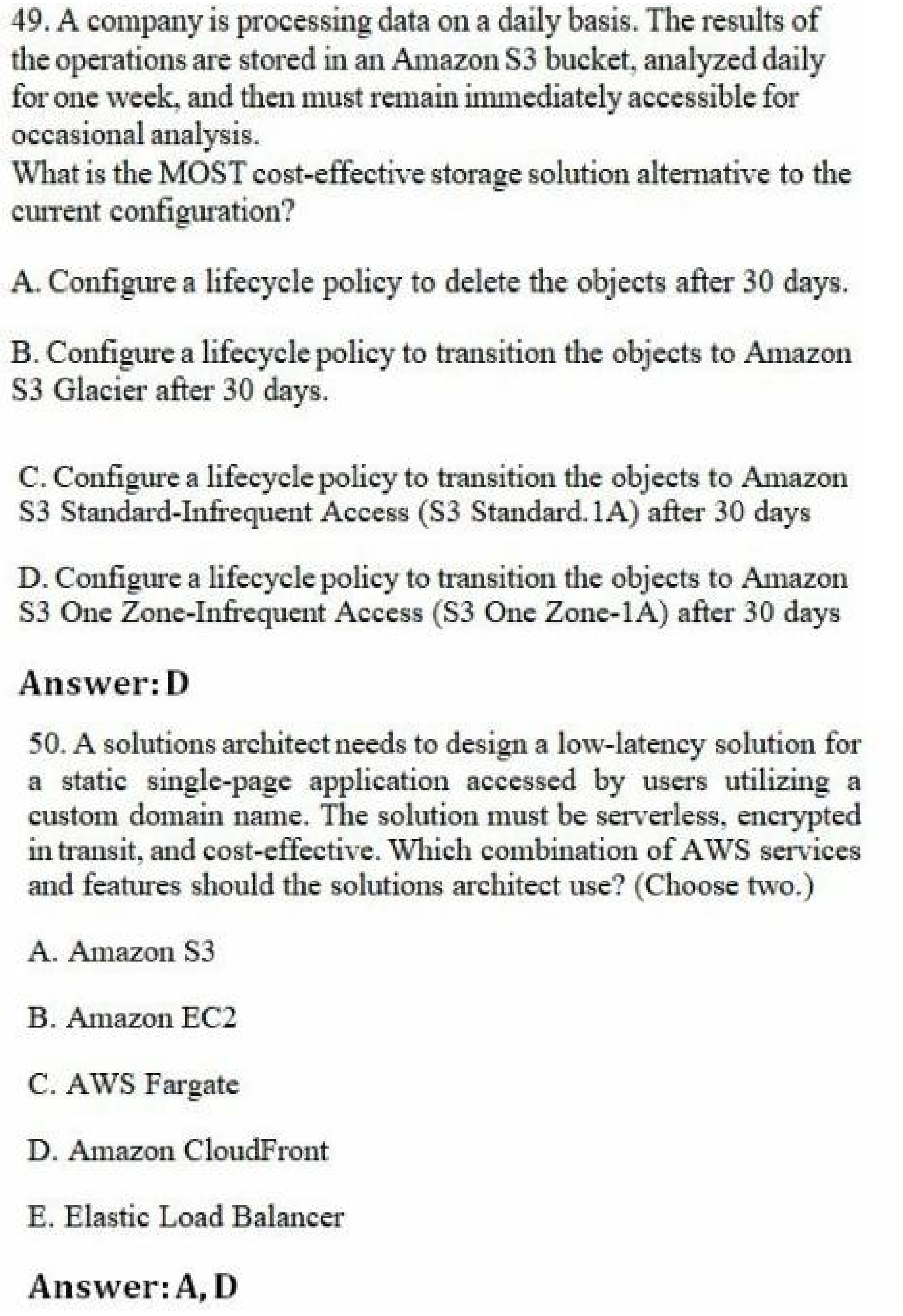
A. Use Amazon SQS with AWS Lambda to generate reports.

B. Increase the idle timeout on the Application Load Balancer to 5 minutes.

C. Update the client-side application code to increase its request timeout to 5 minutes

D. Publish the reports to Amazon S3 and use Amazon CloudFront for downloading to the user.

Answer:A



49. A company is processing data on a daily basis. The results of the operations are stored in an Amazon S3 bucket, analyzed daily for one week, and then must remain immediately accessible for occasional analysis.
What is the MOST cost-effective storage solution alternative to the current configuration?

A. Configure a lifecycle policy to delete the objects after 30 days.

B. Configure a lifecycle policy to transition the objects to Amazon S3 Glacier after 30 days.

C. Configure a lifecycle policy to transition the objects to Amazon S3 Standard-Infrequent Access (S3 Standard.1A) after 30 days

D. Configure a lifecycle policy to transition the objects to Amazon S3 One Zone-Infrequent Access (S3 One Zone-1A) after 30 days

Answer: D

50. A solutions architect needs to design a low-latency solution for a static single-page application accessed by users utilizing a custom domain name. The solution must be serverless, encrypted in transit, and cost-effective. Which combination of AWS services and features should the solutions architect use? (Choose two.)

A. Amazon S3

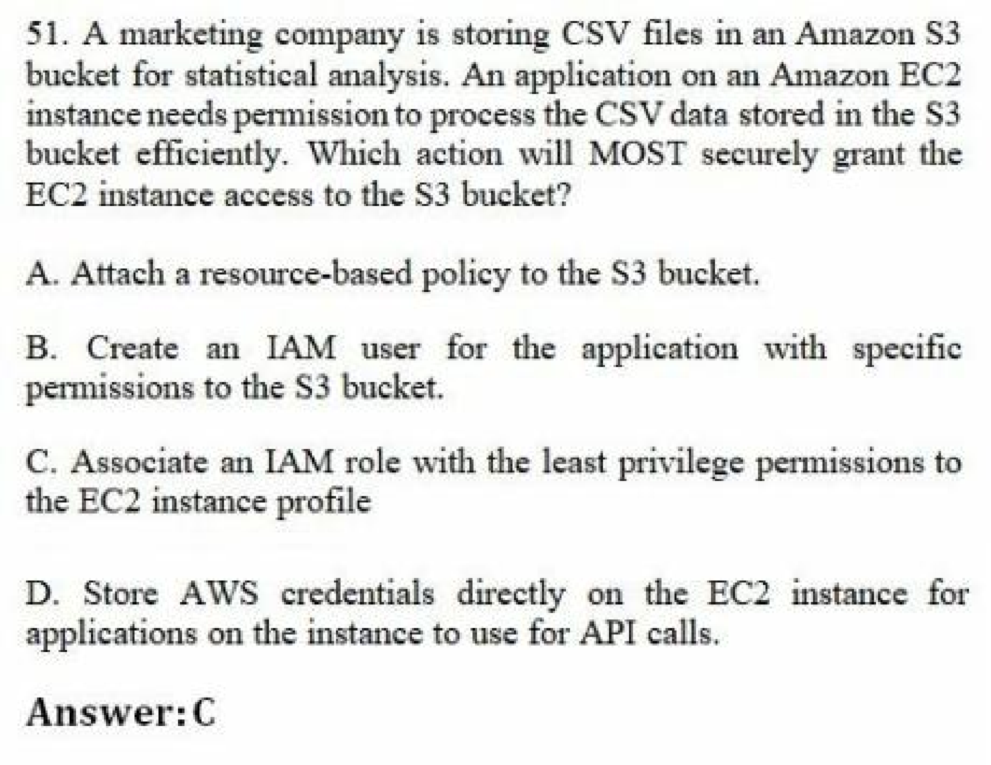
B. Amazon EC2

C. AWS Fargate

D. Amazon CloudFront

E. Elastic Load Balancer

Answer: A, D



51. A marketing company is storing CSV files in an Amazon S3 bucket for statistical analysis. An application on an Amazon EC2 instance needs permission to process the CSV data stored in the S3 bucket efficiently. Which action will MOST securely grant the EC2 instance access to the S3 bucket?

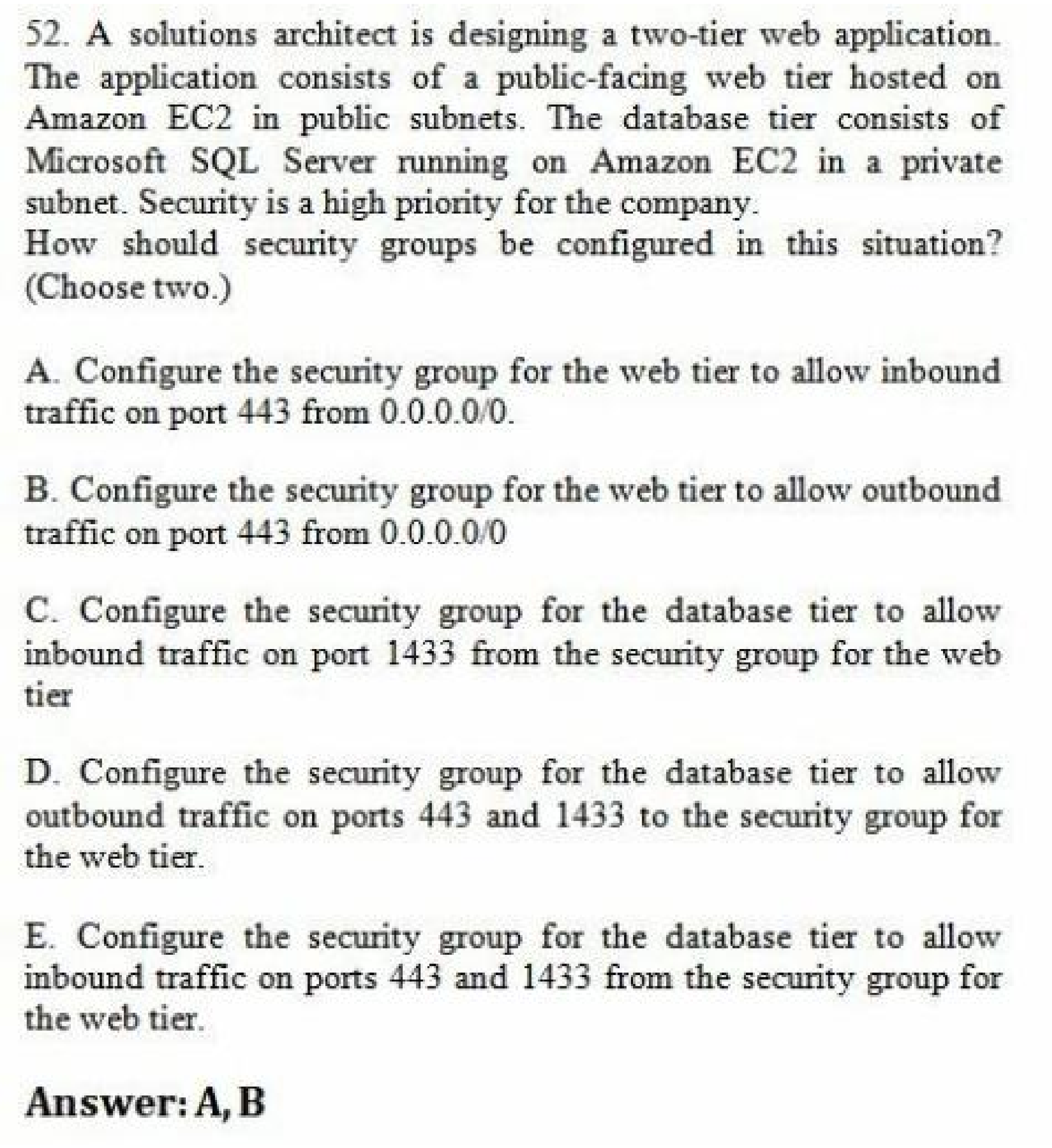
A. Attach a resource-based policy to the S3 bucket.

B. Create an IAM user for the application with specific permissions to the S3 bucket.

C. Associate an IAM role with the least privilege permissions to the EC2 instance profile

D. Store AWS credentials directly on the EC2 instance for applications on the instance to use for API calls.

Answer:C



52. A solutions architect is designing a two-tier web application. The application consists of a public-facing web tier hosted on Amazon EC2 in public subnets. The database tier consists of Microsoft SQL Server running on Amazon EC2 in a private subnet. Security is a high priority for the company.
How should security groups be configured in this situation? (Choose two.)

A. Configure the security group for the web tier to allow inbound traffic on port 443 from 0.0.0.0/0.

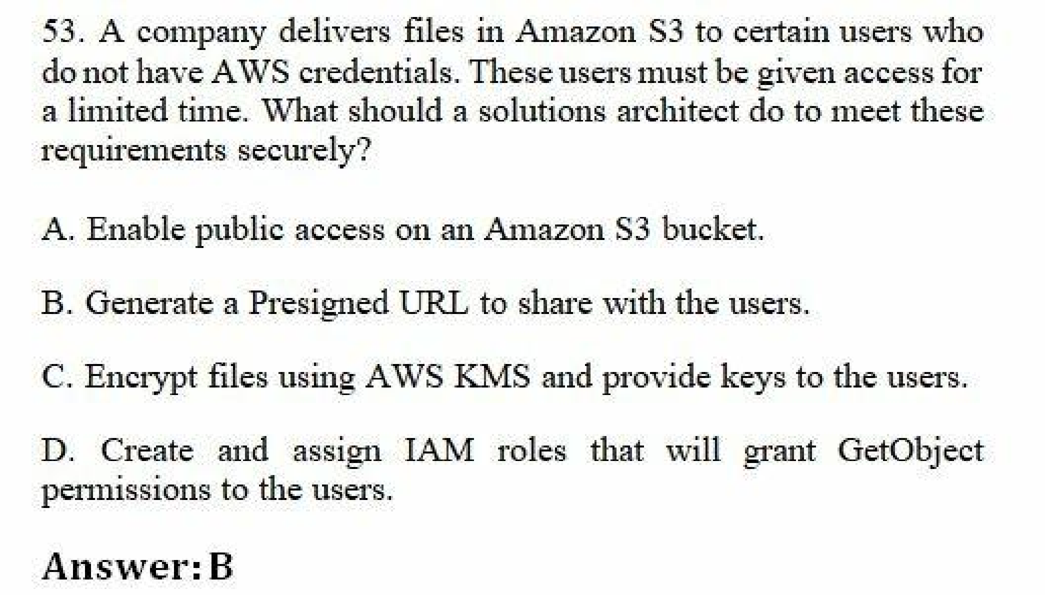
B. Configure the security group for the web tier to allow outbound traffic on port 443 from 0.0.0.0/0

C. Configure the security group for the database tier to allow inbound traffic on port 1433 from the security group for the web tier

D. Configure the security group for the database tier to allow outbound traffic on ports 443 and 1433 to the security group for the web tier.

E. Configure the security group for the database tier to allow inbound traffic on ports 443 and 1433 from the security group for the web tier.

Answer: A, B



53. A company delivers files in Amazon S3 to certain users who do not have AWS credentials. These users must be given access for a limited time. What should a solutions architect do to meet these requirements securely?

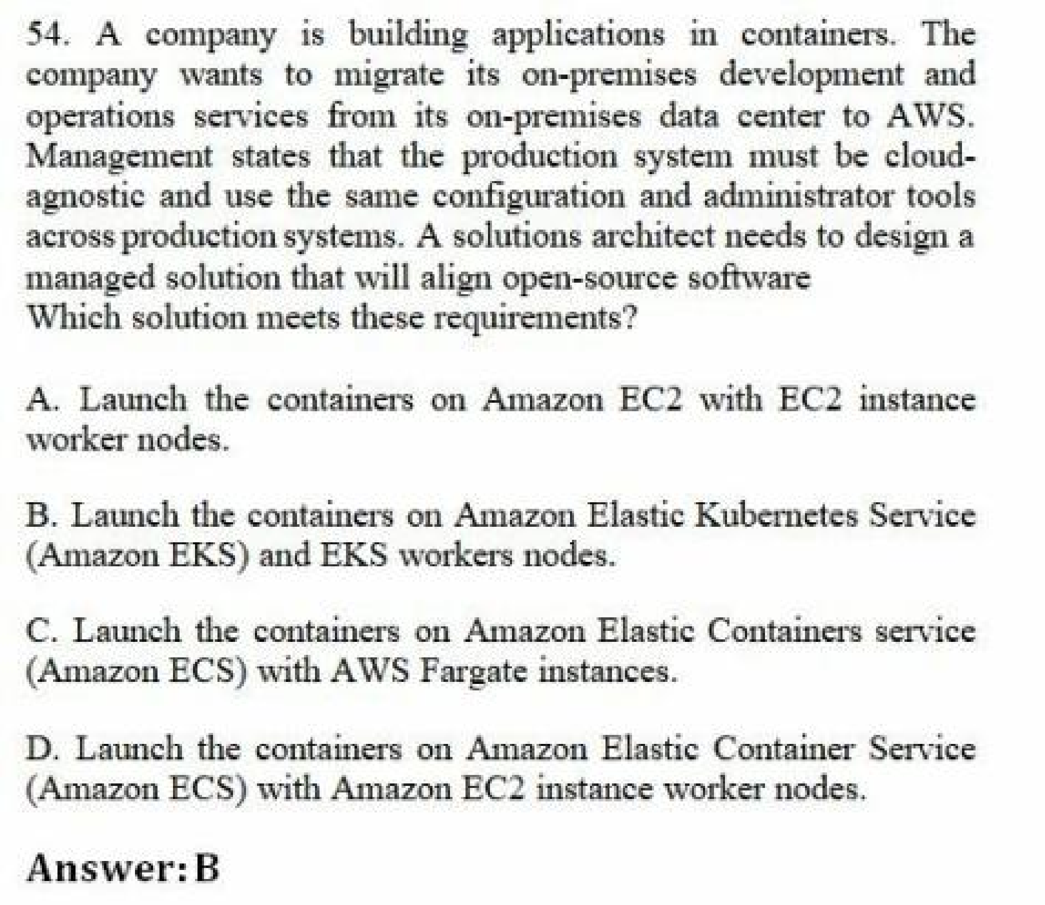
A. Enable public access on an Amazon S3 bucket.

B. Generate a Presigned URL to share with the users.

C. Encrypt files using AWS KMS and provide keys to the users.

D. Create and assign IAM roles that will grant GetObject permissions to the users.

Answer: B



54. A company is building applications in containers. The company wants to migrate its on-premises development and operations services from its on-premises data center to AWS. Management states that the production system must be cloud-agnostic and use the same configuration and administrator tools across production systems. A solutions architect needs to design a managed solution that will align open-source software
Which solution meets these requirements?

A. Launch the containers on Amazon EC2 with EC2 instance worker nodes.

B. Launch the containers on Amazon Elastic Kubernetes Service (Amazon EKS) and EKS workers nodes.

C. Launch the containers on Amazon Elastic Containers service (Amazon ECS) with AWS Fargate instances.

D. Launch the containers on Amazon Elastic Container Service (Amazon ECS) with Amazon EC2 instance worker nodes.

Answer: B

55. A company is planning to use Amazon S3 to store images uploaded by its users. The images must be encrypted at rest in Amazon S3. The company does not want to spend time managing and rotating the keys, but it does want to control who can access those keys.
What should a solutions architect use to accomplish this?

A. Server-Side Encryption with keys stored in an S3 bucket
B. Server-Side Encryption with Customer-Provided Keys (SSE-C)
C. Server-Side Encryption with Amazon S3-Managed Keys (SSE-S3)
D. Server-Side Encryption with AWS KMS-Managed Keys (SSE-KMS)

Answer: D

56. A database is on an Amazon RDS MySQL 5.6 Multi-AZ DB instance that experiences highly dynamic reads. Application developers notice a significant slowdown when testing read performance from a secondary AWS Region. The developers want a solution that provides less than 1 second of read replication latency.
What should the solutions architect recommend?

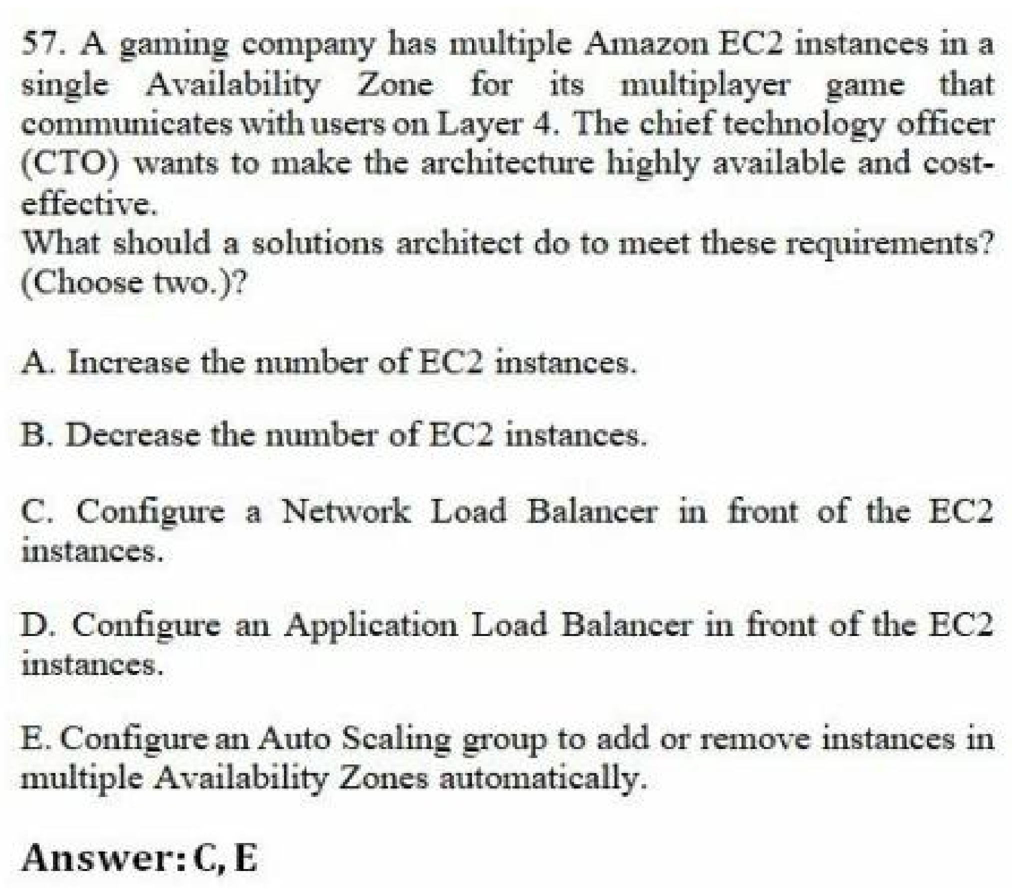
A. Install MySQL on Amazon EC2 in the secondary Region

B. Migrate the database to Amazon Aurora with cross-Region replicas.

C. Create another RDS for MySQL read replica in the secondary Region.

D. Implement Amazon ElastiCache to improve database query performance.

Answer: B



57. A gaming company has multiple Amazon EC2 instances in a single Availability Zone for its multiplayer game that communicates with users on Layer 4. The chief technology officer (CTO) wants to make the architecture highly available and cost-effective.

What should a solutions architect do to meet these requirements? (Choose two.)?

A. Increase the number of EC2 instances.

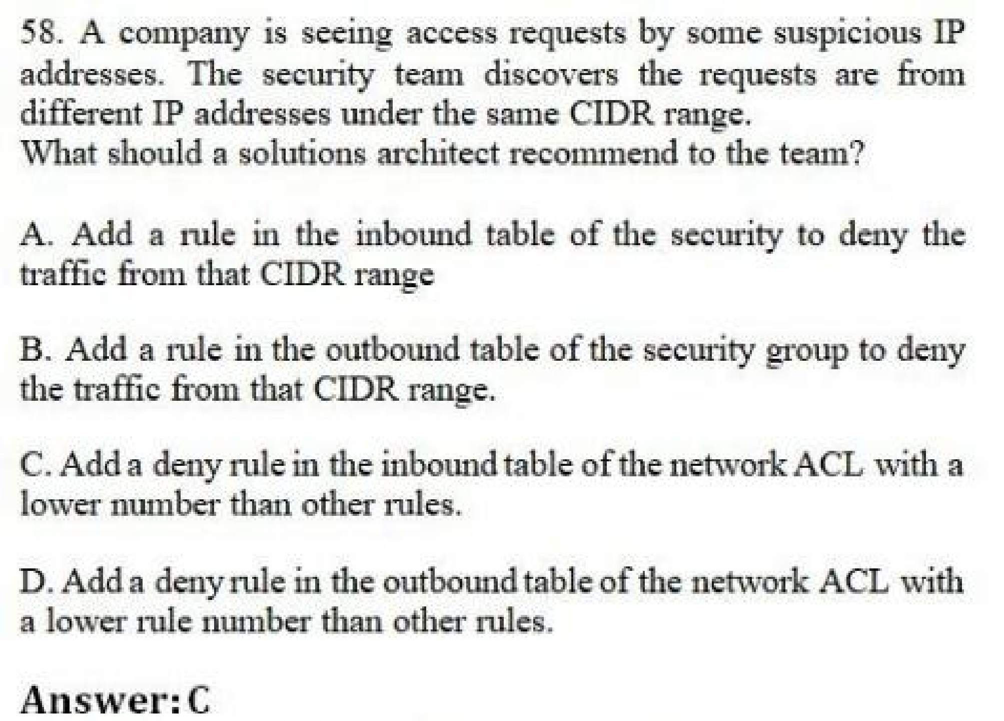
B. Decrease the number of EC2 instances.

C. Configure a Network Load Balancer in front of the EC2 instances.

D. Configure an Application Load Balancer in front of the EC2 instances.

E. Configure an Auto Scaling group to add or remove instances in multiple Availability Zones automatically.

Answer: C, E



58. A company is seeing access requests by some suspicious IP addresses. The security team discovers the requests are from different IP addresses under the same CIDR range.
What should a solutions architect recommend to the team?

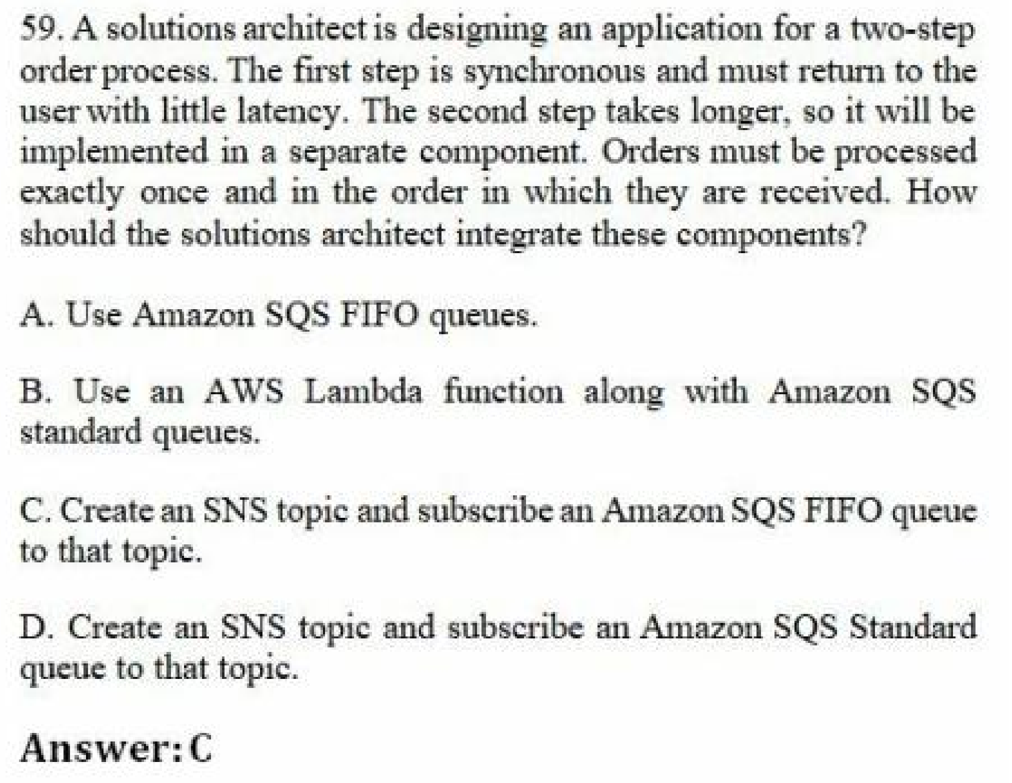
A. Add a rule in the inbound table of the security to deny the traffic from that CIDR range

B. Add a rule in the outbound table of the security group to deny the traffic from that CIDR range.

C. Add a deny rule in the inbound table of the network ACL with a lower number than other rules.

D. Add a deny rule in the outbound table of the network ACL with a lower rule number than other rules.

Answer: C



59. A solutions architect is designing an application for a two-step order process. The first step is synchronous and must return to the user with little latency. The second step takes longer, so it will be implemented in a separate component. Orders must be processed exactly once and in the order in which they are received. How should the solutions architect integrate these components?

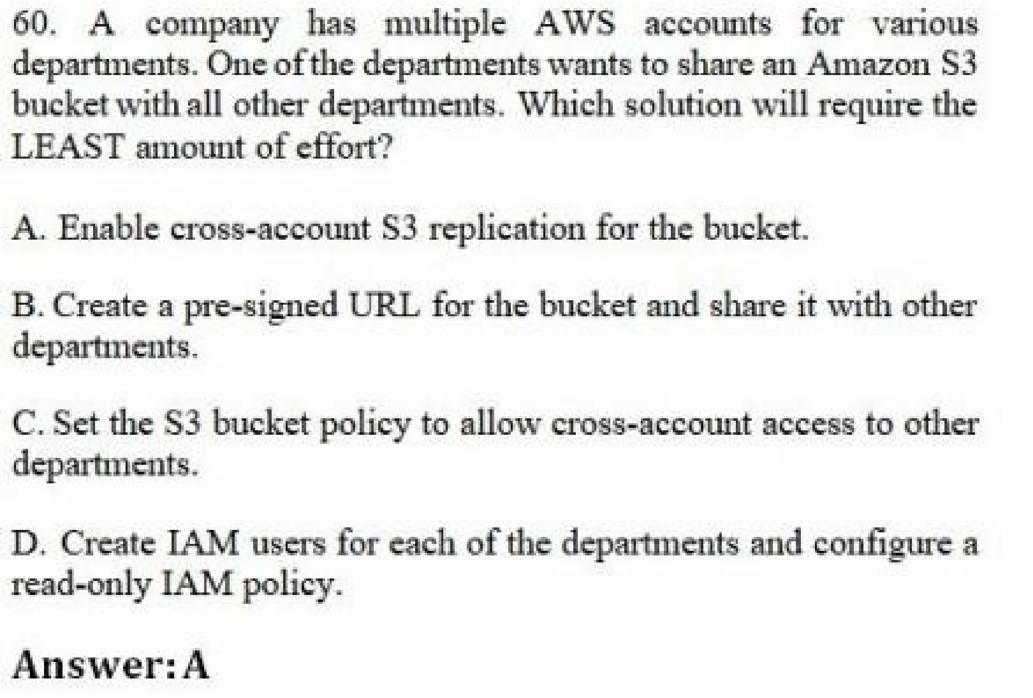
A. Use Amazon SQS FIFO queues.

B. Use an AWS Lambda function along with Amazon SQS standard queues.

C. Create an SNS topic and subscribe an Amazon SQS FIFO queue to that topic.

D. Create an SNS topic and subscribe an Amazon SQS Standard queue to that topic.

Answer: C



60. A company has multiple AWS accounts for various departments. One of the departments wants to share an Amazon S3 bucket with all other departments. Which solution will require the LEAST amount of effort?

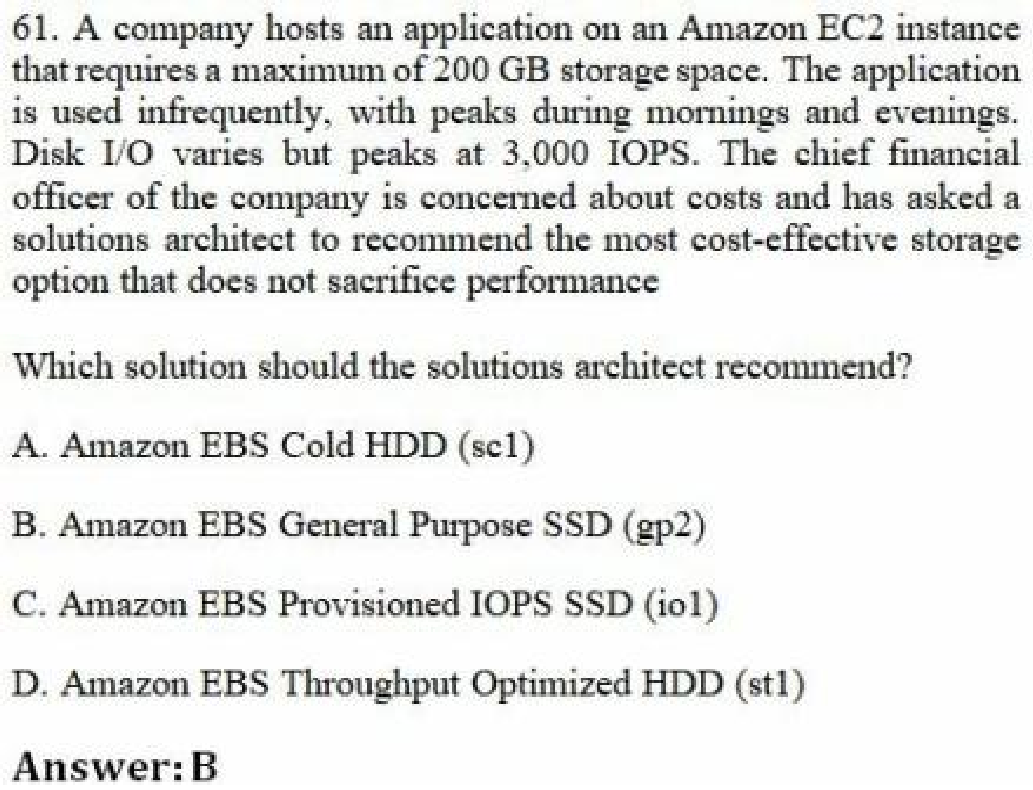
A. Enable cross-account S3 replication for the bucket.

B. Create a pre-signed URL for the bucket and share it with other departments.

C. Set the S3 bucket policy to allow cross-account access to other departments.

D. Create IAM users for each of the departments and configure a read-only IAM policy.

Answer: A



61. A company hosts an application on an Amazon EC2 instance that requires a maximum of 200 GB storage space. The application is used infrequently, with peaks during mornings and evenings. Disk I/O varies but peaks at 3,000 IOPS. The chief financial officer of the company is concerned about costs and has asked a solutions architect to recommend the most cost-effective storage option that does not sacrifice performance

Which solution should the solutions architect recommend?

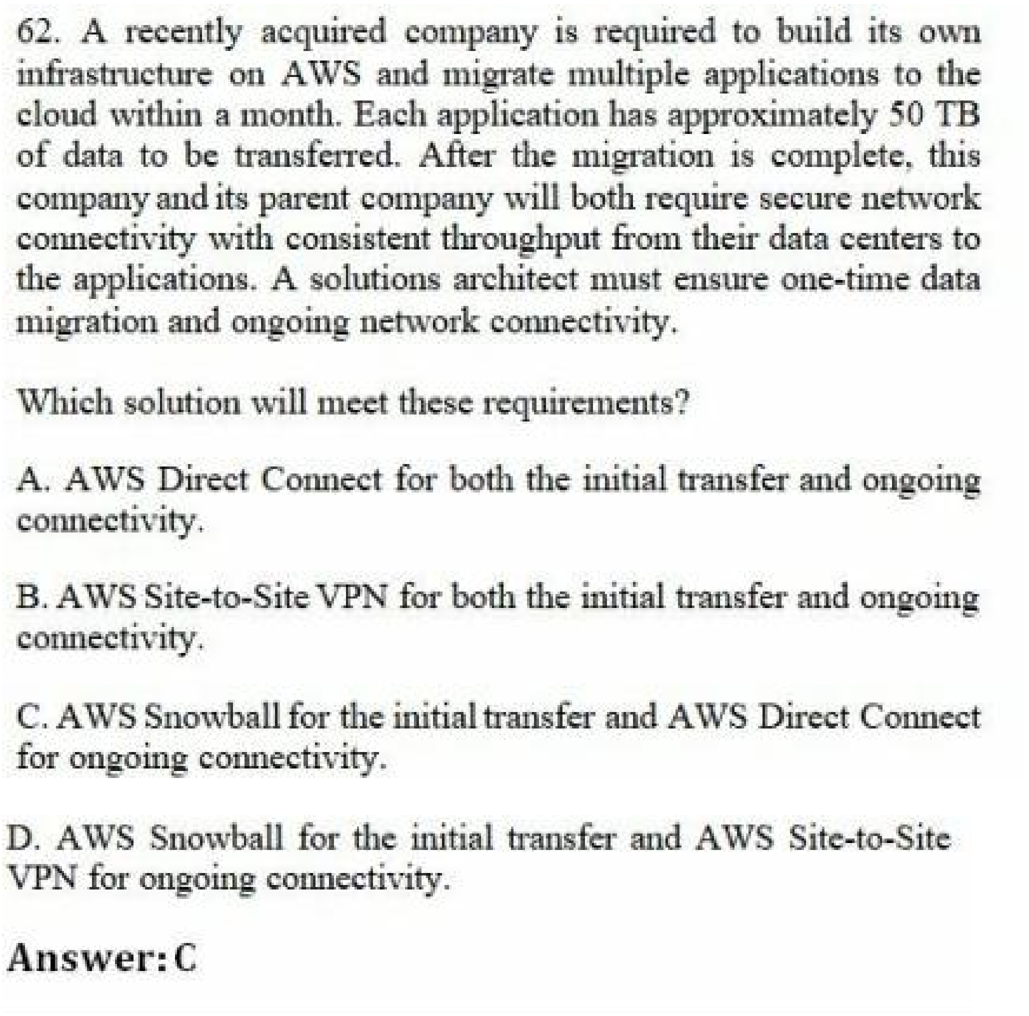
A. Amazon EBS Cold HDD (sc1)

B. Amazon EBS General Purpose SSD (gp2)

C. Amazon EBS Provisioned IOPS SSD (io1)

D. Amazon EBS Throughput Optimized HDD (st1)

Answer: B



62. A recently acquired company is required to build its own infrastructure on AWS and migrate multiple applications to the cloud within a month. Each application has approximately 50 TB of data to be transferred. After the migration is complete, this company and its parent company will both require secure network connectivity with consistent throughput from their data centers to the applications. A solutions architect must ensure one-time data migration and ongoing network connectivity.

Which solution will meet these requirements?

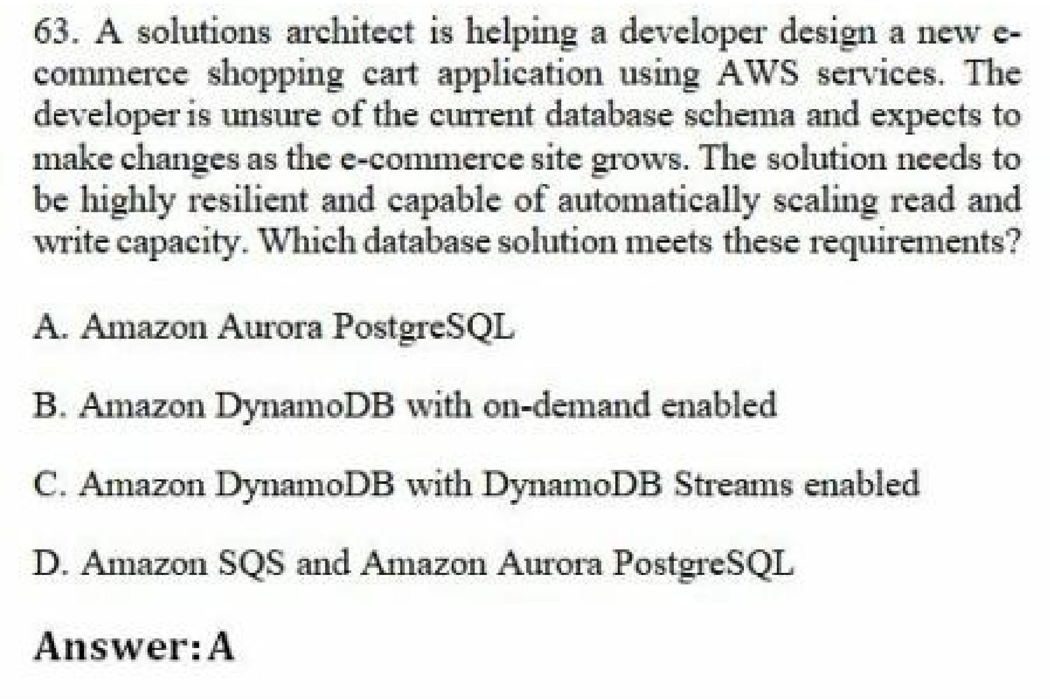
A. AWS Direct Connect for both the initial transfer and ongoing connectivity.

B. AWS Site-to-Site VPN for both the initial transfer and ongoing connectivity.

C. AWS Snowball for the initial transfer and AWS Direct Connect for ongoing connectivity.

D. AWS Snowball for the initial transfer and AWS Site-to-Site VPN for ongoing connectivity.

Answer: C



63. A solutions architect is helping a developer design a new e-commerce shopping cart application using AWS services. The developer is unsure of the current database schema and expects to make changes as the e-commerce site grows. The solution needs to be highly resilient and capable of automatically scaling read and write capacity. Which database solution meets these requirements?

A. Amazon Aurora PostgreSQL

B. Amazon DynamoDB with on-demand enabled

C. Amazon DynamoDB with DynamoDB Streams enabled

D. Amazon SQS and Amazon Aurora PostgreSQL

Answer:A