Solutions Architect Associate exam

Global infrastructure:

26 Launched Regions	84 Availability Zones	410+ Edge Locations

Availability Zones - Distinct locations from within an AWS region that are engineered to be isolated from failures.
A region - is a geographical area divided into Availability Zones. Each region contains at least two Availability Zones
Each AWS Region consists of multiple, isolated, and physically separate Availability Zones within a geographic area.

Identity and Access Management

- New users have no permissions when first created.
- New users are assigned an access key ID and secret access key when first created.
- User access key ID and secret access key are not the same as the password
- Go into CloudWatch, and you create a billing alarm - I can get automatic notifications if my account goes over, like $1,000, or whatever.

S3 (Simple Storage Service)

- S3 - это универсальное пространство имен. Вмена должны быть уникальными во всем мире.
    - [ ] Вопрос со сценарием. В нем рассматривается загрузка файла на S3, и он спрашивает вас, какой код вы получите обратно? Это будет код HTTP 200, который вернется в ваш браузер.
- S3 основан на объектах. И просто думайте об объектах как о файлах. Ключ - имя файла. Значение - его содержимое (данные файла)!

- **Version ID** - у нас есть идентификатор версии, это важно для определения версий.
- **Metadata** - У нас есть метаданные, которые представляют собой данные о данных, которые вы храните.
- **Subresources** - у нас есть списки контроля доступа и торренты. И этот список контроля доступа в основном это разрешения для данного объекта (**Torrents**). И вы можете заблокировать каждый объект по отдельности. Так что вы можете сделать это на уровне ведра, а также на уровне объекта (**Access Control List**).
- **Read after Write consistency for PUTS of new Objects** - как только вы создадите новый объект, вы сможете сразу же прочитать этот объект.
- **Eventual Consistency for overwrite PUTS and DELETES** (can take some time to propagate) - Но если вы обновляете объект или удаляете объект, и попытаетесь прочитать его немедленно, вы получите только конечную согласованность. Таким образом, вы можете получить автоматический объект, или вы можете увидеть удаленный файл. Но если вы подождете около секунды, все будет согласовано. Таким образом, вы получаете конечную согласованность.

S3 has the following features:

1. Tiered Storage Available
2. Lifecycle Management
3. Versioning
4. Encryption
5. MFA Delete
6. Secure your data using **Access Control Lists** and **Bucket Policies**

S3 Storage Classes:

1. **S3 Standard** - это тот, который имеет 99,99% доступности и 11 девяток по долговечности. Она хранится с избытком на нескольких устройствах в нескольких хранилищах и рассчитана на то, чтобы выдержать потери двух объектов одновременно. [miliseconds]

2. **S3 Infrequently Accessed (S3 - IA)** - S3 с редким доступом. Это, в основном, для данных, к которым обращаются реже, но требуют быстрого доступа, когда они вам нужны. В этом случае плата за хранение данных ниже, чем в S3, но взимается плата за извлечение. [miliseconds]
3. **S3 One Zone - IA** - S3 с одной зоной нечастого доступа. И это тот случай, когда вам нужен действительно недорогой вариант для редко используемых данных. И вам даже не нужно, вам не нужно беспокоиться о нескольких зонах доступности. Данные буквально хранятся в одной зоне доступности. И доступ к ним осуществляется нечасто, но вам все равно нужно иметь возможность мгновенного доступа к этим данным. [miliseconds]
4. **S3 Intelligent Tiering** - Она использует машинное обучение. И в основном, что она делает, так это смотрит на то, как часто вы используете свои объекты, и затем перемещает ваши объекты по разным классам хранения на основе полученных знаний. Так, он переместит его из стандартного S3 в S3 с редким доступом, потому что он знает, что вы не обращаетесь к этим файлам. Это и есть интеллектуальная многоуровневая система S3. Это четыре класса хранилищ. [miliseconds]
5. **S3 Glacier** - в основном предназначен для архивирования данных. Так что если вы хотите архивировать свои данные, может быть, они вам не нужны, может быть, вы должны хранить их в течение семи лет из-за какого-то федерального постановления, вы будете использовать Glacier. *Вы можете хранить любой объем данных и это действительно супер, супер дешево. А время поиска данных можно настраивать от нескольких минут до нескольких часов*. [select minutes or hours]
6. **S3 Glacier Deep Archive** - это самый низкий класс хранения, самый дешевый класс хранения, который вы можете купить, но время поиска будет составлять 12 часов. Поэтому если вы хотите получить данные обратно с помощью Deep Archive, вы отправляете запрос и получаете данные через 12 часов. [select hours]

S3 Bill:

1. **Storage** - Чем больше храните в S3, тем больше будет выставлен счет.
2. **Requests** - также взимается плата за количество запросов. Если вы делаете много запросов к этим объектам, это будет стоить дороже.

3. **Storage Management Pricing** - оплата за управление хранением. Таким образом, это различные уровни, которые доступны.
4. **Data Transfer Pricing** - также получаете плату за передачу данных.
5. **Transfer Acceleration** - также получаете плату за ускорение передачи данных
6. **Cross Region Replication Pricing** - Допустим, у вас есть ведро, и оно находится на востоке США. И вы хотите автоматически реплицировать свои объекты в другое ведро, которое находится, скажем, в Сиднее, и вы хотите сделать это для высокой доступности а также для аварийного восстановления. То есть, как только вы загрузите объект в ваше ведро на востоке США, и у вас включена межрегиональная репликация, эти объекты будут автоматически реплицированы в ваше ведро в Сиднее.
7. **S3 Transfer Acceleration** - Это позволяет быстро, легко и безопасно передавать файлов на большие расстояния между вашими конечными пользователями и ведром S3. По сути, это использование преимуществ глобально распределенных пограничных точек Amazon CloudFront. Когда данные поступают на пограничный узел и направляются в Amazon S3 по оптимизированному сетевому маршруту. Все, что они делают, это используют магистральную сеть Amazon. По сути, если вы включите функцию Transfer Acceleration, юзеры загружают свои файлы в Edge Locations, а не в само ведро S3.

How to restrict bucket access?

1. **Bucket Policies** - Первый - это использование политики ведра, которая будет применяться ко всему ведру.
2. **Object Policies** - использование объектных политик. И вместо того, чтобы применять их ко всему ведру, они применяются к отдельным файлам внутри ведра.
3. **IAM Policies to Users & Groups** - мы можем использовать политики IAM для пользователей и групп в вашей учетной записи AWS для контроля доступа к ведру. (Так, возможно, вы хотите, чтобы ваш отдел кадров имел доступ к ведру HR, но вы не хотите, чтобы финансовый отдел или отдел продаж и маркетинга могли читать эту конфиденциальную информацию. Поэтому вы можете использовать политики управления доступом к идентификационным данным и применять их к отдельных пользователей или к группам, которые могут содержать отдельных пользователей).

S3 Security and Encryption:

- You can encrypt individual objects
- You can also encrypt your objects at a bucket level

Versioning:

- каждый раз, когда вы делаете что-то общедоступным, даже если это последняя версия, это не обязательно делает публичными другие версии. Вы должны зайти и сделать это индивидуально.
- Версионность хранит все версии объекта, включая любые права, и даже если вы удалите объект, он поместит маркер удаления на этот объект, но версии, которые существовавшие до этого, все еще будут существовать.
- Версионирование имеет многофакторную аутентификацию и возможность удаления, и это в основном использует многофакторную аутентификацию для удаления файла.
- It's a fantastic backup tool and once enabled version and cannot be disabled, it can only be suspended.

Conclude:

- [ ] Files can be zero bytes to five terabytes in size.
- [ ] There's unlimited storage.
- [ ] S3 is a universal namespace.
- [ ] Your files are stored in these things called buckets.
- [ ] And a bucket is basically just a folder in the cloud.
- [ ] So essentially, it's only used to store files, you're not going to install an operating system on S3 and you're not going to use it to host the database.
- [ ] S3 Storage Classes
- [ ] S3 Bill

S3 Object Lock and Glacier Vault Lock

1. Object Locks come in two modes, governance mode and compliance mode. If you remember governance mode users can't overwrite or delete an object version or alter its lock settings, unless they have special permissions.
2. With compliance mode though, protected objects version, can't be overwritten or deleted by any user, including the root user in your account

3. S3 Glacier Vault Lock is. It allows you to easily deploy and enforce compliance controls for individual S3 Glacier Vaults with a Vault Lock policy and you can specify controls such as WORM in a Vault Lock policy and lock the policy from future edits and once locked, the policy can no longer be changed.

S3 Performance

- Чем больше у нас префиксов, тем большей производительности мы можем добиться. Prefixes - simply then is the pathway between your bucket name and your file.
- Если вы используете шифрование на стороне сервера для KMS, имейте в виду, что при шифровании и расшифровке данных, вы столкнетесь с жесткими ограничениями, которые зависят от региона. If you're using SSE KMS to encrypt your objects in S3, you must keep in mind the KMS limits. Uploading and downloading will count towards the quota. It is region specific.
- Also remember to use multi-part uploads to increase your performance when uploading files to S3, it should be used for any files over 100 megs, and it will be used.

S3 Select and Glacier Select

- S3 Select - it's a way of essentially pulling your data from S3, using SQL.
  - So, S3 Select - It's a way of using SQL to download the data that you need from S3
- Glacier Select - is just like S3 Select, and it allows you to run SQL queries against Glacier directly.

AWS Organizations and Consolidated Billing

- Always enable multi-factor authentication on the root account.
- You should always use a strong and complex password on the root account.
- The paying account should be used for billing purposes only, do not deploy resources into the paying account
- The paying account is simply the root account or master account.
- Enable or disable AWS services using service control policies or SCP

Sharing S3 Buckets Across Accounts

1. `Using Bucket Policies & IAM (applies across the entire bucket). Programmatic Access Only`

2. `Using Bucket ACLs & IAM (individual objects). Programmatic Access Only`
3. `Cross-account IAM Roles. Programmatic and Console access.`

Cross-Region Replication Bucket

- Remember that versioning must be enabled on both the source and the destination buckets in order for replication to work.
- Files in an existing bucket are not replicated automatically. All subsequent or updated files will be replicated automatically.
- Delete markers are not replicated, and then deleting individual versions or delete markers will also not be replicated.
- If you change the permissions of an object in the source bucket, it does not change those permissions in the destination bucket
- The cool thing about cross-region replication is you can do it between buckets in the same AWS account, but you can also do it for buckets in different AWS accounts.
- Cross Region Replication - is at a high level

S3 Transfer Acceleration

- У вас есть пользователи, у вас есть периферийные точки, и ваши пользователи загружают свои большие файлы в пограничные точки, которые затем проходят через магистральную сеть Amazon и загружают эти файлы непосредственно в ваше ведро S3 в указанном вами регионе.
- So you've got your users, you've got your edge locations, and your users upload their big files to the edge locations, which then traverse Amazon's backbone network and will upload those files directly to your S3 bucket in the region that you specify.

AWS DataSync

  - Это способ копирования данных в AWS! Он используется для перемещения больших объемов данных из локальной сети в AWS.

1. Used to move **large amounts** of data from on-premise to AWS.
2. Used with **NFS** and **SMB** compatible file systems.
3. **Replication** can be done hourly, daily, or weekly.
4. Install the **DataSync agent** to start the replication.
5. Can be used to replicate **EFS** to **EFS**

`How you can move from on prem to AWS. DataSync is definitely one of the valid options.`

CloudFront

- **CloudFront** - это сеть доставки контента, или Content Delivery Network (**CDN**), и, по сути, сеть доставки контента это система распределенных серверов или сеть, которая доставляет веб-страницы и другой веб-контент пользователю на основе географического местоположения пользователя, происхождения веб-страницы, и с помощью сервера доставки контента.
    - CloudFront можно использовать для доставки всего вашего сайта, включая динамический, статический, потоковый и интерактивный контент **используя глобальную сеть Edge Locations**, и запрос на ваш контент автоматически автоматически направляется в ближайший Edge Location, таким образом, контент доставляется с наилучшей возможной производительностью.

1. **Origin** - this is the origin of all the files that the CDN will distribute. This can either be an S3 bucket, or it could be an EC2 instance, or an elastic load balancer, or Route53.
2. **Edge Location -** This is location where content will be cached. This is separate to an AWS Region/AZ
3. **Distribution** - is the name that's given to the CDN, which is a collection of Edge Locations.
4. **Web Distribution** - is typically used for websites.
5. **RTMP** - Used for Media Streaming.

- [ ] Edge Locations are not just READ only - you can write to them too.
- [ ] Objects are cached for the life of the TTL (Time To Live.)
- [ ] You can clear cached objects, but you will be charged.

Create CloudFront

- **Web Distribution** and **RTMP**
- You can restrict access using signed URLs or signed cookies.
    - If you push out some data and then you figure out something's wrong and you do an update, but it's not showing up correctly, the way to deal with that is to create an invalidation.

CloudFront Signed URLs and Cookies

- Use signed URLs/cookies ***when you want to secure content*** so that only the people you authorize are able to access it.
    - A signed URL is for individual files. | **1 file = 1 URL.**
    - A signed cookie is for multiple files. | **1 cookie = multiple files.**
    - Origin Access Identity
    - If you origin is EC2, then use CloudFront.
- Поэтому, когда вы идете на экзамен, если они говорят о подписанном URL CloudFront и подписанном URL S3, или просто подумайте о том, могут ли ваши пользователи получить доступ к S3, если они используют OAI через CloudFront, то они не смогут. Поэтому вы будете использовать URL с подписью CloudFront, но если они могут получить доступ к ведру S3 напрямую, и это просто отдельный объект, то вам, вероятно, нужен URL с подписью S3.

Snowball

- **Snowball** - is a petabyte-scale data transport solution that uses secure appliances to transfer large amounts of data into and out of AWS.

Использование Snowball позволяет решить общие проблемы при передаче больших объемов данных включая высокие сетевые затраты,
длительное время передачи данных и проблемы безопасности. Передача данных с помощью Snowball проста, быстро, безопасно и может стоить всего лишь
всего в одну пятую от стоимости использования высокоскоростного Интернета. Snowball в основном поставляется в двух вариантах. У вас есть 50 терабайт или 80 терабайт. В Snowball используется несколько уровней безопасности, предназначенные для защиты ваших данных, включая устойчивые к взлому корпуса, 256-битное шифрование, и стандартный для отрасли модуль Trusted Platform Module, или TPM, который

предназначен для обеспечения как безопасности и полную цепочку хранения ваших данных.

Storage Gateway

- **Storage Gateway** - это, по сути, виртуальное или физическое устройство и он будет реплицировать ваши данные в AWS.
  - **File Gateway** - for flat files, stored directly on S3.
  - **Volume Gateway** - по сути, является способом хранения ваших виртуальных жестких дисков в S3
    - **Stored Volumes** - Entire Dataset is stored on site and is asynchronously backed up to S3.
    - **Cached volumes** - Entire Data is stored on S3 and the most frequently accessed data is cached on site.
    - **Stored Volumes** - это когда у вас есть весь набор данных на месте, а **Cached volumes** - это когда у вас есть только наиболее часто используемые данные кэшируются на месте.
  - **Tape Gateway** - is offers durable, cost-effective solution to archive your data in the AWS cloud.

Если вы работаете архитектором решений, особенно если вы работаете с компанией, которая переходит на AWS, и у них есть виртуальная лента, или есть ленточная библиотека. Вы можете использовать Storage Gateway для переноса этих данных на виртуальные ленты и реплицировать их в облако.

Athena vs. Macie

- **Athena** - is an interactive query service, which enables you to analyze and query data located in S3 using standard SQL. (Это интерактивная служба запросов. Он позволяет вам запрашивать данные, расположенные в S3, используя стандартный SQL, это бессерверный сервис, и он обычно используется для анализа данных журналов, хранящихся в S3.) + Serverless!
    - Well it can be used to query log files stored in S3.
    - So this could be your elastic load balancer logs, could be S3, access logs, etc.

        - You can also use it to generate business reports on data stored in S3.
- **Macie** - is a security service that uses machine learning and **N**atural **L**anguage **P**rocessing or **NLP** to discover, classify and protect sensitive data stored in S3. (Macie по сути является службой безопасности. Он использует искусственный интеллект для анализа ваших данных в S3 и помогает идентифицировать персонально идентифицируемую информацию или PII.)
    - **PII** - Personally Identifiable Information. (Это информация, которая используется для установления личность человека.)

Athena позволяет вам запрашивать ваши данные на S3 на основе команд SQL, которые вы пишете. Однако Macie также запрашивает данные на S3, но он использует машинное обучение и естественный язык для обнаружения информации PII.

EC2