Identity Access Management and S3

Rahul Daware

April 14, 2019

Contents

1 Identity Access Management 101 **3**
- 1.1 Features of IAM . . . 3
- 1.2 Key Terminology for IAM . . . 3

2 What is S3? **3**
- 2.1 Features of S3 . . . 3
- 2.2 How does data consistency work for S3? . . . 4
- 2.3 S3 - Guarantees . . . 4
- 2.4 S3 Storage Classes . . . 4
- 2.5 S3- Charges . . . 4
- 2.6 S3 Transfer Acceleration . . . 4
- 2.7 S3 Security and Encryption . . . 5
- 2.8 S3 -Versioning . . . 5
- 2.9 S3 Lifecycle Management . . . 5
- 2.10 S3 Cross Region Replication . . . 5
- 2.11 S3 Transfer Acceleration . . . 5
- 2.12 CloudFront . . . 6
- 2.13 Snowball . . . 6
- 2.14 Storage Gateway . . . 6
  - 2.14.1 File Gateway . . . 7
  - 2.14.2 Volume Gateway . . . 7
  - 2.14.3 Tape Gateway . . . 7
  - 2.14.4 Exam Tips . . . 8

1 Identity Access Management 101

IAM allows you to manage users and their level of access to the AWS console. It is important to understand IAM and how it works, both for the exam and for administering a company's AWS account in real life.

1.1 Features of IAM

- Centralised control of your AWS account
- Shared access to your AWS account
- Granular permissions
- Identity Federation (including Active Directory, Facebook, Linkedin etc)
- Multifactor authentication
- Provide temporary access for users/devices and services where necessary
- Allows you to set up your own password rotation policy
- Integrates with many different AWS services
- Supports PCI DSS Compliance

1.2 Key Terminology for IAM

- **Users** - End users such as people, employees of an organization etc.
- **Groups** - A collection of users. Each user in the group will inherit permissions of the group.
- **Policies** - Policies are made up of documents, called policy documents. These documents are in a format called JSON and they five permissions as to what a user/group/role is able to do.
- **Roles** - You create roles and then assign them to AWS resources

2 What is S3?

S3 provides developers and IT teams with secure, durable, highly-scalable object storage. Amazon S3 is easy to use, with a simple web services interface to store and retrieve any amount of data from anywhere on the web.

2.1 Features of S3

- S3 is a universal namespace. That is, names must be unique globally.
- When you upload a file to S3, you will receive HTTP 200 code
- S3 is Object based. Think of objects just as files. Objects consist of the following:
  - Key (this is simply the name of the object)
  - Value (This is simply the data and is made up of a sequence of bytes)
  - Version ID (Important for versioning)
  - Metadata (Data about data you are storing)
  - Subresources - Access Control Lists, Torrents
- Tiered Storage Available
- Lifecycle Management
- Versioning

- Encryption
- MFA Delete
- Secure your data using **Access Control Lists** and **Bucket Policies**

2.2 How does data consistency work for S3?

- Read after write consistency for PUTS of new objects
- Eventual consistency for overwrite PUTS and DELETES (can take some time to propagate)

2.3 S3 - Guarantees

- Build for 99.99% availability ofr the S3 platform.
- Amazon guarantee 99.9% availability.
- Amazon guarantees 99.999999999% durability for S3 information. (Remember 11 x 9s)

2.4 S3 Storage Classes

- **S3 Standard** : 99.99% availability, 99.99999999999% durability, stored redundantly across multiple devices in multiple facilities, and is designed to sustain the loss of 2 facilities concurrently.
- **S3 - IA(infrequetly Accessed)**: For data that is accessed less frequently, but requires rapid access when needed. Lower fee than S3, but you are charged a retrieval fee.
- **S3 One Zone - IA**: For where you want a lower-cost option for infrequently accessed data, but do not require the multiple availability zone data resilience.
- **S3 - Intelligent Tiering**: Designed to optimize costs by automatically moving data to the most cost-effective access tier, without performance impact or operational overhead.
- **S3 Glacier**: S3 Glacier is a secure, durable and low-cost storage class for data archiving. You can reliably store any amount of data at costs that are competitive with or cheaper than on-premises solutions. Retrieval times configurable from minutes to hours.
- **S3 Glacier Deep Archive**: S3 Glacier Deep Archive is Amazon S3's lowest cost storage class where a retrieval time of 12 hours is acceptable.

2.5 S3- Charges

- Storage
- Requests
- Storage Management Pricing
- Data Transfer Pricing
- Transfer Acceleration
- Cross Region Replication Pricing

2.6 S3 Transfer Acceleration

Amazon S3 Transfer Acceleration enables fast, easy and secure transfers of files over long distances between your end users and an S3 bucket. Transfer Acceleration takes advantage of Amazon CloudFront's globally distributed edge locations. AS the data arrives at an edge location, data is routed to Amazon S3 over an optimized network path.

2.7 S3 Security and Encryption

By default, all newly created buckets are PRIVATE. You can setup access control to your buckets using Bucket Policies and Access Control Lists. S3 Buckets can be configured to create access logs which log all requests made to the S3 bucket. This can be sent to another bucket and even another bucket in another account.**Encryption in Transit** is achieved by SSL/TLS. **Encryption at Rest (Server Side)** is achieved by:

- S3 Managed Keys - SSE-S3, AWS
- AWS Key Management Service, Managed Keys-SSE-KMS
- Server Side Encryption with Customer Provided Keys-SSE-C

You can also encrypt on client side before sending to S3.

2.8 S3 -Versioning

- Stores all versions of an object (including all writes and even if you delete an object)
- Great Backup Tool
- Once enabled, **versioning cannot be disabled**, only suspended
- Integrates ith **Lifecycle** Rules
- Versioning's **MFA Delete** capability, which uses multi-factor authentication, can be used to provide an additional layer of security.

2.9 S3 Lifecycle Management

- Automates moving your objects between the different storage tiers.
- Can be used in conjunction with versioning.
- Can be applied to current versions and previous versions.

2.10 S3 Cross Region Replication

- Versioning must be enabled on both the source and destination buckets.
- Regions must be unique.
- Files in an existing bucket are not replicated automatically.
- All subsequent updated files will be replicated automatically.
- Delete markers are not replicated.
- Deleting individual versions or deleting delete markers will not be replicated.

2.11 S3 Transfer Acceleration

S3 Transfer Acceleration utilises the CloudFront Edge NEtwork to accelerate your uploads to S3. Instead of uploading directly to your S3 bucket, you can use a distinct URL to upload directly to an edge location which will then transfer that file to S3. You will get a distinct URL to upload to in the format ¡bucket-name¿.s3-accelerate.aws.amazon.com

2.12 CloudFront

A content delivery network (CDN) is a system of distributed servers (network) that deliver webpages and other web content to a user based on the geographic locations of the user, the origin of the webpage, and a content delivery server.

- **Edge Location** - This is the location where content will be cached. This is separate to an AWS Region/AZ
- **Origin** - This is the origin of all the files that the CDN will distribute. This can be an S3 bucket, an EC2 instance, an elastic load balancer, or Route53.
- **Distribution** - This is the name given to the CDN which consists of a collection of edge locations.

Amazon CloudFront can be used to deliver your entire website, including dynamic, static, streaming, and interactive content using a global network of edge locations. Tequests for your content are automatically routed to the nearest edge location, so content is delivered with the best possible perfomance.Types of CloudFront Distributions:

- Web Distribution - Typically used for websites
- RTMP - Used for media streaming

2.13 Snowball

Snowball is a petabyte-scale data transport solution that uses secure appliances to transfer large amounts of data into and out of AWS. USing Snowball addresses common challenges with large-scale data transfers including high network costs, long transfer times, and security concerns. Transferring data with snowball is simple, fast, secure and can be as little as one-fifth the cost of high speed internet. Snowball comes in either a 50TB or 80TB size. Snowball uses multiple layers of security designed to protect your data including tamper-resistant enclosures, 256-bit encryption, and an industry standard trusted platform module (TPM) designed to ensure boths security and full chain of custody of your data. Once the data transfer job has been processed and verfified, AWS performs a software erasure of the Snowball appliance. AWS Snowball Edge is a 100TB data transfer device with ob-board storage and compute capabilities. You can use Snowball Edge to move lage amounts of data into and out of AWS, as a temporary storage tier for large local datasets, or to support local workloads in remote or offline locations. Snowball Edge connects to your existing applications and infrastructure using standard storage interfaces, streamlining the data transfer process and minimizing setup and integration. Snowball Edge can cluster together to form a local storage tier and process your data on-premises, helping ensure your applications continue to run even when they are not able to access the cloud. AWS Snowmobile is a Exabyte-scale data transfer service used to move extremely large amounts of data to AWS. You can transfer up to 100PB per snowmobile, a 45-foot long ruggedizds shipping container, pulled by a semi-trailer truck. Snowmobile makes it easy to move massive volumes of data to the cloud, including video libraries, image repositories, or even a complete data center migration. Transferring data with Snowmobile is secure, fast and cost effective.

Available Internet Connection	Theoretical Min. No. of days to transfer 100TB at 80% network utilization	When to consider AWS Import/Export Snowball?
T3(44.736 Mbps)	269 days	2TB or more
100Mbps	120 days	5TB or more
1000Mbps	12 days	60TB or more

2.14 Storage Gateway

AWS Storage Gateway is a service that connects an on-premises software appliance with cloud-based storage to provide seamless and secure integration between an organization's on-premise IT environment and AWS's storage infrastructure. The service enables you to securely store data to the AWS cloud for scalable and cost

effective storage. AWS storage gateway's software appliance is available for you to download as a virtual machine (VM) image that you install on a host in your datacenter. Storage Gateway supports either VMWare ESXi or Microsoft Hyper-V. Once you have installed your gateway and associated it with your AWS account through the activate process, you can use the AWS Management Console to create the storage gateway option that is right for you. The three different types of Storage Gateway are as follows:

- File Gateway (NFS)
- Volume Gateway (iSCSI)
  - Stored Volumes
  - Cached Volumes
- Tape Gateway

2.14.1 File Gateway

Files are stored as objects in your S3 buckets, accessed through a Network File System (NFS) mount point. Ownership, permissions, and timestamps are durable stored in S3 in the user-metadata of the object associated with the file. Once objects are transferred to S3, the ycan be managed as native S3 objects, and bucket policieis such as versioning, lifecycle management, and cross-region replication apply directly to objects stored in your bucket.

2.14.2 Volume Gateway

The volume interface presents your applications with disk volumes using the iSCSI block protocol. Data written to these volumes can be asynchronously backed as point-in-time snapshots of your volumes , and stored in the cloud as Amazon EBS snapshots. Snapshots are incremental backups that capture only changed blocks. All snapshot storage is also compressed to minimize your storage charges.

Stored Volumes let you stored your primary data locally, while asynchronously backing up that data to AWS. Stored volumes provide your on-premises applications with low-latency access to their entire datasets, while providing durable, off-site backups. You can create storage volumes and mount them as iSCSI devies from your on-premises application dservers. Data written to your stored volumes is stored on your on-premises storage hardware. This data is asynchronously backed up to Amazon Simple Storage Service (Amazon S3) in the form of Amazon Elastic Block Store (Amazon EBS) snapshots. 1GB -16TB in size for stored volumes.

Cached volumes let you use Amazon Simple Storage Service (Amazon S3) as your primary data storage while retaining frequently accessed data locally in your storage gateway. Cached volumes minimize the need to scale your on-premises storage infrastructure, while still providing your applications with low-latency access to their frequently accessed data. You can create storage volumes up to 32TB in size and attach them as iSCSI devices from your on-premises application servers. Your gateway stored data that you write to these volumes in Amazon S3 and retains recently read data in your on-premises storage gateway's cache and upload bufer storage. ! GB-32TB in size for cached volumes.

2.14.3 Tape Gateway

Tape Gateway offers a durable, cost-effective solution to archive your data in the AWS Cloud. The VTL interface it provides lets you leverage your existing tape-based backup application infrastructure to store your data on virtual tape cartridges that you create on your tape gateway. Each tape gateway is preconfigured with a media changer and tape drives, which are available to your existing client backup applications as iSCSI devices. Your add tape cartridges as you need to archive your data. Supported by NetBackup, BackupExec, Veeam etc.

2.14.4 Exam Tips

- File Gateway - For flat files stored directly on S3
- Volume Gateway (Stored Volumes) - Entire Data se is stored on site and is asynchronously backedup to S3.
- Volume Gateway (Cached Volumes) - Entire dataset is stored on S3 and the most frequently accessed data is cached on site.
- Gateway Virtual Tape Library