Simple Storage Service (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.
- S3 is a safe place to store your files.
- It is Object-based storage. Allows you to upload files
- The data is spread across multiple devices and facilities.
- Files can be from 0 Bytes to 5 TB.
- There is unlimited storage.
- Files are stored in Buckets.
- S3 is a Universal namespace. That is, names must be unique globally.
- It is one of the Oldest Service

Data Consistency Model for S3

- Read after Write consistency for PUTS of new Objects.
- Eventual Consistency for overwrite PUTS and DELETES (can take some time to propagate)
- S3 is Object Based. Objects consist of the following
    - Key (the name of the Object)
    - Value (The data and is made up of a sequence of bytes)
    - Version ID
    - Metadata (Data about data you are storing)
    - Sub resources:
        - Access Control Lists

The Basics

- Built for 99.99% availability for the S3 platform.
- Amazon Guarantee 99.9% availability.
- Amazon guarantees 99.99999999999% durability for S3 information.
- Tiered Storage Available.
- Life Cycle Management
- Versioning
- Encryption

- Secure your data using
    - Access Control Lists and Bucket Policies
- Object-based storage only (for files)
- Not suitable to install an Operating System on

Storage Tiers/Classes

- **General Purpose - 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.
- **Unknown or Changing access – S3 Intelligent Tiering:**
    - The S3 Intelligent-Tiering storage class is designed to optimize costs by automatically moving data to the most cost-effective access tier, without performance impact or operational overhead.
    - It works by storing objects in two access tiers: one tier that is optimized for frequent access and another lower-cost tier that is optimized for infrequent access.
    - For a small monthly monitoring and automation fee per object, Amazon S3 monitors access patterns of the objects in S3 Intelligent-Tiering, and moves the ones that have not been accessed for 30 consecutive days to the infrequent access tier.
    - If an object in the infrequent access tier is accessed, it is automatically moved back to the frequent access tier. There are no retrieval fees when using the S3 Intelligent-Tiering storage class, and no additional tiering fees when objects are moved between access tiers. It is the ideal storage class for long-lived data with access patterns that are unknown or unpredictable.

- **S3 Standard -IA (Infrequent 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:** want a lower cost option for infrequently accessed data, but do not require the multiple Availability Zone data resilence.
- **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.
    - To keep costs low yet suitable for varying needs, S3 Glacier provides three retrieval options that range from a few minutes to hours.

- You can upload objects directly to S3 Glacier, or use S3 Lifecycle policies to transfer data between any of the S3 Storage Classes for active data (S3 Standard, S3 Intelligent-Tiering, S3 Standard-IA, and S3 One Zone-IA) and S3 Glacier.

- **S3 Glacier Deep Archive:**
    - S3 Glacier Deep Archive is Amazon S3's lowest-cost storage class and supports long-term retention and digital preservation for data that may be accessed once or twice in a year.
    - It is designed for customers — particularly those in highly-regulated industries, such as the Financial Services, Healthcare, and Public Sectors — that retain data sets for 7-10 years or longer to meet regulatory compliance requirements.
    - S3 Glacier Deep Archive can also be used for backup and disaster recovery use cases, and is a cost-effective and easy-to-manage alternative to magnetic tape systems, whether they are on-premises libraries or off-premises services.
    - S3 Glacier Deep Archive complements Amazon S3 Glacier, which is ideal for archives where data is regularly retrieved and some of the data may be needed in minutes.
    - All objects stored in S3 Glacier Deep Archive are replicated and stored across at least three geographically-dispersed Availability Zones, protected by 99.999999999% of durability, and can be restored within 12 hours.

	S3 Standard	S3 Intelligent-Tiering*	S3 Standard-IA	S3 One Zone-IA†	S3 Glacier	S3 Glacier Deep Archive
Designed for durability	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)
Designed for availability	99.99%	99.9%	99.9%	99.5%	99.99%	99.99%
Availability SLA	99.9%	99%	99%	99%	99.9%	99.9%
Availability Zones	≥3	≥3	≥3	1	≥3	≥3
Minimum capacity charge per object	N/A	N/A	128KB	128KB	40KB	40KB
Minimum storage duration charge	N/A	30 days	30 days	30 days	90 days	180 days
Retrieval fee	N/A	N/A	per GB retrieved	per GB retrieved	per GB retrieved	per GB retrieved
First byte latency	milliseconds	milliseconds	milliseconds	milliseconds	select minutes or hours	select hours
Storage type	Object	Object	Object	Object	Object	Object
Lifecycle transitions	Yes	Yes	Yes	Yes	Yes	Yes

Charges

- Charged for:
    - Storage
    - Requests
    - Storage Management Pricing

- Data Transfer Pricing
- 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.
- File Uploads to Edge Location from there it uploads to S3 Bucket

Security and Encryption

- By default, all newly created buckets are PRIVATE.
- You can Setup access control to your buckets using
    - Bucket Policies (Applies to entire Bucket)
    - Access Control Lists (Applies to Object level within Buckets)
- S3 Buckets can be configured to create access logs which log all requests made to the S3 bucket. This can be done to another bucket.
- Encryption:
    - In-Transit (Data Transfer to and From your bucket)
        - SSL/TLS
    - At Rest
        - Server Side Encryption
            - S3 Managed Keys – SSE-S3
            - AWS Key Management Service, Managed Keys – SSE-KMS
            - Server Side Encryption with Customer Provided Keys – SSE-C
        - Client Side Encryption

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 with Life cycle rules

- Versioning's MFA Delete capability, which uses multi-factor authentication, can be used to provide an additional layer of security.

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.
- You cannot replicate to multiple buckets or use daisy chaining
- Delete markers are NOT replicated
- Deleting individual versions of delete markers will not be replicated.

Life Cycle Management

- Can be used in conjunction with Versioning.
- Can be applied to current versions and previous versions
- Following actions can now be done:
    - Transition to the Standard – Infrequent Access Storage class (30 days after the Creation date)
    - Archive to the Glacier Storage Class (30 days after IA, if relevant)
    - Permanently Delete

CDN & its Key Terminology

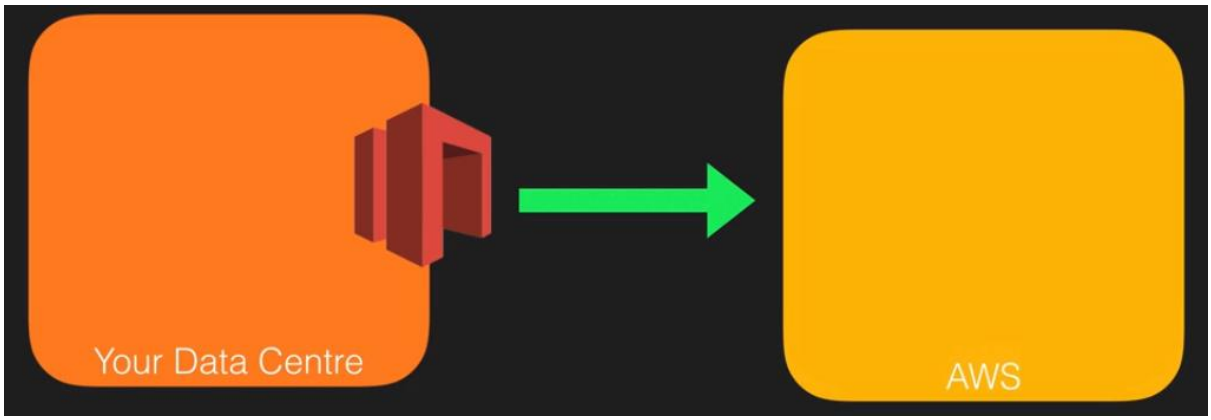
- CDN - A Content Delivery Network (CDN) is a system of distributed servers(network) that deliver webpages and other web content to a user based on the geographic locations of the user, the origin of the webpage and a content delivery server.
- Edge Location – This is the location where content will 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 either an S3 Bucket, an EC2 Instance, an Elastic Load Balancer or Route53.
- Distribution – This is the name given the CDN which consists of a collection of Edge Locations.

CloudFront and its Key Terminology

- Amazon CloudFront can be used to deliver your entire website, including dynamic, static, streaming and interactive content using a global network of Edge locations.
- Requests for your content are automatically routed to the nearest edge location, so content is delivered with the best possible performance.
- Is Optimized to work with other Amazon Webservices, like S3, EC2, Amazon Elastic Load Balancing and Amazon Route 53. Amazon CloudFront also works seamlessly with any non-AWS origin server, which stores the original, definitive versions of your files.
- Web Distribution – Typically used for Websites.
- RTMP – Used for Media Streaming.
- Edge Locations are not just READ only, you can write to them too. (i.e put an object on to them)
- Objects are cached for the life of the TTL (Time To Live)

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-premises 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.



Types of Storage Gateways

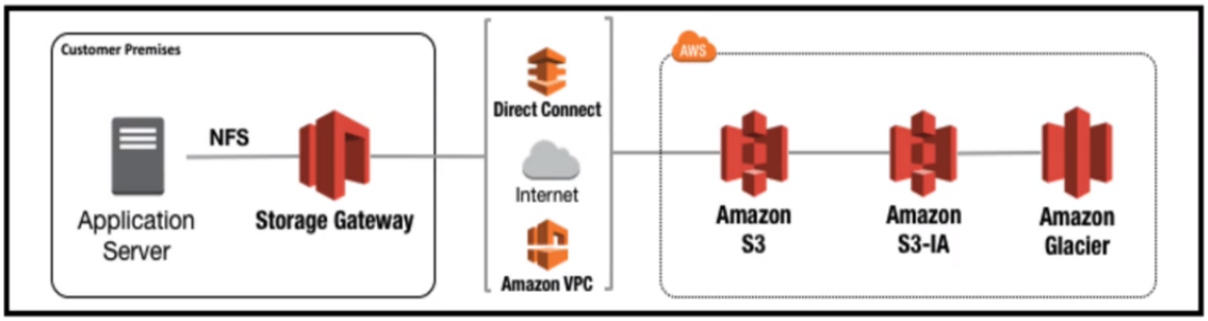
- File Gateway (NFS) - For flat files, stored directly on S3.
- Volumes Gateway (iSCSI) –Block based storage

- Stored Volumes- Entire Dataset is stored on site and is asynchronously backed up to S3.
- Cached Volumes – Entire Dataset is stored on S3 and the most frequently accessed data is cached on site.

- Tape Gateway (VTL) – Backup and archiving solutions

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 durably stored in S3 in the user-metadata of the object associated with the file.
- Once objects are transferred to S3, they can be managed as native S3 objects, and bucket policies such as versioning, life cycle management, and cross-region replication apply directly to objects stored in your bucket.



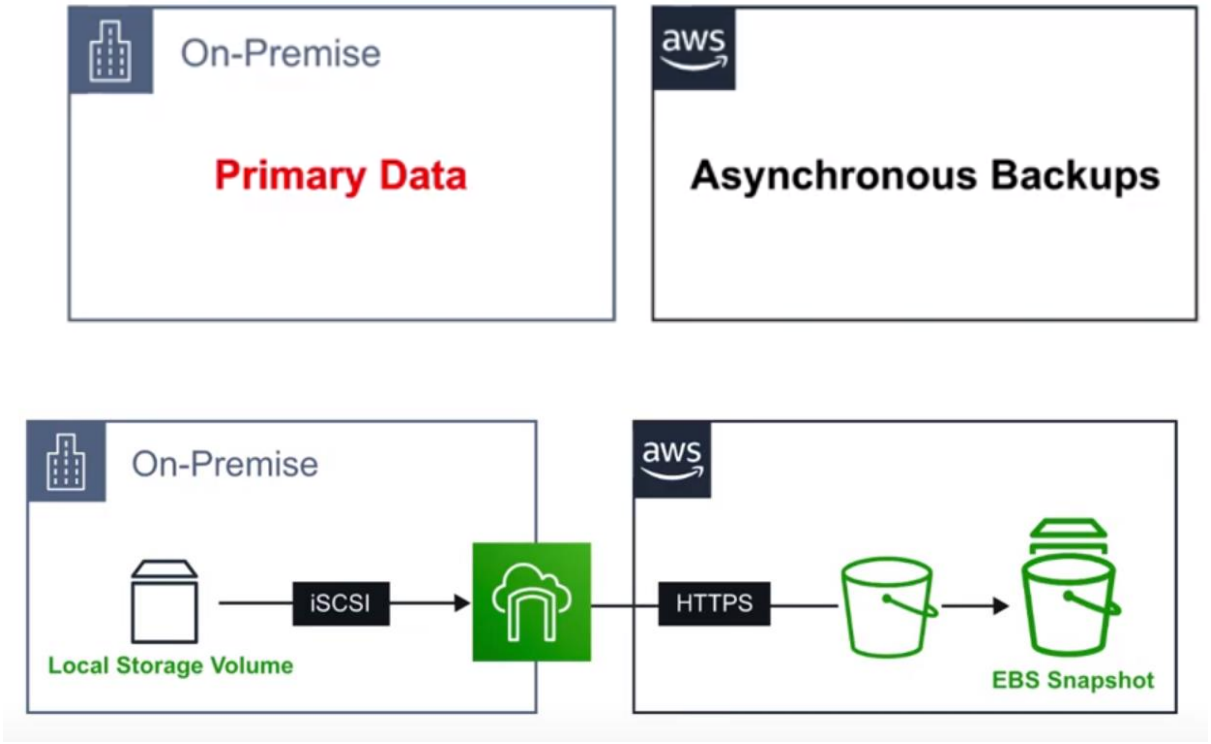
Volume Gateway

- The Volume Interface presents your applications with disk volumes using the iSCSI block protocol
- Data written to these volumes can be asynchronously backed up 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

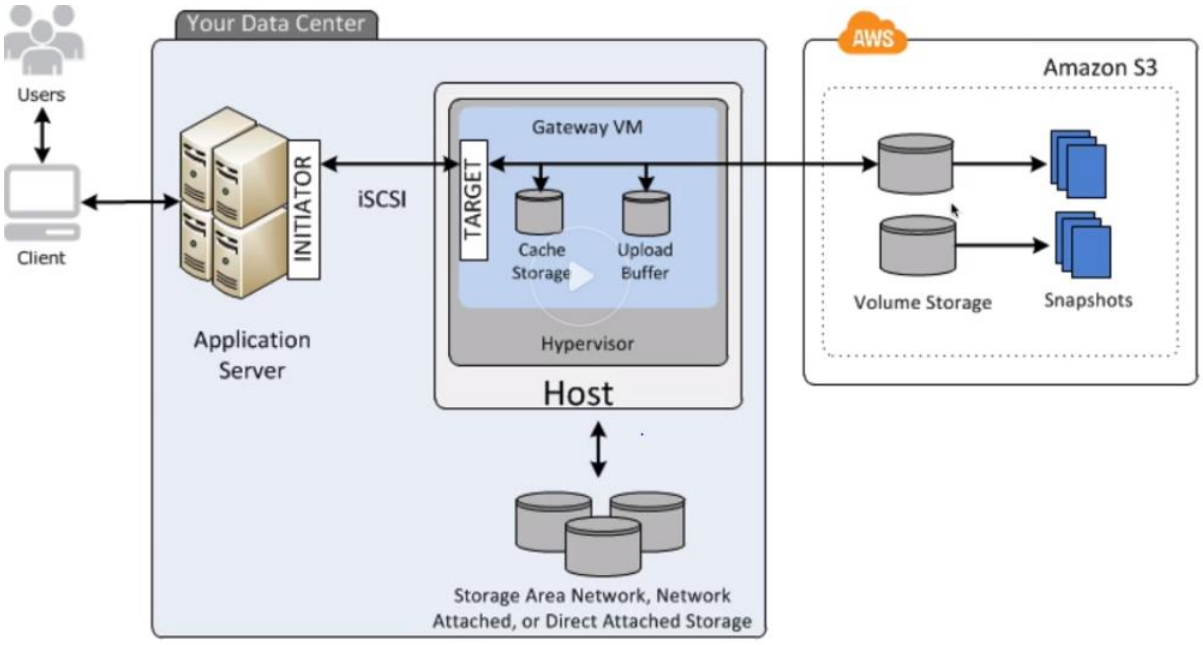
- Stored Volumes let you store your primary data locally, while asynchronously backing up that data to AWS.

- 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 devices from your on-premises application servers.
- Data written to your stored volumes is stored on your on-premises storage hardware.
- This data is asynchronously backed up to Amazon S3 in the form of Amazon EBS snapshots. 1 GB – 16 TB in size for Stored Volumes.



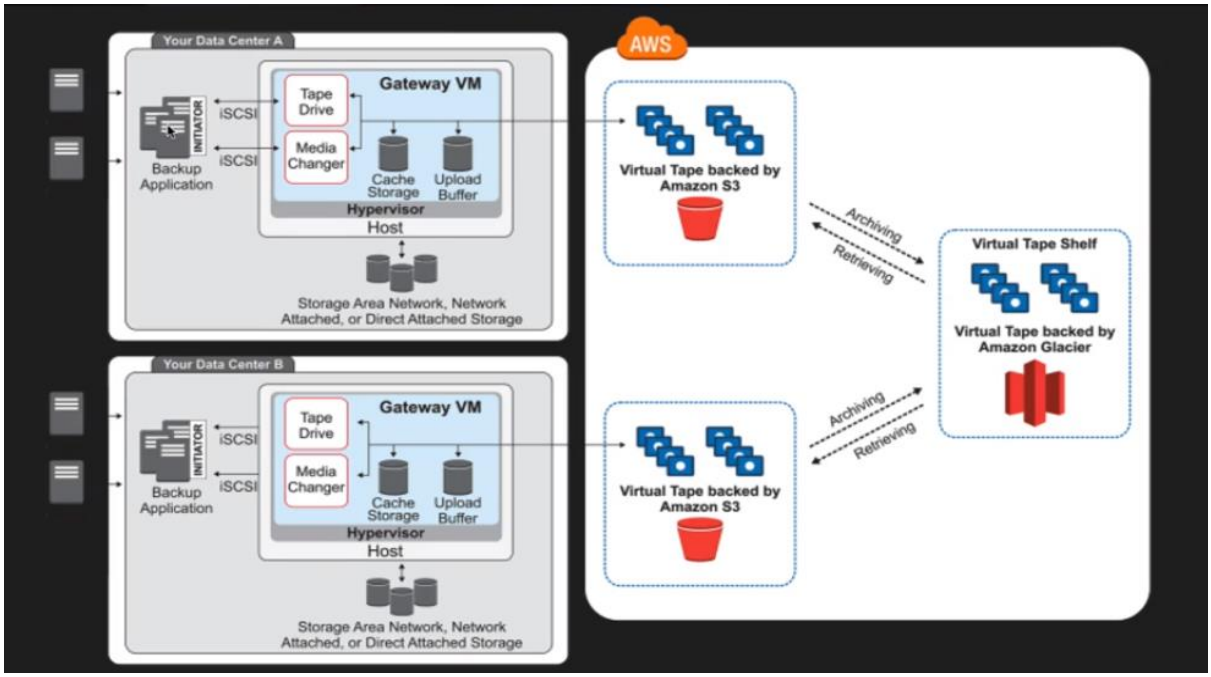
Cached Volumes

- Cached Volumes let you use 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 32 TB in size and attach to them as iSCSI devices from your on-premises application servers. Your gateway stores 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 buffer storage. 1 GB – 32 TB in size for Cached Volumes.



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 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. You add tape cartridges as you need to archive your data. Supported by NetBackup, Backup Exec, Veeam etc.



Snowball

- Before Snowball there was a service called Import/Export Disk which accelerates moving large amounts of data into and out of the AWS cloud using portable storage devices for transport.
- AWS Import/Export Disk transfers your data directly onto and off of storage devices using Amazon's high-speed internal network and bypassing the internet.
- Types of Snowball
    - Snowball (Reinvent 2015)
    - Snowball Edge (Reinvent 2016)
    - Snowmobile (Reinvent 2016)
- Snowball can
    - Import to S3
    - Export from S3

- 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.
- 80 TB Snowball in all regions. Snowball uses multiple layers of security designed to protect your data including tamper-resistant enclosures, 256-bit encryption, and an industry-standard Trusted Standard Module (TPM) designed to ensure both security and full-chain-of-custody of your data.
- Once the Data transfer job has been processed and verified, AWS performs a software erasure of the snowball appliance.
- On-board storage

Snowball Edge

- AWS Snowball Edge is a 100 TB data Transfer device with On-board storage and Compute capabilities.
- You can use Snowball edge to move large amounts of data into and out of AWS, as a temporary storage tier for large local database, 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.

Snowmobile

- AWS Snowmobile is an 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 ruggedized 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.