

Module 8: Data Storage Scaling

Identify the correct statement

Multiple EBS Volumes may be attached to the same EC2 instance

EC2 Instance Store(ephemeral volumes) are a local resource directly attached to the host that an EC2 instance runs in

Because of its scalable and durable nature, S3 is a good candidate for hosting live database transaction log and data files

CloudFront supports caching of static objects as well RTMP and RTMPE streaming and HTTP progressive download

EBS volume have a maximum size of **2TB**

Sequential IO is a better fit for EBS volumes compared to EC2 Instance Store(ephemeral volumes)

Frequent EBS snapshots increase EBS volume durability

Multiple EC2 instances with multiple EBS volumes can be pooled to create a multi-petabyte Distributed Network File System (DNFS)



Topics

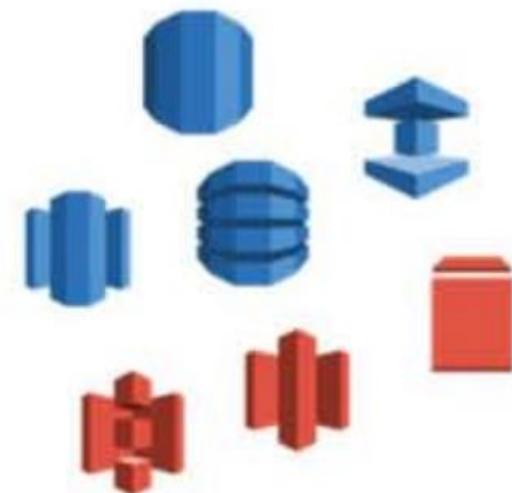
- Data storage options
- Amazon EBS
- Instance storage
- Amazon S3
- AWS Snowball
- Amazon EFS
- AWS Storage Gateway

Topics

- Data storage options
 - AWS storage options
 - Best practices for data storage
- Amazon EBS
- Instance storage
- Amazon S3
- AWS Snowball
- Amazon EFS
- AWS Storage Gateway

AWS Storage Options

- Block Storage: Instance Store, Amazon EBS
- Object Storage: Amazon S3, Amazon Glacier, Amazon CloudFront
- Sync Volumes: AWS Storage Gateway
- Relational Databases: Amazon RDS
- NoSQL Database: Amazon DynamoDB
- In-memory Cache: ElastiCache
- Content Cache: Amazon CloudFront



Best practices on choosing the right data storage solution

- Understand the variety of storage options on AWS
- Access performance, durability, cost, and interface
- POSIX versus Object store-choose where appropriate
- Use multiple cloud storage-storage hierarchy
- Horizontal versus Vertical scaling
- Be creative. Use storage alternatives like in-memory caches

It's all about performance-oriented and cost-oriented choice.

Topics

- Data storage options
- Amazon Elastic Block Storage(EBS)
 - Benefits of using Amazon EBS
 - Standard volumes versus Provisioned IOPS volumes
 - Amazon EBS pricing
- Instance storage
- Amazon S3
- AWS Snowball
- Amazon EFS
- AWS Storage Gateway



AWS Solutions Architect Associate

Elastic Block Store



Elastic Block Store Introduction

Elastic Block Store (EBS)



**A virtual hard drive in the cloud.
Create new volumes attach to EC2 instances,
backup via snapshots and easy encryption.**



Introduction to EBS

What is IOPS?

IOPS stands for Input/Output Per Second. It is the speed at which **non-contiguous reads and writes** can be performed on a storage medium. high I/O = lots of small fast reads and writes

What is Throughput?

The **data transfer rate to and from the storage** medium in megabytes per second.

What is Bandwidth?

bandwidth is **the measurement of the total possible speed of data movement along the network**



Think of **Bandwidth** as the **Pipe** and **Throughput** as the **Water**.



Introduction to EBS

Elastic Block Store (EBS) is a highly available and durable solution for attaching persistent block storage volumes to an EC2 instance. Volumes are automatically replicated within their Availability Zone (AZ) to protect from component failure.

There are  **5 Types** of EBS Storage

1. **General Purpose (SSD)** (gp2) for general usage without specific requirements
2. **Provisioned IOPS (SSD)** (io1) when you require really fast input & output
3. **Throughput Optimized HDD** (st1) magnetic drive optimised for quick throughput
4. **Cold HDD** (sc1) Lowest cost HDD volume for infrequently access workloads
5. **EBS Magnetic** (standard) previous generation HDD



AWS Solutions Architect Associate

Elastic Block Store



Volume Types



EBS - Volume Type Usage

	Solid State Drives (SSD)		Hard Disk Drives (HDD)		
Volume Type	General Purpose	Provisioned IOPS SSD	Throughput Optimized HDD	Cold HDD	EBS Magnetic
API Names	gp2	io1	st1	sc1	standard
Description	Balances price and performance	Highest SSD performance for Mission-critical low latency or high throughput	Low-cost. Designed for frequently accessed, throughput intensive workloads	Lowest HDD cost. Less frequently used workloads	
Use Cases	Most Workloads	Large Databases IOPS greater than 16,000 or Throughput greater than 250 MiB	Data Warehouses Big Data Log Processing	File Storage	Archival Storage
Volume Size	1GiB - 16TiB	4GB - 16 TiB	500GiB - 15TiB	500GiB - 15TiB	500GiB - 15TiB
Max IOPS	16,000	64,000	500	250	40-200

- 1. General Purpose (SSD) (gp2)** for general usage without specific requirements
- 2. Provisioned IOPS (SSD) (io1)** when you require really fast input & output
- 3. Throughput Optimized HDD (st1)** magnetic drive optimised for quick throughput
- 4. Cold HDD (sc1)** Lowest cost HDD volume for infrequently access workloads
- 5. EBS Magnetic (standard)** previous generation HDD



AWS Solutions Architect Associate

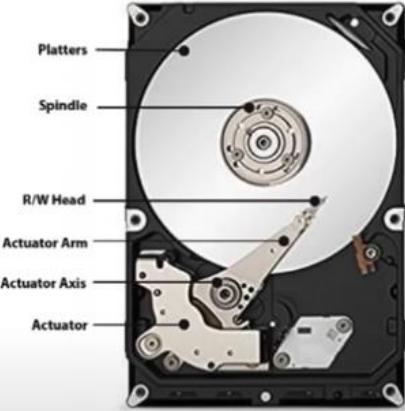
Elastic Block Store



Medium HDD



Storage Volumes



Hard Disk Drive (HDD)

HDD is magnetic storage that uses rotating platters an actuator arm and a magnetic head (similar to record player).

HDD is very good at writing a continuously amount of data.

HDD not great for writing many small reads and write (think of the arm of record player having to lift up and down and move around)

- Better for Throughput
- Physical Moving Part



AWS Solutions Architect Associate

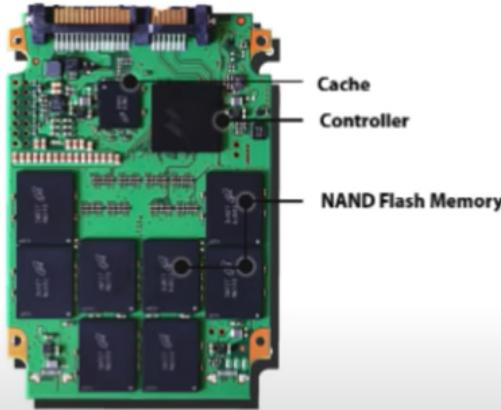
Elastic Block Store



Medium SSD



Storage Volumes



Solid State Drive (SSD)

Uses integrated circuit assemblies as memory to store data persistently, typically using flash memory. SSDs are typically more resistant to physical shock, run silently, and have quicker access time and lower latency.

- Very good frequently reads and writes (I/O)
- No physical moving parts



AWS Solutions Architect Associate

Elastic Block Store



Magnetic Tapes



Storage Volumes



Magnetic Tape

A large reel of magnetic tape. A tape drive is used to write data to the tape. Medium and large-sized data centers deployed both tape and disk formats. They normally come in the form of a cassettes. Magnetic tape is very cheap to produce and can store considerable amount of data.

- Durable for decades
- cheap to produce



A Tape drive used to read a modern cassette



Cassettes containing magnetic tape



AWS Solutions Architect Associate

Elastic Block Store



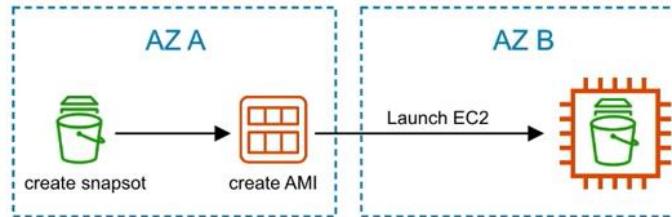
Moving Volumes



EBS - Moving Volumes

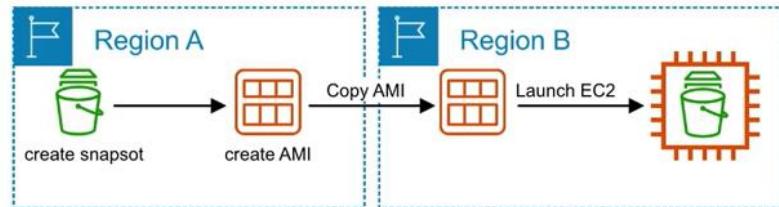
From one AZ to another

1. take a Snapshot of the volume
2. create an AMI from the Snapshot
3. launch new EC2 instance in desired AZ



From one Region to another

1. take a Snapshot of the volume
2. create an AMI from the Snapshot
3. copy the AMI to another region
4. launch a new EC2 instance from the copied AMI.





AWS Solutions Architect Associate

Elastic Block Store



Encrypted Root Volumes



EBS - Encrypted Root Volume

When you are through the wizard launching an EC2 instance you can **encrypt the volume on creation**

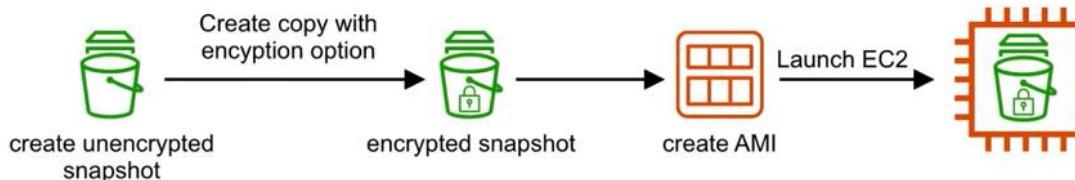
The screenshot shows the 'Create New Instance' wizard's volume configuration section. It lists a single 'Root' volume with the following details:

Volume Type	Device	Snapshot	Size (GiB)	Volume Type	IOPS	Throughput (MB/s)	Delete on Termination	Encryption
Root	/dev/xvda	snap-0e4c15b8cba3e8ae6	8	General Purpose SSD (gp2)	100 / 3000	N/A	<input checked="" type="checkbox"/>	<input type="button" value="Not Encrypted"/>

An 'Add New Volume' button is also visible.

If you want to encrypt an **existing volume** you'll have to do the following:

- Take a Snapshot of the unencrypted volume
- Create a copy of that Snapshot and select the **Encryption option**
- Create a new AMI from the encrypted Snapshot
- Launch a new EC2 instance from the created AMI





AWS Solutions Architect Associate

Elastic Block Store



EBS vs Instance Store Volumes



EBS vs Instance Store Volumes

An EC2 instance can be backed (root device) by an **EBS Volume** or **Instance Store Volume**



EBS Volumes

A **durable**, block-level storage device that you can attach to a single EC2 instance

EBS Volume created from an EBS Snapshot

- Can start and stop instances.
- Data will persist if your reboot your system

Ideal for when you want data to persist. In most use cases you'll want EBS backed volume

lasting for a very short time



Instance Store Volumes (Ephemeral)

A **temporary** storage type located on disks that are physically attached to a host machine.

An Instance Store Volume is created from a template stored in S3

- Cannot stop instances can only terminate.
- Data will be lost in case of health host fails or instance is terminated

Ideal for temporary backup, and for storing an application's cache, logs, or other random data.



AWS Solutions Architect Associate

Elastic Block Store



EBS Cheat Sheet



EBS CheatSheet

- **Elastic Block Store (EBS)** is a virtual hard disk. Snapshots are a point-in-time copy of that disk.
- Volumes exist on EBS. Snapshots exist on S3.
- Snapshots are incremental, only changes made since the last snapshot are moved to S3.
- Initial Snapshots of an EC2 instance will take longer to create than subsequent Snapshots
- If taking Snapshot of a root volume, the EC2 instance should be stopped before Snapshotting
- You can take Snapshots while the instance is still running.
- You can create AMIs from Volumes, or from Snapshots.
- **EBS Volumes** A **durable**, block-level storage device that you can attach to a single EC2 instance
- **EBS Volumes** can be modified on the fly eg. storage type or volume size.
- Volumes always exist in the same AZ as the EC2 instance.
- **Instance Store Volumes** A **temporary** storage type located on disks that are physically attached to a host machine.
- **Instance Store Volumes** (ephemeral) cannot be stopped. If the host fails then you lose your data.
- EBS Backed instances can be stopped and you will not lose any data.
- By default root volumes are deleted on termination.
- **EBS Volumes** can have termination protection (don't delete the volume on termination)
- Snapshots or restored encrypted volumes will also be encrypted.
- You cannot share a snapshot if it has been encrypted.
- Unencrypted snapshots can be shared with other AWS accounts or made public.

Topics

- Data storage options
- Amazon Elastic Block Storage (EBS)
- Instance storage
 - Benefits of Amazon Instance Store
 - Instance Store Best Practices
- Amazon S3 and Amazon CloudFront

Instance Store (1 of 3)

- No additional charge beyond your Amazon EC2 Instance
- Number and size of volumes varies by Amazon EC2 instance type
 - Larger instances have larger/more volumes
 - Hi1.4xlarge=2x1024GB SSD
 - C1.xlarge=4x450GB
 - T1.micro=none
- Not automatically attached; Must request at instance launch
- Volatile
 - No persistence
 - Data is gone when an Amazon EC2 instance stops, fails or is terminated

Instance Store (2 of 3)

- Zero network overhead; local, direct attached resource
 - No network variability
 - Not optimized for random I/O
 - Generally better for sequential I/O
- Root volume and data volume are lost on physical disk failure, stopping, or terminating of instance
- Ideal for storing temporary data like buffers, caches, scratch data, and other temporary content, or for data that is replicated across a fleet of instances, such as a load-balanced pool of web servers

Instance Store (3 of 3)

- High-performance SSD option
 - H1.4xlarge EC2 instance type
 - (2) x 1TB SSD local to instance
 - ~120,000 random read IOPS (4 KB blocks)
 - ~10,000-85,000 random write IOPS(4 KB blocks)
- High-storage
 - Hs1.8xlarge EC2 instance type
 - (24) x 2TB disks local to instance = 48 TiB

Instance Store Anti-Patterns

- Persistent storage
 - Consider Amazon EBS
- Database/ Structure Storage
 - Consider Amazon RDS, DynamoDB, and so on
- Shareable storage
 - Local instance storage volumes cannot be shared
 - Consider Amazon EBS
- Backups
 - Consider Amazon EBS and Amazon Snapshots

Instance Store Review Questions

- Describe benefits of Amazon Instance Store
- Describe best practices of using Instance Store

Topics

- Data storage options
- Amazon Elastic Block Storage (EBS)
- Instance storage
- Amazon S3
- AWS Snowball
- Amazon EFS
- AWS Storage Gateway

Simple Storage Service (S3)



Object-based storage service.

Serverless storage in the cloud.

Don't worry about filesystems or disk space.



Introduction to S3

What is Object Storage (Object-based Storage)?

data storage architecture that manages data as objects, **as opposed** to other storage architectures:

- **file systems** which manages data as a files and file hierarchy, and
- **block storage** which manages data as blocks within sectors and tracks.

S3 provides you with **unlimited storage**. You don't need to think about the underlying infrastructure
The S3 Console provides an interface for you to upload and access your data



S3 Object

Objects contain your data. They are like files.

Object may consist of:

- **Key** this is the name of the object
- **Value** the data itself made up of a sequence of bytes
- **Version ID** when versioning enabled, the version of object
- **Metadata** additional information attached to the object

You can store data from **0 Bytes** to **5 Terabytes** in size



S3 Bucket

Buckets hold objects. Buckets can also have folders which in turn hold objects

S3 is a universal namespace so bucket names must be unique (think like having a domain name)



AWS Solutions Architect Associate

Simple Storage Service (S3)



Storage Classes



S3 - Storage Classes

Trade **Retrieval Time, Accessibility and Durability** for **Cheaper Storage**

11 9's (Eleven Nines) = 99.999999999%

9 9's (Nine Nines) = 99.9999999%

Cheaper it gets

Standard (default) Fast! 99.99% Availability, **11 9's** Durability. Replicated across at least three AZs

Intelligent Tiering* Uses ML to analyze your object usage and determine the appropriate storage class. Data is moved to the most cost-effective access tier, without any performance impact or added overhead.

Standard Infrequently Accessed (IA) Still Fast! Cheaper if you access files less than once a month. Additional retrieval fee is applied. **50% less** than Standard (reduced availability)

One Zone IA Still Fast! Objects only exist in one AZ. Availability (is 99.5%). but cheaper than Standard IA by 20% less (Reduce durability) Data could get destroyed. A retrieval fee is applied.



Glacier

For long-term cold storage. Retrieval of data can take minutes to hours but the off is very cheap storage

Glacier Deep Archive The lowest cost storage class. Data retrieval time is 12 hours.





S3 - Storage Classes Comparison

	Standard	Intelligent Tiering	Standard IA	One-Zone IA	Glacier	Glacier Deep Archive
Durability	11 9's	11 9's	11 9's	11 9's	11 9's	11 9's
Availability	99.99%	99.9%	99.9%	99.5%	N/A	N/A
Availability SLA	99.99%	99%	99%	99%	N/A	N/A
AZs	>3	>3	>3	1	>3	>3
Min Capacity charge per object	N/A	N/A	128kb	128kb	40kb	40kb
Min storage duration charge	N/A	30 days	30 days	30 days	90 days	180 days
Retrieval fee	N/A	N/A	Per GB	Per GB	Per GB	Per GB
First byte latency	ms	ms	ms	ms	mins to hours	hours

S3 Guarantees

Platform is built for 99.99% availability
Amazon guarantee 99.9% availability
Amazon guarantees 11' 9s of durability





AWS Solutions Architect Associate

Simple Storage Service (S3)



S3 Security



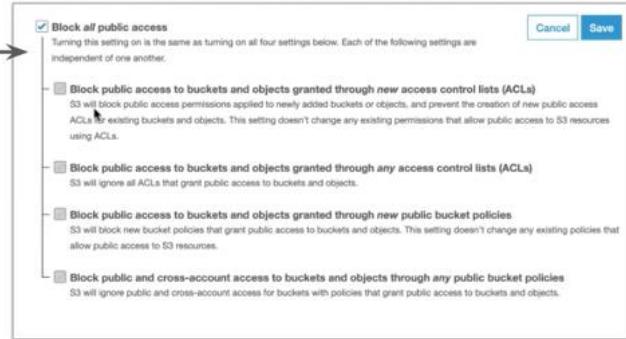
S3 - Security

All new buckets are **PRIVATE** when created by default

Logging per request can be turned on a bucket

Log files are generated and saved in a different bucket.
(even a bucket in a different AWS account if desired)

Access control is configured using **Bucket Policies** and
Access Control Lists (ACL)



The screenshot shows the 'Block all public access' checkbox checked. Below it, four other options are listed:

- Block public access to buckets and objects granted through **new access control lists (ACLS)**: S3 will block public access permissions applied to newly added buckets or objects, and prevent the creation of new public access ACLs for existing buckets and objects. This setting doesn't change any existing permissions that allow public access to S3 resources using ACLs.
- Block public access to buckets and objects granted through **any access control lists (ACLs)**: S3 will ignore all ACLs that grant public access to buckets and objects.
- Block public access to buckets and objects granted through **new public bucket policies**: S3 will block new bucket policies that grant public access to buckets and objects. This setting doesn't change any existing policies that allow public access to S3 resources.
- Block public and cross-account access to buckets and objects through **any public bucket policies**: S3 will ignore public and cross-account access for buckets with policies that grant public access to buckets and objects.

Buttons at the top right: Cancel, Save.

Access Control Lists



The screenshot shows the 'Everyone' ACL configuration. It includes sections for 'Access to the objects' (List objects, Write objects) and 'Access to this bucket's ACL' (Read bucket permissions, Write bucket permissions). Buttons at the bottom: Cancel, Save.

Legacy feature (but not deprecated)
of controlling access to buckets and
objects.

Simple way of granting access

Bucket Policies

Use a policy to define complex rule access.

```
1 {  
2   "Version": "2012-10-17",  
3   "Statement": [  
4     {  
5       "Sid": "PublicReadGetObject",  
6       "Effect": "Allow",  
7       "Principal": "*",  
8       "Action": "s3:GetObject",  
9       "Resource": "arn:aws:s3:::www.exampro.co/*"  
10    }  
11  ]  
12 }
```



AWS Solutions Architect Associate

Simple Storage Service (S3)



S3 Encryption



S3 – Encryption

Encryption In Transit

Traffic between your local host and S3 is achieved via **SSL/TLS**

Server Side Encryption (SSE) - Encryption At Rest

Amazon help you encrypt the object data

S3 Managed Keys - (Amazon manages all the keys)

SSE-AES S3 handles the key, uses AES-256 algorithm

SSE-KMS Envelope encryption, AWS KMS and you manage the keys

SSE-C Customer provided key (you manage the keys)

Client-Side Encryption

You encrypt your own files before uploading them to S3





AWS Solutions Architect Associate

Simple Storage Service (S3)



Data Consistency



S3 - Data Consistency

New Objects (PUTS)

Read After Write Consistency

When you upload a new S3 object you are able **read immediately** after writing.

Overwrite (PUTS) or Delete Objects (DELETES)

Eventual Consistency

When you overwrite or delete an object it takes time for S3 to replicate versions to AZs.

If you were to read immediately, S3 may return you an old copy. You need to generally wait a few seconds before reading.



AWS Solutions Architect Associate

Simple Storage Service (S3)

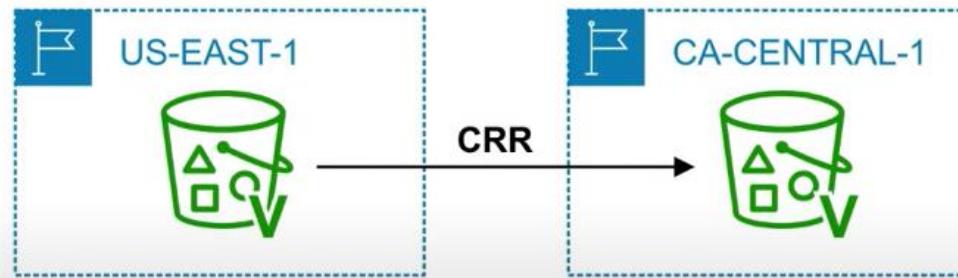


Cross-Region Replication



S3 – Cross Region Replication (CRR)

When enabled, any object that is uploaded will be **automatically replicated** to another region(s)
Provides higher durability and potential disaster recovery for objects



You must have **versioning** turned on both the **source** and **destination** buckets.
You can have CRR replicate to another AWS account



AWS Solutions Architect Associate

Simple Storage Service (S3)

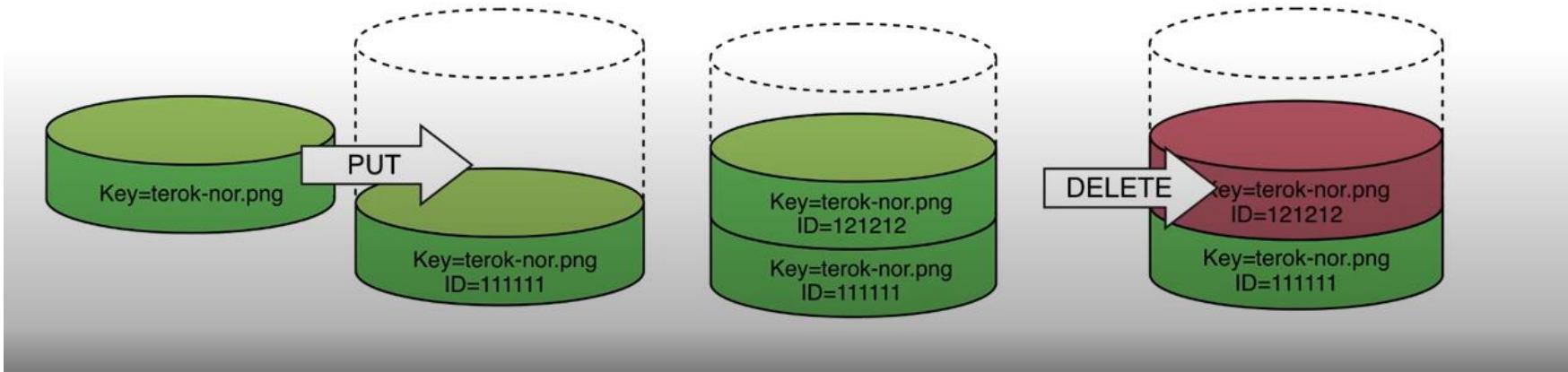


Versioning



S3 Versioning

- Store all versions of an object in S3
- Once enabled it cannot be disabled, only suspended on the bucket
- Fully integrates with S3 Lifecycle rules
- MFA Delete feature provides extra protection against deletion of your data





AWS Solutions Architect Associate

Simple Storage Service (S3)



Lifecycle Management



S3 Lifecycle Management

Automate the process of moving objects to different Storage classes or deleting objects all together.

Can be used together with **versioning**

Can be applied to both **current** and **previous** versions



Lifecycle rule

(1) Name and scope (2) Transitions (3) Expiration (4) Review

Storage class transition

There are per-request fees when using lifecycle to transition data to any S3 or S3 Glacier storage class. [Learn more](#) or see [Amazon S3 pricing](#)

Current version Previous versions

For current versions of objects [+ Add transition](#)

Object creation Days after creation

Transition to Glacier after 7 X

⚠️ Transitioning small objects to Glacier or Glacier Deep Archive will increase costs.
Before creating a lifecycle rule that transitions small objects to Glacier or Glacier Deep Archive, consider how many objects will be transitioned and how long you plan to keep the objects. Lifecycle request charges for these objects will increase your costs. [Learn more](#) or see [Amazon S3 pricing](#)

I acknowledge that this lifecycle rule will increase the one-time lifecycle

Previous Next



AWS Solutions Architect Associate

Simple Storage Service (S3)



Transfer Acceleration



S3 – Transfer Acceleration

Fast and secure transfer of files **over long distances** between your end users and an S3 bucket.

Utilizes  **CloudFront's distributed ** **Edge Locations.**

Instead of uploading to your bucket, users use a **distinct URL** for an Edge Location

As data arrives at the Edge Location it is automatically routed to S3 over a specially optimized network path. (Amazon's backbone network)





AWS Solutions Architect Associate

Simple Storage Service (S3)



Presigned URLs



S3 – PresignedUrls

Generate a url which provides you temporary access to an object to either upload or download object data. Presigned URLs are commonly used **to provide access to private objects**. You can use AWS CLI or AWS SDK to generate Presigned URLs.

```
aws s3 presign s3://mybucket/myobject --expires-in 300
```



```
https://mybucket.s3.amazonaws.com/myobject?AWSAccessKeyId=AKIAJXXXXXXXXXXXXXX&Expires=1503602631&Signature=ib0GfAovnhIF13DALdAgstdg2s%3D
```

You have a web-application which needs to allow users to download files from a password protected part of your web-app. Your web-app generates presigned url which expires after 5 seconds. The user downloads the file.



AWS Solutions Architect Associate

Simple Storage Service (S3)



MFA Delete



S3 - MFA Delete

MFA Delete ensures users cannot delete objects from a bucket unless they provide their MFA code.

MFA Delete can only be enabled under these conditions

1. The **AWS CLI** must be used to turn on MFA
2. The bucket must have **versioning turned on**



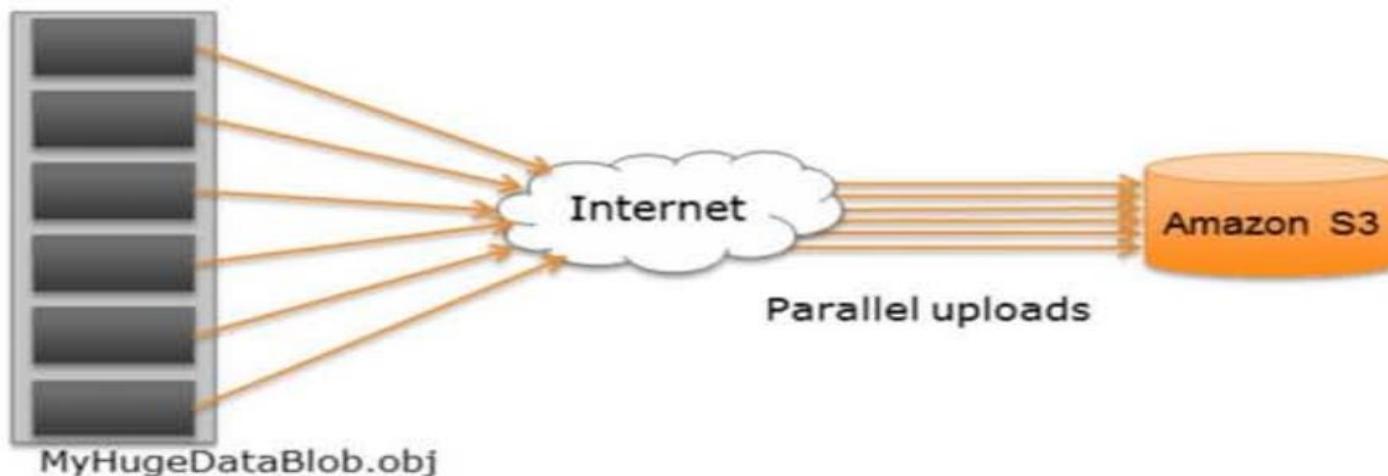
```
aws s3api put-bucket-versioning \
    --bucket bucketname \
    --versioning-configuration Status=Enabled,MFADelete=Enabled \
    --mfa "your-mfa-serial-number mfa-code" \
```

Only the bucket owner logged in as Root User can DELETE objects from bucket

Multipart Upload



- Upload as 100MB per part
- Multipart Upload supports a maximum of 10,000 parts



Website hosting using Amazon S3

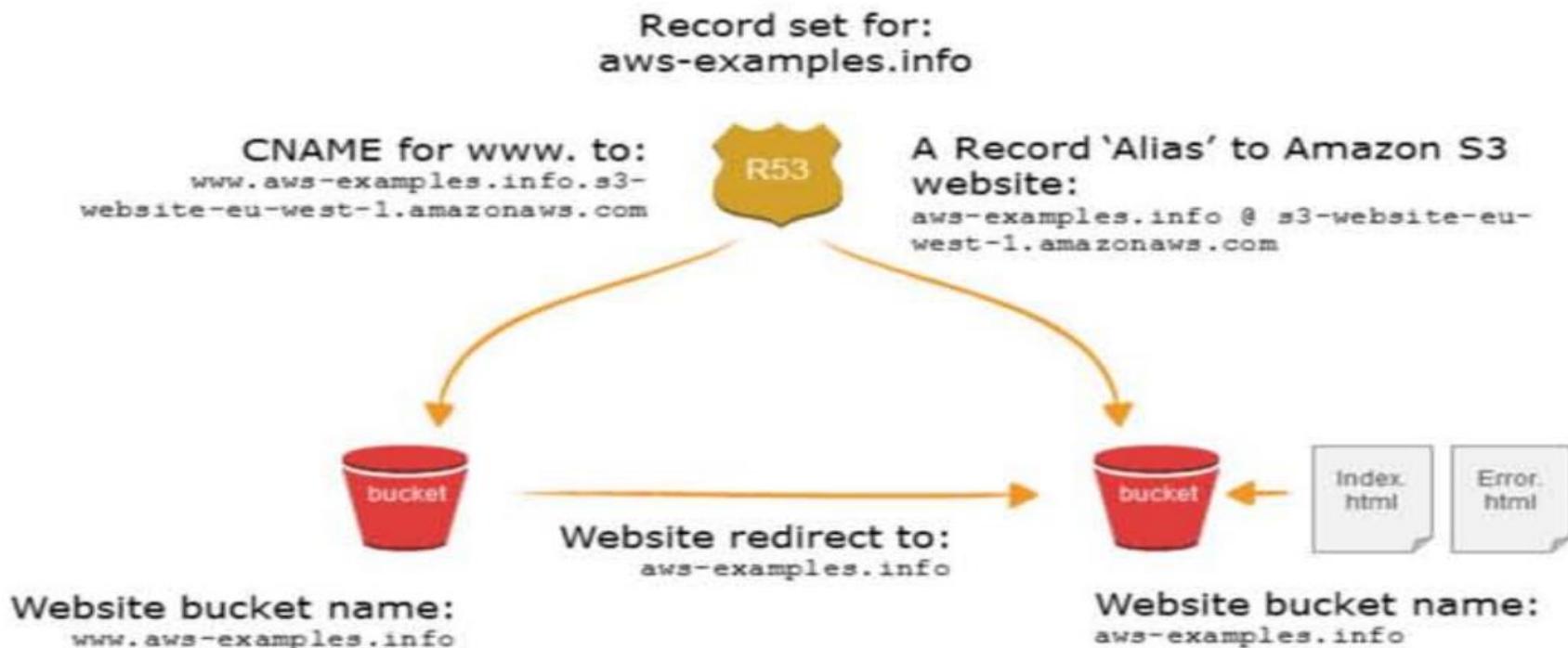


- Static sites with client-side scripts

The screenshot shows the AWS S3 console interface for enabling static website hosting. On the left, a sidebar lists 'Enable website hosting', 'Index Document: index.html', 'Error Document: error.html', and 'Edit Redirection Rules: You can set custom rules to automatically redirect web page requests for specific URLs'. A yellow box highlights the 'Setting default documents' section, and an orange arrow points from it to the 'Index Document' field. On the right, the main panel is titled 'Static Website Hosting' and contains the following information:

- Endpoint:** `www.aws-examples.info.s3-website-eu-west-1.amazonaws.com`
- Description:** You can host your static website entirely on Amazon S3. Once you enable your bucket for static website hosting, all your content is accessible to web browsers via the Amazon S3 website endpoint for your bucket.
- Configuration Options:**
 - Do not enable website hosting
 - Enable website hosting
 - Redirect all requests to another host name
- Target Bucket:** To redirect requests to another bucket, enter the name of the target bucket below. If you are redirecting to a root domain address (e.g. example.com), see our walkthrough for configuring root domain website hosting.
Redirect all requests to: `www-examples.info`
- Buttons:** Save and Cancel

Website hosting using Amazon S3 (continue)





AWS Solutions Architect Associate

Simple Storage Service (S3)



S3 Cheat Sheet



S3 CheatSheet

- **Simple Storage Service (S3)** Object-based storage. Store **unlimited** amount of data without worry of underlying storage infrastructure
- S3 replicates data across at least 3 AZs to ensure 99.99% Availability and 11' 9s of durability
- Objects contain your data (they're like files)
- Objects can be size anywhere from **0 Bytes** up to 5 Terabytes
- Buckets contain objects. Buckets can also contain folders which can in turn contain objects.
- Bucket names are unique across all AWS accounts. Like a domain name.
- When you upload a file to S3 successfully you'll receive a HTTP 200 code
- **Lifecycle Management** Objects can be moved between storage classes or objects can be deleted automatically based on a schedule
- **Versioning** Objects are given a Version ID. When new objects are uploaded the old objects are kept. You can access any object version. When you delete an object the previous object is restored. Once Versioning is turned on it cannot be turned off, only suspended.
- **MFA Delete** enforces DELETE operations to require MFA token in order to delete an object. Must have versioning turned on to use. Can only turn on MFA Delete from the AWS CLI. Root Account is only allowed to delete objects
- All new buckets are **private by default**
- Logging can be turned on for a bucket to log operations performed on objects
- **Access control** is configured using **Bucket Policies** and **Access Control Lists (ACL)**
- **Bucket Policies** are JSON documents which let you write complex control access
- **ACLs** are the legacy method (not deprecated) where you grant access to objects and buckets with simple actions



S3 CheatSheet

- **Security in Transit** Uploading files is done over SSL
- **SSE** stands for Server Side Encryption. S3 has **3 options** for SSE.
- **SSE-AES** S3 handles the key, uses AES-256 algorithm
- **SSE-KMS** Envelope encryption via AWS KMS and you manage the keys
- **SSE-C** Customer provided key (you manage the keys)
- **Client-Side Encryption** You must encrypt your own files before uploading them to S3
- **Cross Region Replication (CRR)** allows you to replicate files across regions for greater durability. You must have versioning turned on in the source and destination bucket. You can have CRR replicate to bucket in another AWS Account
- **Transfer Acceleration** provide faster and secure uploads from anywhere in the world. data is uploaded via distinct url to an Edge Location. Data is then transported to your S3 bucket via AWS backbone network.
- **PresignedUrls** is a url generated via the AWS CLI and SDK. It provides temporary access to write or download object data. PresignedUrls are commonly used to access private objects.



S3 CheatSheet

- S3 has **6 different** Storage Classes:
 - **Standard** Fast! 99.99% Availability, 11 9's Durability. Replicated across at least three AZs
 - **Intelligent Tiering** Uses ML to analyze your object usage and determine the appropriate storage class. Data is moved to the most cost-effective access tier, without any performance impact or added overhead.
 - **Standard Infrequently Accessed (IA)** Still Fast! Cheaper if you access files less than once a month. Additional retrieval fee is applied. 50% less than Standard (reduced availability)
 - **One Zone IA** Still Fast! Objects only exist in one AZ. Availability (is 99.5%). but cheaper than Standard IA by 20% less (Reduce durability) Data could get destroyed. A retrieval fee is applied.
 - **Glacier** For long-term cold storage. Retrieval of data can take minutes to hours but the off is very cheap storage
 - **Glacier Deep Archive** The lowest cost storage class. Data retrieval time is 12 hours.

AWS Snowball



Petabyte-scale data transfer service.

Move data onto AWS via physical briefcase computer



Snowball

Low Cost It cost thousands of dollars to transfer 100TB over high speed internet.
Snowball can reduce that costs by **1/5th**

Speed It can take 100 TB over 100 days to transfer over high speed internet
Snowball can reduce that transfer time by **less than a week**

Snowball features and limitations:

- E-Ink display (shipping information)
- Tamper and weather proof
- Data is encrypted end-to-end (256-bit encryption)
- uses Trusted Platform Module (TPM) ←
- For security purposes, data transfers must be completed within **90 days** of the Snowball being prepared.
- Snowball can **Import** and **Export** from  **S3**



a specialized chip on an endpoint device that stores RSA encryption keys specific to the host system for hardware authentication.

Snowballs come in two sizes:

- **50 TB** (42 TB of usable space)
- **80 TB** (72 TB of usable space)



AWS Solutions Architect Associate

Snowball



Snowball Edge





Snowball Edge

Similar to Snowball but with **more storage** and with **local processing**

Snowball Edge features and limitations:

- LCD display (shipping information and other functionality)
- can undertake local processing and edge-computing workloads
- Can use in a cluster in groups of 5 to 10 devices
- three options for device configurations
 - storage optimized (24 vCPUs)
 - compute optimized (54 vCPUs)
 - GPU optimized (54 vCPUs)



Snowball Edge come in  two sizes:

- **100 TB** (83 TB of usable space)
- **100 TB Clustered** (45 TB per node)



Snowball & Snowball Edge & Snowmobile *CheatSheet*

- **Snowball** and **Snowball Edge** is a rugged container which contains a storage device
- **Snowmobile** is a 45-foot long ruggedized shipping container, pulled by a semi-trailer truck.
- Snowball and Snowball Edge is for **peta-scale** migration. Snowmobile is for **exabyte-scale** migration
- **Low Cost** thousands of dollars to transfer 100TB over high speed internet. Snowball is **1/5th**
- **Speed** 100 TB over 100 days to transfer over high speed internet, Snowball takes **less than a week**
- **Snowball come in two sizes:**
 - **50 TB** (42 TB of usable space)
 - **80 TB** (72 TB of usable space)
- **Snowball Edge comes in two sizes:**
 - **100 TB** (83 TB of usable space)
 - **100 TB Clustered** (45 TB per node)
- **Snowmobile comes in one size:** 100PB
- You can both **export** or **import** data using Snowball or Snowmobile
- You can import into **S3** or **Glacier**
- **Snowball Edge** can undertake local processing and edge-computing workloads
- **Snowball Edge** Can use in a cluster in groups of 5 to 10 devices
- **Snowball Edge** provides three options for device configurations
 - storage optimized (24 vCPUs)
 - compute optimized (54 vCPUs)
 - GPU optimized (54 vCPUs)



AWS Solutions Architect Associate

Elastic File System



Introduction to Elastic File System

Elastic File System (EFS)



Scalable, elastic, cloud-native NFS file system.

Attach a single file system to multiple EC2 Instances

Don't worry about running out or managing disk space



Introduction to Elastic File System (EFS)

EFS is a file storage service for EC2 instances

Storage capacity grows (upto petabytes) and shrinks automatically based on data stored (elastic)

Multiple EC2 instances in same VPC **can mount a single EFS Volume** (volume must be in same VPC)

EC2 instances install the **NFSv4.1 client** and can then mount the EFS volume

EFS is using Network File System version 4 (**NFSv4**) protocol

EFS creates multiple **mount targets** in all your VPC subnets.

You based per space used starting at \$0.30 GB / month

Mount targets							
VPC	Availability Zone	Subnet	IP address	Mount target ID	Network interface ID	Security groups	Mount target state
vpc-75fcfd90f (default)	us-east-1b	subnet-edd727b1 (default)	172.31.36.225	fsmt-1a56eef0	eni-089be3fede4855115	sg-c4b0668c - default	Available
	us-east-1a	subnet-4d1f9b07 (default)	172.31.24.213	fsmt-1c56eefc	eni-07f1538413c5581a7		Creating
	us-east-1c	subnet-9706c1f0 (default)	172.31.10.213	fsmt-1e56eef0	eni-0a685b10dbc4144aa		Creating
	us-east-1e	subnet-cd8e99f2 (default)	172.31.63.220	fsmt-1f56eef0	eni-0120337a18e205006		Creating
	us-east-1d	subnet-f7684d1 (default)	172.31.85.52	fsmt-e157ef01	eni-043ea2b3a223e4944	sg-c4b0668c - default	Available
	us-east-1f	subnet-fa0e81f5 (default)	172.31.76.141	fsmt-e257ef02	eni-04b4c60064f530831	sg-c4b0668c - default	Available



AWS Solutions Architect Associate

Elastic File System



EFS Cheat Sheet



EFS *CheatSheet*

- **Elastic File System (EFS)** supports the Network File System version 4 (NFSv4) protocol.
- You pay GB of storage per month
- Volumes can scale to petabyte size storage
- Volumes will shrink and grow to meet current data stored (elastic)
- Can support thousands of concurrent connections over NFS.
- Your data is stored across multiple AZs within a region.
- Can mount multiple EC2 instance to a single EFS (as long as they are all in the same VPC)
- Creates Mount Points in all your VPC subnets so you can mount from anywhere within your VPC
- Provides Read After Write Consistency.



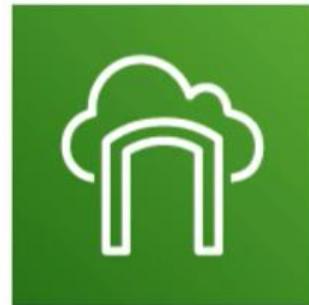
AWS Solutions Architect Associate

Storage Gateway



Storage Gateway Introduction

AWS Storage Gateway



Extending, backing up **on-premise storage to the cloud**



Introduction to Storage Gateway

AWS Storage Gateway **connects an on-premises software appliance with cloud-based storage.**

Provides you seamless and secure integration between your organization's on-premises IT environment and AWS's storage infrastructure.

Securely store your data to the AWS Cloud for **scalable** and **cost effective** storage.

Software appliance is available as a **virtual machine (VM) image.**

Supports both **VMware ESXi** and **Microsoft Hyper-V**

Once installed and activated you can use the **AWS Console** to create your gateway

There are **3 Types** of Gateways

Select gateway type

File gateway

Store files as objects in Amazon S3, with a local cache for low-latency access to your most recently used data.

Volume gateway

Block storage in Amazon S3 with point-in-time backups as Amazon EBS snapshots.

Tape gateway

Back up your data to Amazon S3 and archive in Amazon Glacier using your existing tape-based processes.

Cancel

Next





Storage Gateway Types

There are **3 Types** of Gateways

File Gateway (NFS) (store your files in S3)

Volume Gateway (iSCSI) (store copies of your hard disk drives in S3)



- **Stored Volumes**
- **Cached Volumes**

Tape Gateway (VTL) (virtual tape library)



AWS Solutions Architect Associate

Storage Gateway

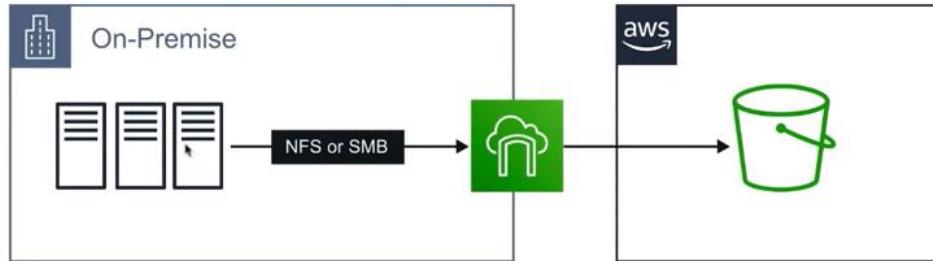


File Gateway



File Gateway (NFS)

Your files are stored as objects inside your S3 buckets.
Access your files through a **Network File System (NFS)** or **SMB** mount point.



Ownership, permissions, and timestamps are all stored within **S3 metadata** of the object associated with the file.

Once a file is transferred to S3, it can be managed as a **native S3 object**.

Bucket Policies, Versioning, Lifecycle Management, and Cross-Region Replication apply directly to objects stored in your bucket.



AWS Solutions Architect Associate

Storage Gateway

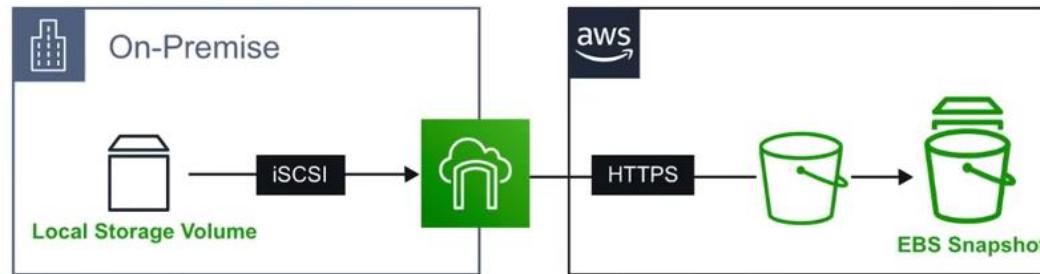


Volume Gateway



Volume Gateway (iSCSI)

Volume Gateway presents your applications with disk volumes using the **Internet Small Computer Systems Interface (iSCSI)** block protocol.



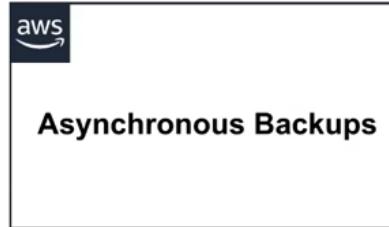
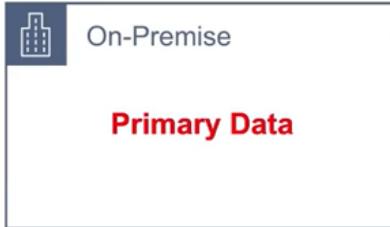
Data that is written to volumes can be asynchronously backed up as **point-in-time snapshots** of the volumes, and stored in the cloud as **AWS EBS Snapshots**.

Snapshots are incremental backups that capture only changed blocks in the volume.

All Snapshot storage is also **compressed** to help minimize your storage charges.



Volume Gateway - Stored Volumes



Primary data is **stored locally**, while **asynchronously backing up** that data to AWS.

Provide your on-premises applications with low-latency access to their entire datasets, while still providing durable off-site backups.

Create storage volumes and **mount them as iSCSI devices** from your on-premises servers.

Any data written to stored volumes in **stored on your on-premises** storage hardware.

Amazon Elastic Block Store (**EBS**) snapshots are backed up to **AWS S3**.

Stored Volumes can be between **1GB - 16TB** in size



Volume Gateway - Cached Volumes



On-Premise

Cache Most Frequently Accessed Files



Primary Data

Lets you **use AWS S3 as your primary data storage**, while retaining **frequently accessed data locally** in your storage gateway.

Minimizes the need to scale your on-premises storage infrastructure, while still providing your applications with low latency data access.

Create storage volumes up to 32TB in size and **attach them as iSCSI devices** from your **on-premises servers**.

Your gateway stores data that you write to these volumes in S3, and retains recently read data in your on-premises storage gateway cache and upload buffer storage.

Cached volumes can be between **1GB - 32GB** in size



AWS Solutions Architect Associate

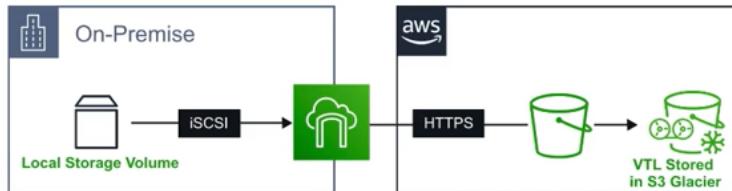
Storage Gateway



Tape Gateway (VTL)



Tape Gateway (VTL)



A durable, cost-effective solution to archive your data in the AWS Cloud

The **VTL** interface it provides let you leverage existing tape-based backup application infrastructure.

Store data on **virtual tape cartridges** that you create on your tape gateway.

Each tape gateway is **pre-configured 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.



Veritas Backup Exec™

VERITAS
NetBackup

veeam



Supported by **NetBackup**, **Backup Exec**, and **Veeam**.



AWS Solutions Architect Associate

Storage Gateway



Storage Gateway Cheat Sheet



Storage Gateway CheatSheet

- **Storage Gateway** connects on-premise storage to cloud storage (hybrid storage solution)
- There are three types of Gateways: File Gateway, Volume Gateway, Tape Gateway
- **File Gateway** lets S3 act a local file system using NFS or SMB, extends your local hard drive to S3
- **Volume Gateway** is used for backups and has two types: **Stored** and **Cached**
- **Stored Volume Gateway** continuously backups local storage to S3 as EBS Snapshots **Primary Data on-Premise**
- Stored Volumes are **1GB to 16TB** in size
- **Cached Volume Gateway** caches the frequently used files on-premise. **Primary Data** is stored on S3
- Cached Volumes are **1GB to 32GB** in size
- **Tape Gateway** backups up virtual tapes to S3 Glacier for long archive storage