



Object Storage options In AWS



Atul Kumar



LinkedIn QR code

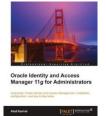
Scan My code **Atul Kumar** Founder at K21Academy: Learn Cloud From Experts

- > Author & Cloud Architect
- ➤ 21+ Years working in IT & Certified Cloud Architect
- > Helped **8500+ individuals** to learn Cloud & Cloud Native











WhatsApp & Ticketing System

support@k21academy.com









Module Agenda





- Traditional Storage
- AWS Cloud Storage
- Different Storage Options Available On AWS
- Simple Storage Service (S3) and its components
- Working Of S3
- Difference Between S3, EBS, And EFS
- Bucket Policy

Module: Agenda



- Access Control List(ACL)
- Hosting A Static Website On Amazon S3
- Versioning
- Cross-Region Replication (CRR) and its Use case
- Amazon S3 Transfer Acceleration
- Choice of Storage Classes on S3
- Life Cycle Policy Of S3 Bucket
- Access S3 And Glacier
- Cost Optimization Of S3





- Upload A File To AWS S3 Through A Website
- Cloud Front And Snowball
- > AWS FSx
- Storage Gateway
- Different modes of Gateway: File, Volume and Tape Gateway





Traditional Storage





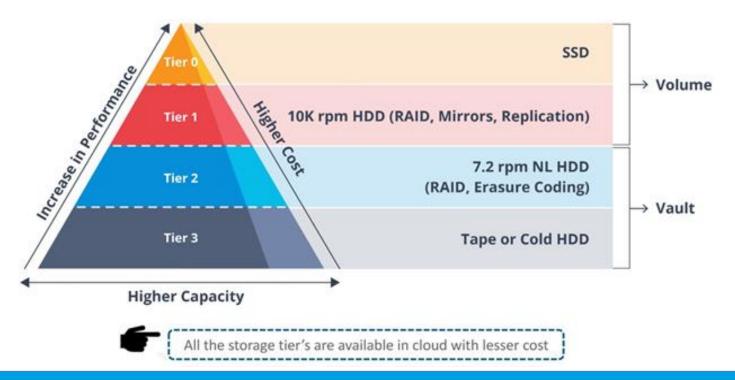
Volume is a storage device that is formatted to store directories and files for frequent use. e.g. fixed disk, CD-ROM etc.

Vault is a storage box or a container which stores the archive data for longer period of time









Why We Need To Cloud Storage?



Solution provided by cloud Solution provided by cloud Storage is sitting idle in the Backup processes slow storage during day data center Pay for infrastructure Fast services with low cost you use with no up 50% and risk front payment On average nearly 40% of Nearly 50% of organizations storage punchased is not used need to reduce backup times Data protection strategies are Inactive data is sitting on incomplete coastly storage Data reduction technique Fast services with low cost 40% and archiving to store and low risk inactive cold data Almost 40% of respondents have Up to 95% of data is cold only one backup method Migrations are frequent, costly Data silos duplicate & lengthy management, hardware 2015 2018 Easy cloud-based backup Easy migration of and data archiving solutions Documents, images & other Plan for storage migration every Files are growing rapidly & 3 years multiply data silos





AWS Cloud Storage



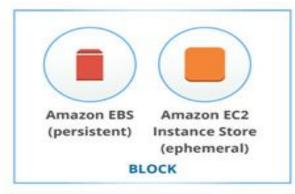


Cloud storage is a way of storing the data over the internet, which is highly















Storage Services



Block Storage

- Data used here is raw and unformatted
- It is accessed by only one instance at a time
- Instance store is used here to store the temporary data whereas EBS is used to store the persistent data





File Storage

- Data structures that keep track of the related set of data
- EFS can be accessed by the multiple instance at a time through NFS protocol
- It is used as clustered database and document sharing

Object Storage

- Direct access to the data without traversing through directories
- It can be accessed by the users who have the access to the bucket through http or https or API





Cloud Data Migration

- ➤ Cloud Data Migration is a storage system which connects an on-premises software appliance with cloud-based storage
- > Data can be transferred through internet also, but it will be slow and costly

Cloud Storage	What is it?	
Storage Gateway	Integrates on-premises IT environments with AWS storage (Storage transfer limit = 32 TB)	
Snowball \$	A service that enables large-volume data transfer (Storage transfer limit = 80 TB/Snowball)	





S3 (Simple Storage Service)



S3 (Simple Storage Service)

➤S3 is an interface that help to store and retrieve any amount of data, at any point of time, from anywhere through internet.

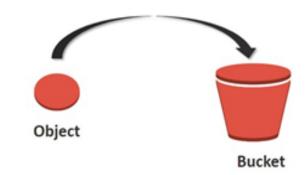












- > Fundamental entities stored in Amazon S3.
- ➤ Each Amazon S3 object has data, a key, and metadata.
- Additional information that is used to refers the data like the name, format and the time at which they the file have been added is known as metadata.
- ➤ Object can be Uniquely identified within a bucket by a key(name) and a version ID.
- ➤ Each object can contain up to 5 TB of data.









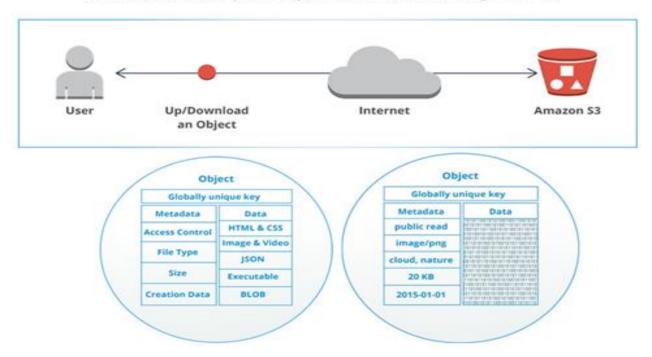


- > Used to store objects, which consist of data and metadata
- > The bucket can be configured and created in any specific region .
- ➤When an object is added to the bucket, Amazon S3 generates a unique version ID and assigns it to the object.
- ➤ By default, only 100 buckets can be created in each AWS accounts.
- ➤ It provides the unlimited storage of data.





User Download or Upload object into S3 Bucket through internet





Difference Between EFS, EBS, and S3

Amazon S3	Amazon EBS	Amazon EFS
Can be publicly accessible	Accessible only via the EC2 Machine	Accessible via several EC2 machine and AWS Services
Web interface	File system interface	Web and file system interface
Object storage	Block storage	File storage
Scalable	Hardly scalable	Scalable
Slowest among the all	Fastest among the all	Faster than S3, slower than EBS
Good for storing backups	Is meant to be EC2 drive	Good for sharable applications and workloads





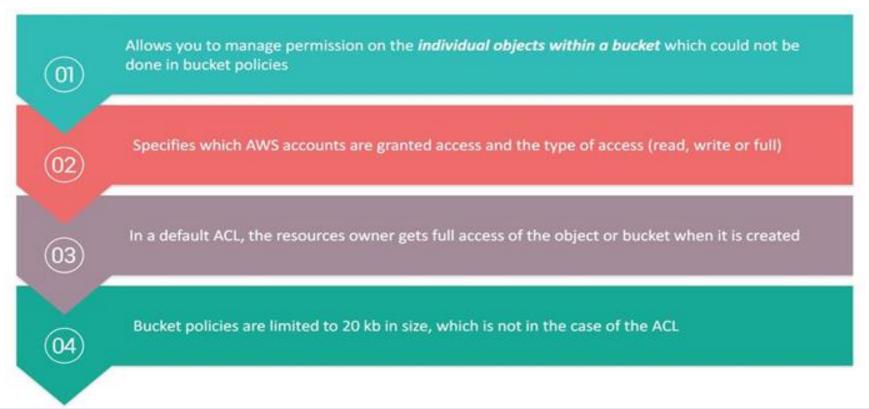
- ➤ Bucket policy allows users to authorize policies which either grant or deny access to any number of accounts and across a range or set of keys.
- ➤ It allows centralized management of permission that is the security rule is applied to all the files within the bucket.

Sample JSON script

If we want to restrict an user Alice from accessing the bucket, then we can apply the security with the help of the JSON script such that the user will not be able to access the bucket

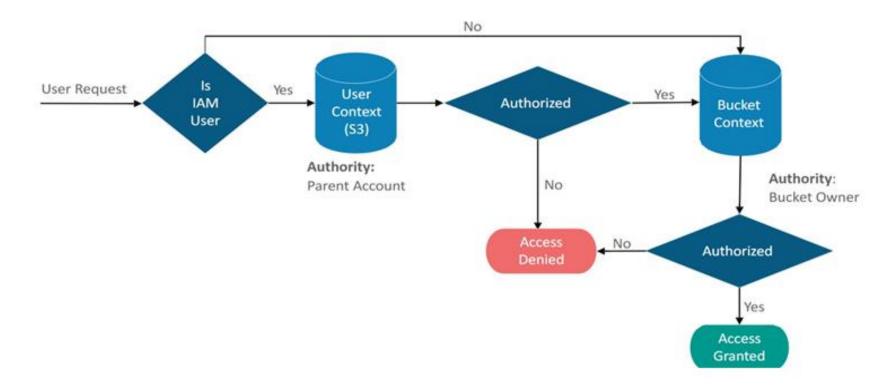












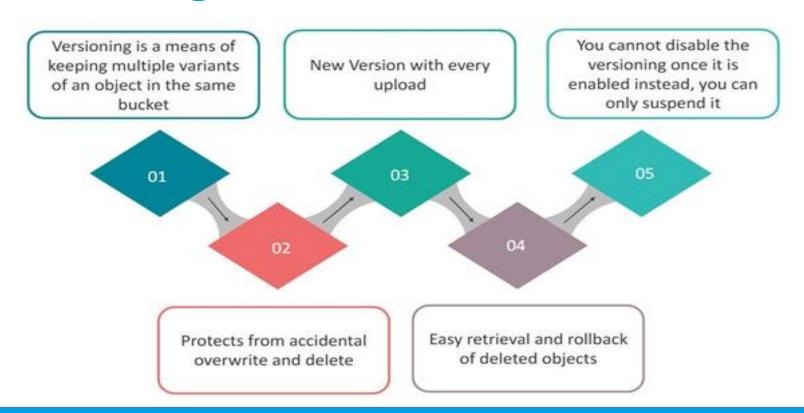




AWS Object Storage Options Versioning

Versioning







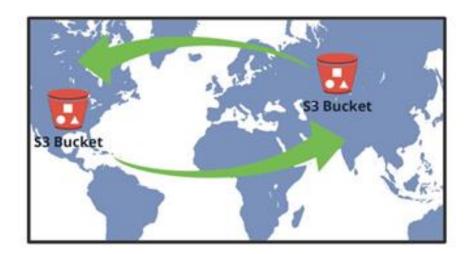


S3 Cross-Region Replication (CRR)



Cross-Region Replication (CRR)

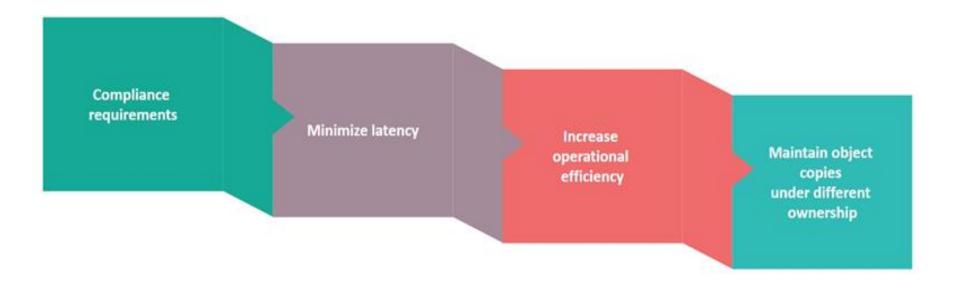
- ➤ It is a bucket-level feature that enables automatic, asynchronous copying of objects across bucket.
- In the object replicas in destination bucket = replicas of the objects in the source bucket.



Replicating objects in two different region with the same key name and metadata

Use Case of CRR





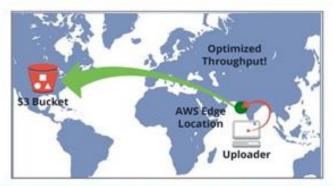




Transfer Acceleration In AWS S3



Amazon S3 Transfer Acceleration



- 01 Enables easy, fast and secure way of transferring files over a distance between the client and the bucket
- O2 The data is sent to S3 using 54 existing edge location with no changes required in firewall or software
- 03 It is faster because the data is routed to the S3 over an optimised routing network path
- O4 S3 transfer acceleration pricing is added in addition to the data transfer pricing and it is charges \$0.04/GB



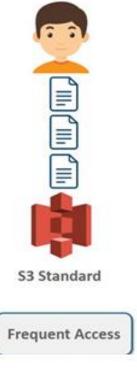


Choice Of Storage Classes On S3













S3 Standard



S3 standard

S3 Infrequent Access

Amazon Glacier



S3 Standard is used for frequent access of data

It is highly throughput and low latency performance

To automate the migration of objects life cycle policy is used

It is used for Content Distribution, Big Data analytics, Mobile and Gaming Applications





S3 standard

S3 Infrequent Access

Amazon Glacier



Used for data that is less frequently but requires rapid access when needed

It can be applied in object level as well as bucket level

Designed for large objects with minimum object storage

Same durability and throughput as S3 Standard

Amazon Glacier



S3 standard

S3 Infrequent Access

Amazon Glacier

An economical storage pricing service that provides secure, durable, and flexible storage for data backup and archival data



Unlimited amount of data can be stored in it

Flexible storing and retrieving Option

Single archive data can be as large as 40 terabyte

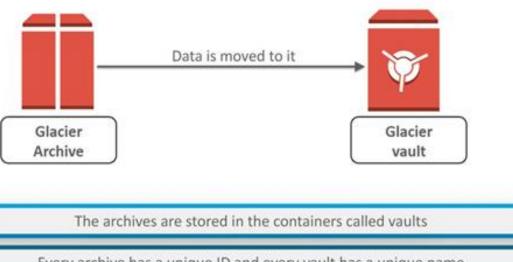






S3 Infrequent Access

Amazon Glacier



Every archive has a unique ID and every vault has a unique name

Updating of a existing archive is not possible, so we have to delete and upload it

The data which moved from S3 to Glacier, can be retrieved only using S3





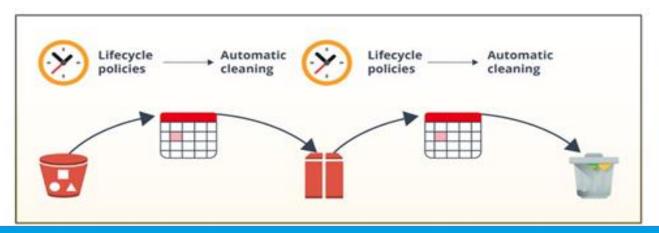
Characteristics	Standard	Standard - Infrequent Access	Glacier
Durability	99.99999999%	99.99999999%	99.99999999%
Availability	99.99%	99.90%	N/A
Minimum Object Size	No limit	128KB	No limit
Minimum Storage Duration	No minimum duration	30 Days	90 Days
First Byte Latency	milliseconds	milliseconds	4 hours
Retrieval Fee	No Fee	per GB retrieved	per GB retrieved

You can retrieve up to 5% of your average monthly Glacier storage for free each month, after this limit a retrieval fee starting at \$0.01 per gigabyte has to be paid



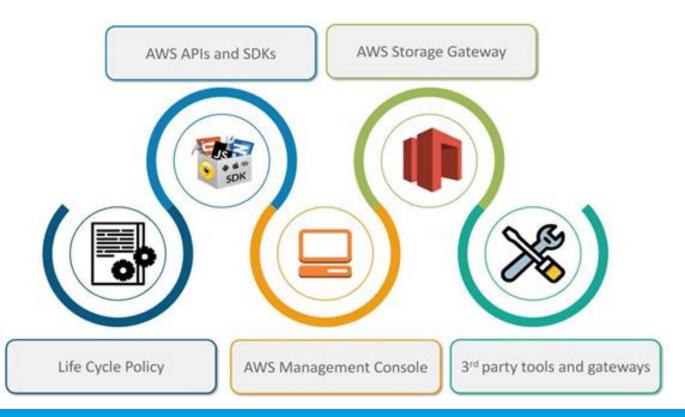
Life Cycle Policy of S3 Bucket

- ➤ *Life Cycle* policy is a set of rules that automate the defined actions (transition or expiration) on the group of objects present in S3.
- Helps to manage the storage space and saves the cost.
- It can be applied to all the object in the bucket or the subset of objects in the bucket.
- Also, be applied to the versions of the object.









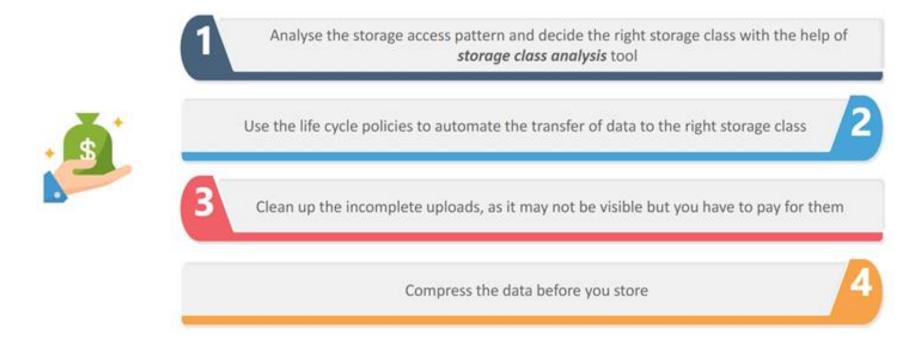




S3 Cost Optimization











CloudFront Content Delivery Network (CDN)

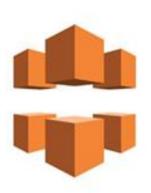


Content Delivery Network (CDN)

- ➤ A content delivery network (CDN) is a group of geographically distributed servers that speed up the delivery of web content by bringing it closer to where users are. Data centers across the globe use caching, a process that temporarily stores copies of files, so that you can access internet content from a web-enabled device or browser more quickly through a server near you.
- CDN services were created to solve the problem of network congestion caused by delivering rich web content, such as graphics and video over the internet - much like a traffic jam. Getting content from centrally located servers to individual users simply took too long.
- > By providing solutions for performance, availability, security, and intelligence, CDNS help the world's top companies and organizations do business successfully online.

CloudFront





CloudFront acts like an intermediate point between the main server and the end user

The data gets cached to speed up the delivery process of your static or dynamic content

Delivers the content through edge location

Reduces the no of networks that a users request must pass through

Increases the data transfer rate





AWS Snow Family

AWS Snow Family



- Applications are moving to the cloud faster today than ever before. A new category of applications requires increased capabilities and performance at the edge of the cloud, or even beyond the edge of the network.
- The Snow Family, comprised of AWS Snow Cone, AWS Snowball, and AWS Snowmobile, offers several physical devices and capacity points, most with built-in computing capabilities. These services help physically transport up to exabytes of data into and out of AWS.
- > Snow Family devices are owned and managed by AWS and integrate with AWS security, monitoring, storage management, and computing capabilities.

Snow Cone



➤ AWS Snow cone is the smallest member of the AWS Snow Family of edge computing, edge storage, and data transfer devices, weighing in at 4.5 pounds (2.1 kg) with 8 terabytes of usable storage. Snow cone is ruggedized, secure, and purpose-built for use outside of a traditional data center.

AWS Snow cone is currently available in the US East (N. Virginia), US West (Oregon), Europe (Ireland), Europe (Frankfurt), Asia Pacific (Sydney), and Canada (Central) Regions.



Snowball



- ➤ AWS Snowball, is an edge computing, data migration, and edge storage device that comes in two options.
- > Snowball Edge Storage Optimized devices provide both block storage.
- Amazon S3-compatible object storage, and 40 vCPUs of compute capacity coupled with 80 terabytes of usable block or Amazon S3-compatible object storage. They are well suited for local storage and large scale-data transfer.



Snowball

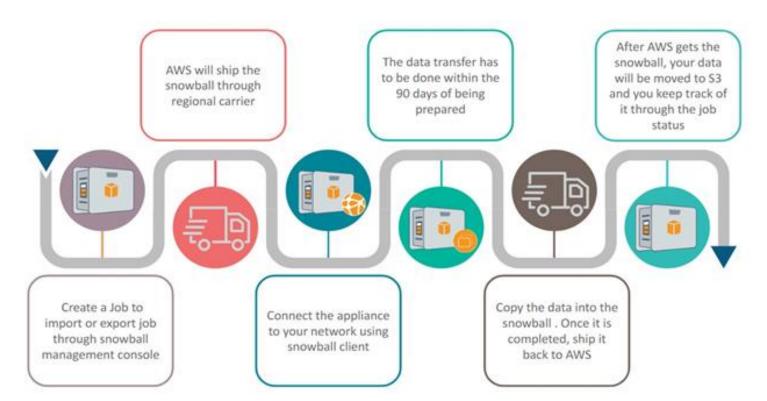


> Snowball Edge Compute Optimized devices provide 52 vCPUs, 42 terabytes of usable block or object storage and an optional GPU for use cases like advanced machine learning and full motion video analysis in disconnected environment.



Snowball: How Does It Work?





Snow Mobile



- 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 more secure, fast and cost effective.
- ➤ When your Snowmobile is on site, AWS personnel will work with your team to connect a removable, high-speed network switch from Snowmobile to your local network and you can begin your high-speed data transfer from any number of sources within your data center to the Snowmobile.

Snow Mobile

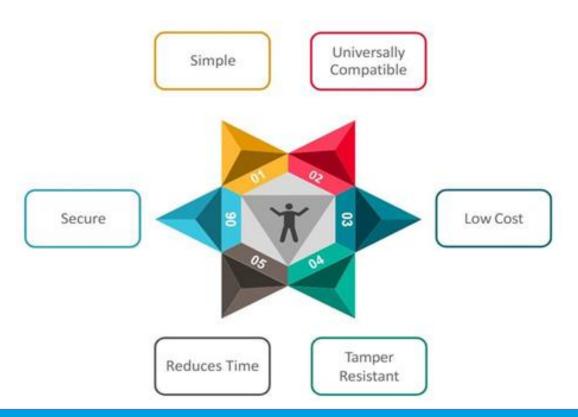


> Snowmobile uses multiple layers of security to help protect your data including dedicated security personnel, GPS tracking, alarm monitoring, 24/7 video surveillance, and an optional escort security vehicle while in transit. All data is encrypted with 256-bit encryption keys you manage through the AWS Key Management Service (KMS) and designed for security and full chain-of-custody of your data.











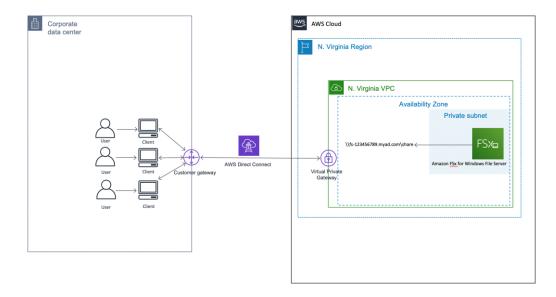


AWS FSx





> Amazon FSx makes it easy and cost effective to launch, run, and scale feature-rich, highperformance file systems in the cloud. It supports a wide range of workloads with its reliability, security, scalability, and broad set of capabilities.



Benefits of AWS FSx

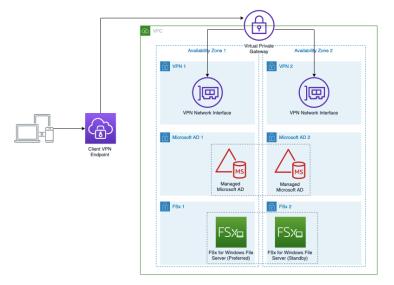


- Fully featured and fully managed Amazon FSx makes it easy to provide broadly-accessible and highly-performant file storage that enables a wide variety of use cases. Amazon FSx file systems support industry-standard protocols that offer connectivity to Linux, Windows, and macOS users and applications.
- ➤ **Highly available and protected -** To ensure high availability and durability, Amazon FSx automatically replicates your data within or across AWS Availability Zones to protect it from component failure, continuously monitors for hardware failures, and automatically replaces infrastructure components or switches to a stand-by file server in the event of a failure.
- ➤ **Cost effective** Amazon FSx lets you optimize your price and performance to support a broad spectrum of use cases, from small user shares to the most demanding compute-intensive workloads.
- ➤ **Hybrid-enabled -** Amazon FSx makes it easy to do more with your data. You can migrate and synchronize data from on premises to AWS and make it immediately available to a broad set of integrated AWS services.





- Migrate your workloads to the cloud
- > Build bleeding-edge machine learning, analytics, and HPC applications
- Simplify business continuity
- Increase development and test agility
- Accelerate Media & Entertainment workloads







AWS Storage Gateway



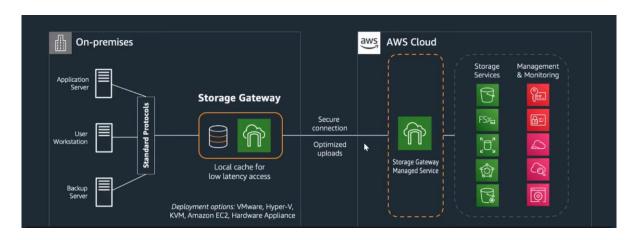


- ➤ AWS Storage Gateway is a hybrid cloud storage service that gives you onpremises access to virtually unlimited cloud storage. Customers use Storage Gateway to simplify storage management and reduce costs for key hybrid cloud storage use cases.
- Acts as a bridge between on-premises software appliance and cloud-based storage. Your applications connect to the service through a virtual machine or gateway hardware appliance using standard storage protocols, such as NFS, SMB, and iSCSI.



Storage Gateway

➤ The gateway connects to AWS storage services, such as Amazon S3, Amazon S3 Glacier, Amazon S3 Glacier Deep Archive, Amazon FSx for Windows File Server, Amazon EBS, and AWS Backup, providing storage for files, volumes, snapshots, and virtual tapes in AWS. The service includes a highly-optimized and efficient data transfer mechanism.







- > Storage Gateway helps you reduce cost, maintenance, and scaling challenges associated with managing on-premises storage environments.
- ➤ **Standard Storage Protocols:** Storage Gateway seamlessly connects to your local production or backup applications with NFS, SMB, iSCSI, or iSCSI-VTL, so you can adopt AWS Cloud storage without needing to modify your applications.
- ➤ **Fully Managed Cache**: The local gateway appliance maintains a cache of recently written or read data so your applications can have low-latency access to data that is stored durably in AWS.

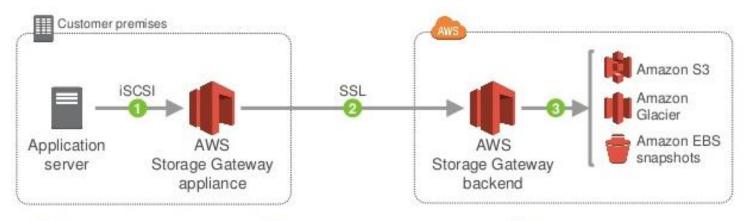




- ➤ **Optimized and Secure Data Transfer:** Storage Gateway provides secure upload of changed data and secure downloads of requested data, encrypting data in transit between any type of gateway appliance and AWS using SSL. Storage Gateway delivers end-to-end protection of customer data from the Storage Gateway in the enterprise network to the data residing in AWS. The service supports security features, access controls, and supplies compliances and certifications that address enterprise customers' real and perceived security concerns
- ➤ **High Availability on VMware**: Storage Gateway provides high availability on VMware through a set of health-checks integrated with VMware vSphere High Availability (VMware HA).

What happens when my application writes data?

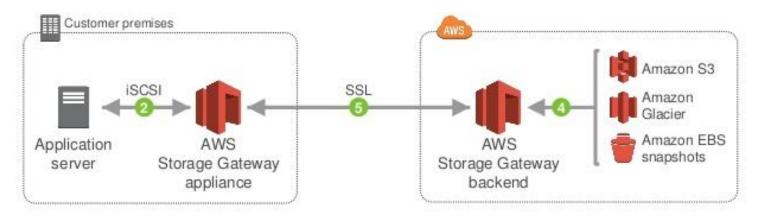




- Application writes data to AWS Storage
 Gateway using the iSCSI block protocol
- AWS Storage Gateway stores blocks locally, and asynchronously compresses and securely uploads the changed data
- Backend takes compressed data and stores in Amazon S3, Amazon Glacier, or Amazon EBS, with AES-256 encryption



What happens when my application reads data?



- Application reads data from AWS Storage Gateway using the iSCSI block protocol
- 2 AWS Storage Gateway returns requested data from local storage
- Data not in local storage is requested from backend
- 6 AWS Storage Gateway receives data from backend, decompresses, stores locally, and responds to the application
- Backend fetches compressed data from Amazon S3, Amazon Glacier, or Amazon EBS



AWS Storage Gateway Configuration

Gateway-stored volumes

Low-latency for all your data with point-in-time backups to AWS

Gateway-cached volumes

Low-latency for frequently used data with all data stored in AWS

Gateway-virtual tape library (VTL)

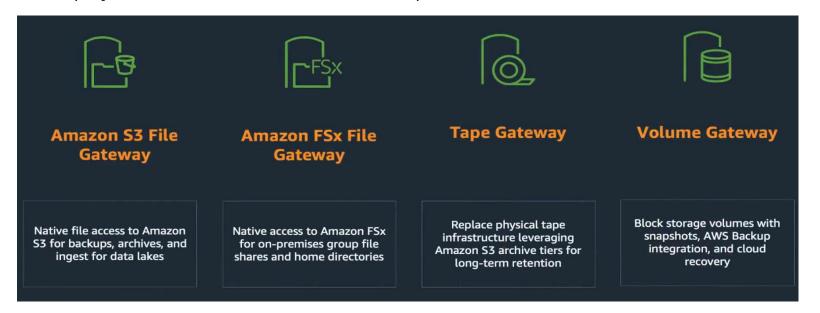
Replacement for on-premises physical tape infrastructure for backup and archive iSCSI block storage

iSCSI virtual tape storage



Storage Gateway

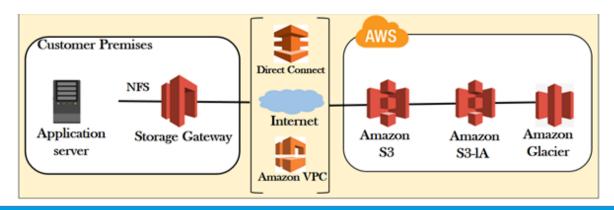
You can order and deploy a hardware appliance into your on-premises environment, deploy as a virtual machine into your on-premises environment (VMware ESXi, Microsoft Hyper-V, Linux KVM), or deploy in cloud as an Amazon Elastic Compute Cloud (EC2) instance.





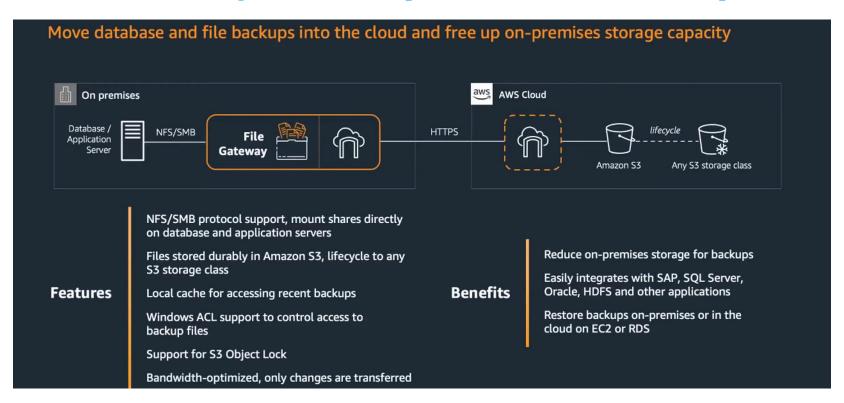
Amazon S3 File Gateway

- Amazon S3 File Gateway provides a seamless way to connect to the cloud in order to store application data files and backup images as durable objects in Amazon S3 cloud storage.
- Amazon S3 File Gateway offers SMB(Server Message Block) or NFS-based access to data in Amazon S3 with local caching. It can be used for on-premises data-intensive Amazon EC2-based applications that need file protocol access to S3 object storage.
- Used for Data Analytics ML.





File Gateway for on-premises backup





Benefits of Amazon S3 File Gateway

> Access to S3 capabilities

Storing data in Amazon S3 means you have access to the latest AWS developer tools and S3 API. Amazon S3 storage management tools like versioning, Cross-Region replication (CRR), and lifecycle management policies can lower the cost of long-term archiving, simplify audit and compliance requirements, and safeguard all your data, not just the data kept on-premises.

Object cloud storage on-demand

Amazon S3 File Gateway provides a seamless way to connect to the cloud in order to store backup images and durable objects in Amazon S3 cloud storage. Customers of all sizes and industries can use Amazon S3 to store and protect your data.



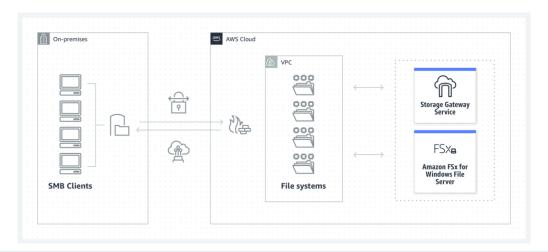
Benefits of Amazon S3 File Gateway

➤ Integration with AWS services for cloud-native use cases
File data can be stored as objects in S3 and use native AWS services to run big data analytics, serverless, high-performance computing (HPC), and machine learning (ML) to gain insights from your unstructured data sets, while still integrating with your on-premises environments.



Amazon FSx File Gateway

- > Amazon FSx File Gateway optimizes on-premises access to fully managed, highly reliable file shares in Amazon FSx for Windows File Server.
- Customers with unstructured or file data, whether from SMB(Server Message Block)-based group shares, or business applications, may require on-premises access to meet low-latency requirements. Amazon FSx File Gateway helps accelerates your file-based storage migration to the cloud to enable faster performance, improved data protection, and reduced cost.





Benefits of FSx S3 File Gateway

> Improved latency with local caching

Amazon FSx File Gateway provides low-latency performance for frequently accessed data. This enables seamless read and write activity when files are shared between their on-premises locations and the cloud.

Windows-native compatibility

With Amazon FSx File Gateway, you can seamlessly access the Windows-native SMB capabilities of Amazon FSx for Windows File Server such as NTFS, Active Directory integration, and data deduplication.

> Enhanced data protection

With access to Amazon FSx for Windows File Server, you can take advantage of automated, crash-consistent backups, and use AWS Backup for centralized backup and retention.



Choosing the Right Gateway

- You may be aware of Amazon S3 File Gateway (originally named File Gateway), and might now be wondering which type of workload is best suited for the two gateways:
- With **Amazon S3 File Gateway**, you can access data stored in Amazon Simple Storage Service (S3) as files, and it's also a solution for file ingestion into S3 for use in running object-based workloads and analytics, and for processing data that exists in on-premises files. More of a back up Solutions.



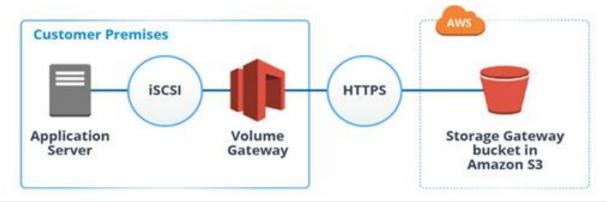
Choosing the Right Gateway

- Amazon FSx File Gateway, on the other hand, is a solution for moving networkattached storage (NAS) into the cloud while continuing to have low-latency, seamless access for your on-premises users. This includes two general-purpose NAS use-cases that use the SMB file protocol:
- 1. end-user home directories and
- 2. departmental or group file shares.
- Amazon FSx File Gateway supports multiple users sharing files, with advanced data management features such as access controls, snapshots for data protection, integrated backup, and more.





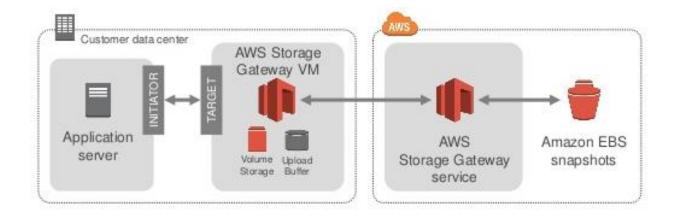
- ➤ Volume gateway is mounted to your on-premise application servers through Internet Small Computer System Interface (iSCSI) devices iSCSI devices enable you to access the network drive remotely through your system.
- >The data in the drive, is taken as a snapshot and stored in the S3.
- >We make use of this snapshots to create a volume and attached to the instance or an on-premise server.
- ➤ Different configuration of Volume gateway are:
- 1. Stored mode
- 2. Cached mode





Volume Gateway - Storage Mode

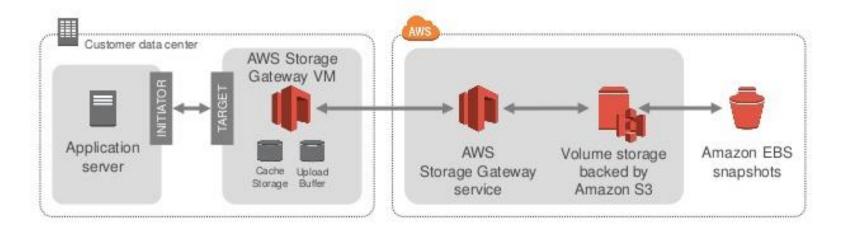
- Primary Data is stored on Premises.
- Asynchronous backup to AWS.
- Point-in-Time Backups stored as Amazon EBD Snapshots.
- Up-to 12 Volumes, up-to 16 TB for up-to 192 TB Gateway.





Volume Gateway - Cached Mode

- Primary data stored in AWS.
- Frequently accesses data cached on-premises.
- Point-in-time backups stored as Amazon EBS snapshots.
- Up to to 32 volumes, up to 32 TB each, for up to 1 PB per Gateway.





Tape Gateway

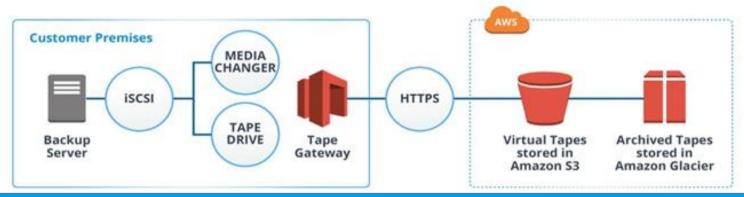
Tape Gateway enables you to replace using physical tapes on premises with virtual tapes in AWS without changing existing backup workflows.

- Virtual tapes stored in AWS
- Frequently accesses data cached on-premises
- Up to 1500 tapes, up to 2.5 TB each for up-to 150 TB per Gateway
- Unlimited number of tapes in virtual tape shelf (VTS).



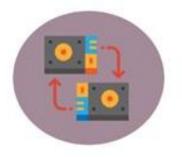
Tape Gateway

- Support many leading storage solution Veritas, Veeam, Dell EMC.
- ➤ It is mounted to your on-premise application servers through ISCSI devices, which Is preconfigured with tape drive and media changer .
- Tape drive helps to perform the I/O and seek permission operation on tape.
- Media changer helps to manage the tapes in Virtual Tape library (VTL).















Backup & archive

Disaster Recovery Data Migration or mirroring

Replace or expand on-premises storage

Lab Exercises







- Grant Public Access Manually and Via Bucket Policy
- Hosting a Static Website
- Versioning in AWS S3





Contents

1	Introduction	3
2	Documentation Links	5
3	Pre-requisite	6
4	Create an S3 Bucket	9
5	Grant Public Access Manually	25
6	Grant Public Access via Bucket Policy	33
7	Hosting A Static Website	48
8	Versioning In AWS S3	58
	Deleting S3 bucket	
10		
_		



S3 Cross Region Replication

Steps to perform Cross Region Replication







Contents

1	Introduction	3
2	Documentation Links	5
3	Pre-requisite	6
4	S3 Cross Region Replication	8
	4.1 Create a Source Bucket	9
	4.1 Create a Source Bucket	19
	4.3 Create a Rule for Replication	23
	4.4 Upload New Objects to Source S3 Bucket	31
	4.5 Test the Replication in our Destination Bucket	33
5	Delete/Cleanun	35
	5.1 Delete our Destination Bucket	35
	5.1 Delete our Destination Bucket	43
6		53





S3 Lifecycle Management on S3

Steps to set Lifecycle configuration in a Bucket





Contents

1	Introduction	3
2	Documentation Links	5
3	Prerequisites	6
	Set a lifecycle configuration on a bucket	
4	4.1 Add Sample Data to S3 Bucket	8
4	4.2 Create a Lifecycle Rule	12
5	Summary	23

Quiz



- **Q.** Which of these S3 storage classes is the most durable?
- A. S3
- B. S3-IA
- C. S3 One Zone-IA
- D. All these classes are equally durable

Answer: D.

Explanation: All S3 storage classes share the same durability: 11 9s(99.99999999%). That's often unintuitive, so it's best to recall that all S3 classes have durability in common and decrease in availability from S3 to S3-IA to S3 One Zone-IA.

Find Us





https://www.facebook.com/K21Academy



http://twitter.com/k21Academy



https://www.linkedin.com/company/k21academy



https://www.youtube.com/k21academy



https://www.instagram.com/k21academy/