# Day - 06 | Amazon S3 | AWS Cloud Practitioner Certification CLF-C02

Created on 2024-07-21 17:13
Published on 2024-07-23 20:04

## Amazon S3

☁ S3 Use cases

☁ Amazon S3 Overview - Buckets

☁ Amazon S3 Overview - Objects

☁ S3 Security

☁ S3 Bucket Policies

☁ Bucket settings for Block Public Access

☁ S3 Websites

☁ S3 - Versioning

☁ S3 Access Logs

☁ S3 Replication (CRR & SRR)

☁ S3 Storage Classes

✱ S3 Durability and Availability

✱ S3 Standard General Purpose

✱ S3 Storage Classes - Infrequent Access

ɘ S3 Standard Infrequent Access (S3 Standard-IA)

ɘ S3 One Zone Infrequent Access (S3 One Zone-IA)

✱ Amazon S3 Glacier Storage Classes

ɘ Amazon S3 Glacier Instant Retrieval

ɘ Amazon S3 Glacier Flexible Retrieval (formerly Amazon S3 Glacier)

ɘ Amazon S3 Glacier Deep Archive - for long term storage

✱ S3 Intelligent-Tiering

☁ S3 Object Lock & Glacier Vault Lock

☁ Shared Responsibility Model for S3

☁ AWS Snow Family

✱ Data Migrations with AWS Snow Family

✱ Time to Transfer

✱ Snowball Edge (for data transfers)

✱ AWS Snowcone

✱ AWS Snowmobile

✱ Snow Family - Usage Process

☁ What is Edge Computing?

☁ Snow Family - Edge Computing

☁ AWS OpsHub

☁ Hybrid Cloud for Storage

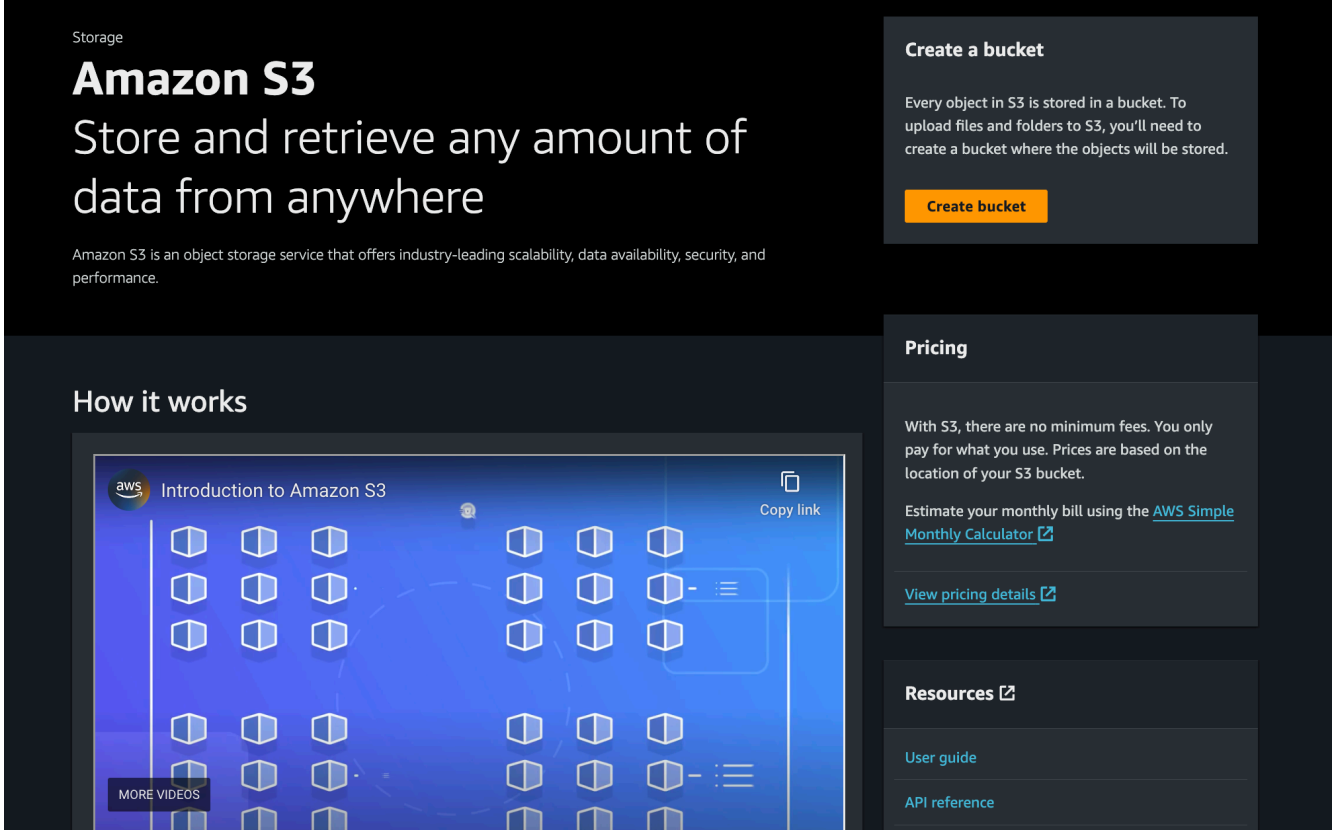
☁ AWS Storage Gateway

☁ Amazon S3 - Summary

---

## Amazon S3

Amazon Simple Storage Service (Amazon S3) is an object storage service that offers industry-leading scalability, data availability, security, and performance. Customers of all sizes and industries can use S3 to store and protect any amount of data for a range of use cases, such as websites, mobile applications, backup and restore, archive, enterprise applications, IoT devices, and big data analytics.



Amazon S3

## S3 Use cases

- Backup and storage
- Disaster Recovery
- Archive
- Hybrid Cloud storag e
- Application hosting
- Media hosting
- Data lakes & big data analytics
- Software delivery

- Static website

### Amazon S3 Overview - Buckets

**Buckets** are the fundamental containers in Amazon S3 for storing data (objects/files). Each bucket can hold an unlimited number of objects and serves as a namespace for objects within it. Buckets are identified by a globally unique name (across all regions all accounts). Buckets are defined at the region level. S3 looks like a global service but buckets are created in a region. Naming convention:

☆ No uppercase

☆ No underscore

☆ 3-63 characters long

☆ Not an IP

☆ Must start with lowercase letter or number

### Amazon S3 Overview - Objects

- Objects (files) have a Key

- The key is the FULL path:

✯ s3://my-bucket/my_file.txt

✯ s3://my-bucket/my_folder1/another_folder/my_file.txt

- The key is composed of **prefix** + **object name**

✯ s3://my-bucket/my_folder1/another_folder/my_file.txt

- There’s no concept of “directories” within buckets (*although the UI will trick you to think otherwise*)
- Just keys with very long names that contain slashes (′/′)
- Object values are the content of the body:

✯ Max Object Size is 5TB (5000GB)

✯ If uploading more than 5GB, must use “multi-part upload”

- Metadata (list of text key / value pairs – system or user metadata)

✯ Tags (Unicode key / value pair – up to 10) – useful for security / lifecycle

✯ Version ID (if versioning is enabled)

### S3 Security

- **User based** -

✯ IAM policies - which API calls should be allowed for a specific user from IAM console

- **Resource Based** -

✯ Bucket Policies - bucket wide rules from the S3 console - allows cross account

✯ Object Access Control List (ACL) – finer grain

✯ Bucket Access Control List (ACL) – less common

- **Note:** an IAM principal can access an S3 object if

☆ the user IAM permissions allow it OR the resource policy ALLOWS it

☆ AND there’s no explicit DENY

- **Encryption:** encrypt objects in Amazon S3 using encryption keys

## S3 Bucket Policies

Bucket policies are JSON-based access policy language that you can use to manage permissions for S3 buckets. They define what actions are allowed or denied for which principals (users) on the specified resources

Use S3 bucket for policy to:

☆ Grant public access to the bucket

☆ Force objects to be encrypted at upload

☆ Grant access to another account (Cross Account)

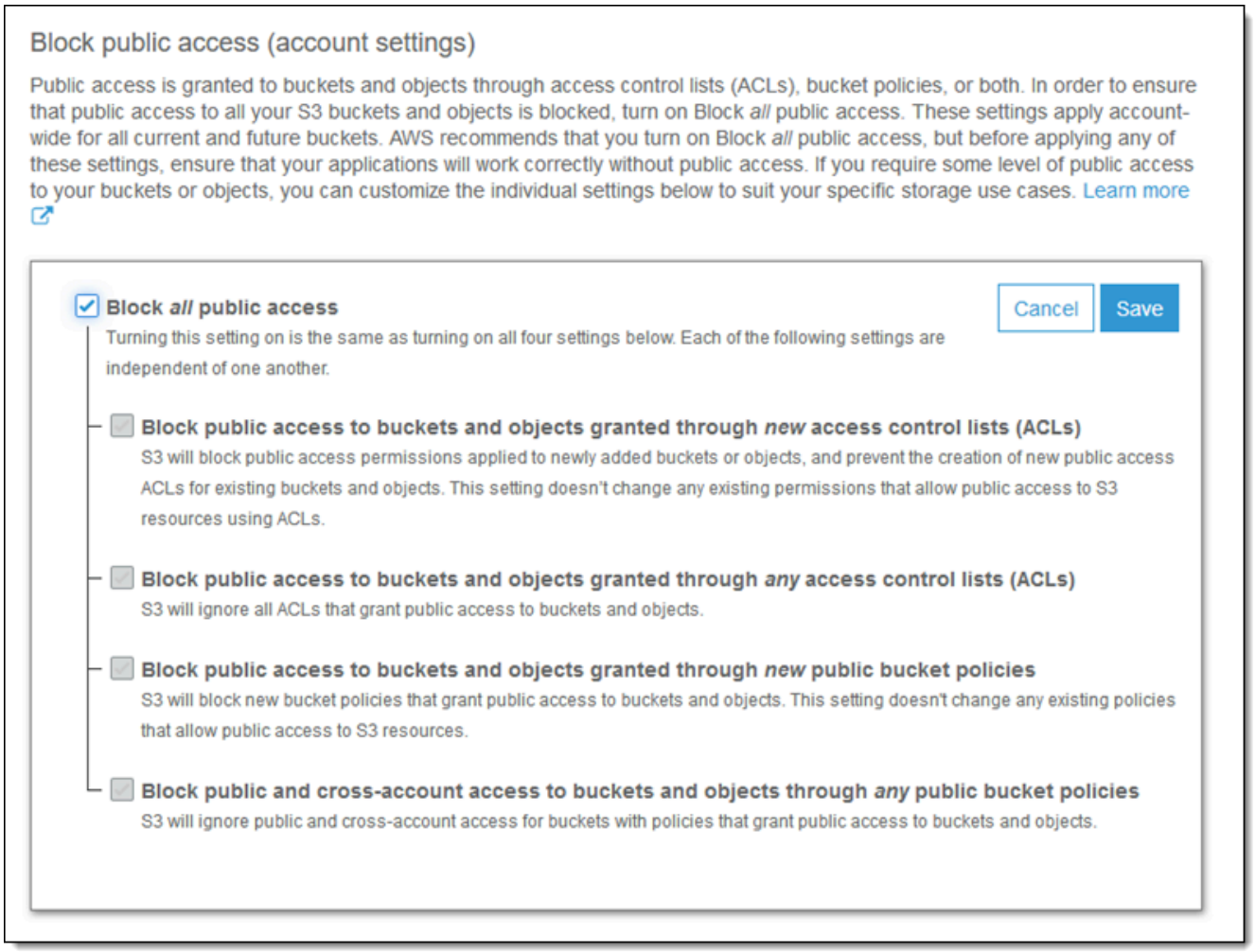
S3 Bucket Policies

### Bucket settings for Block Public Access

Amazon S3 provides settings to block public access to your S3 resources. This feature helps prevent unintended public access and helps you adhere to best practices for securing your S3 data.

- Block all public access: On
- These settings were created to prevent company data leaks
- If you know your bucket should never be public, leave these on

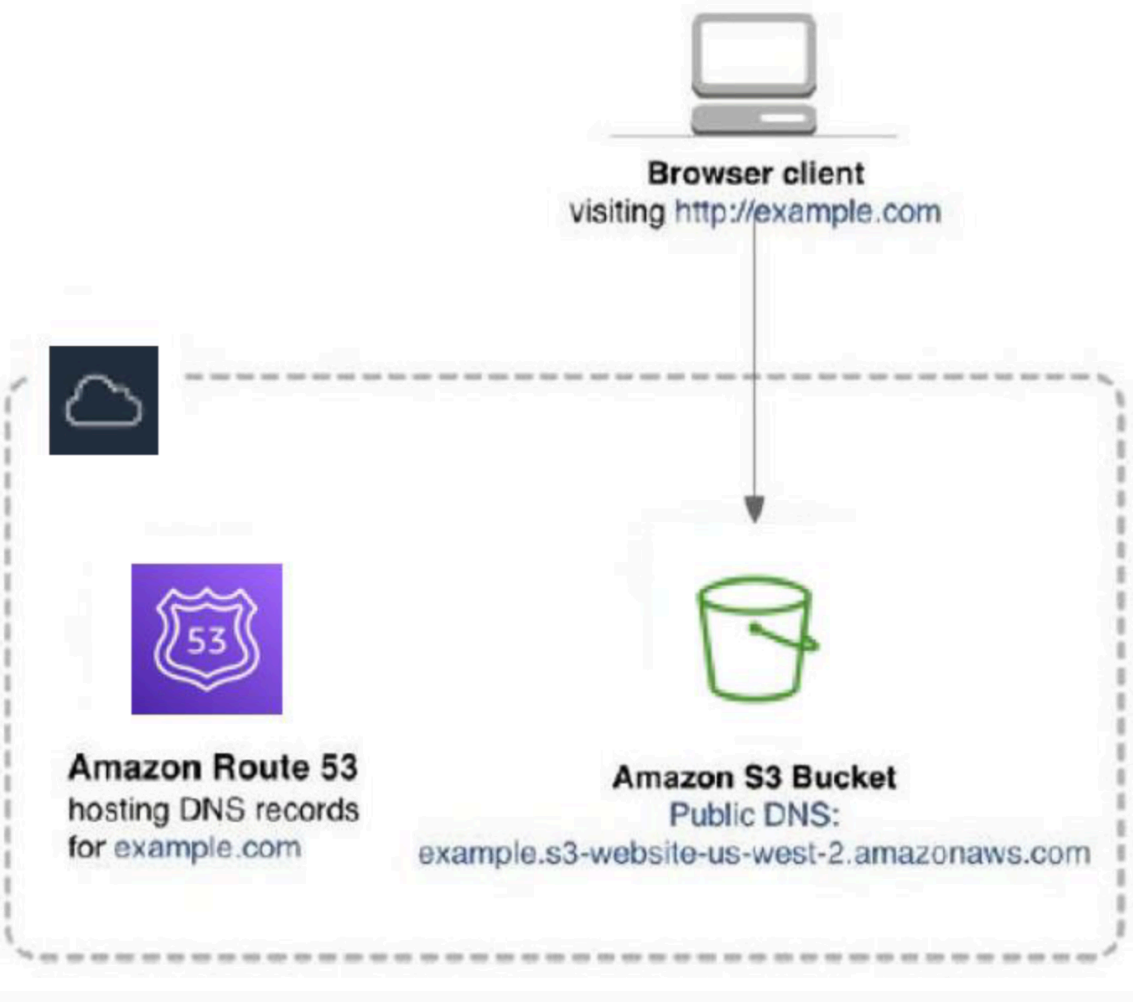
- Can be set at the account level




Bucket settings for Block Public Access

## S3 Websites

Amazon S3 can host static websites over www. You can configure your bucket to serve static web content, set up an index document, and manage error documents. **If you get a 403 (Forbidden) error, make sure the bucket policy allows public reads!**




S3 Websites

## S3 - Versioning

Versioning in Amazon S3 allows you to keep multiple versions of an object in the same bucket. This feature helps protect against accidental overwrites and deletions. Any file that is not versioned prior to enabling versioning will have version “null”. If the user suspends versioning, this will not delete the previous versions.

## S3 Access Logs

Amazon S3 provides the capability to log all access requests made to your S3 buckets. Access logs can be analyzed to track and audit usage patterns and permissions. Any request made to S3, from any account, authorized or denied, will be logged into another S3 bucket. That data can be analyzed using data analysis tools.

## S3 Replication (CRR & SRR)

- **Cross-Region Replication (CRR):** Automatically replicates objects across different AWS regions. **Use cases:** compliance, lower latency access, replication across accounts

- **Same-Region Replication (SRR):** Replicates objects within the same region. **Use cases:** log aggregation, live replication between production and test accounts.

- Must enable versioning in source and destination.

- Buckets can be in different accounts

- Copying is asynchronous

- Must give proper IAM permissions to S3

---

## S3 Storage Classes

- Amazon S3 Standard - General Purpose

- Amazon S3 Standard - Infrequent Access (IA)

- Amazon S3 One Zone - Infrequent Access

- Amazon S3 Glacier Instant Retrieval

- Amazon S3 Glacier Flexible Retrieval

- Amazon S3 Glacier Deep Archive

- Amazon S3 Intelligent Tiering

***NOTE -> Can move between classes manually or using S3 Lifecycle configurations***

### S3 Durability and Availability

- **Durability:**

✪ High durability (99.999999999%, 11 9's) of objects across multiple AZ

✪ If you store 10,000,000 objects with Amazon S3, you can on average expect to incur a loss of a single object once every 10,000 years

✪ Same for all storage classes

- **Availability:**

✪ Measures how readily available a service is

✪ Varies depending on storage class

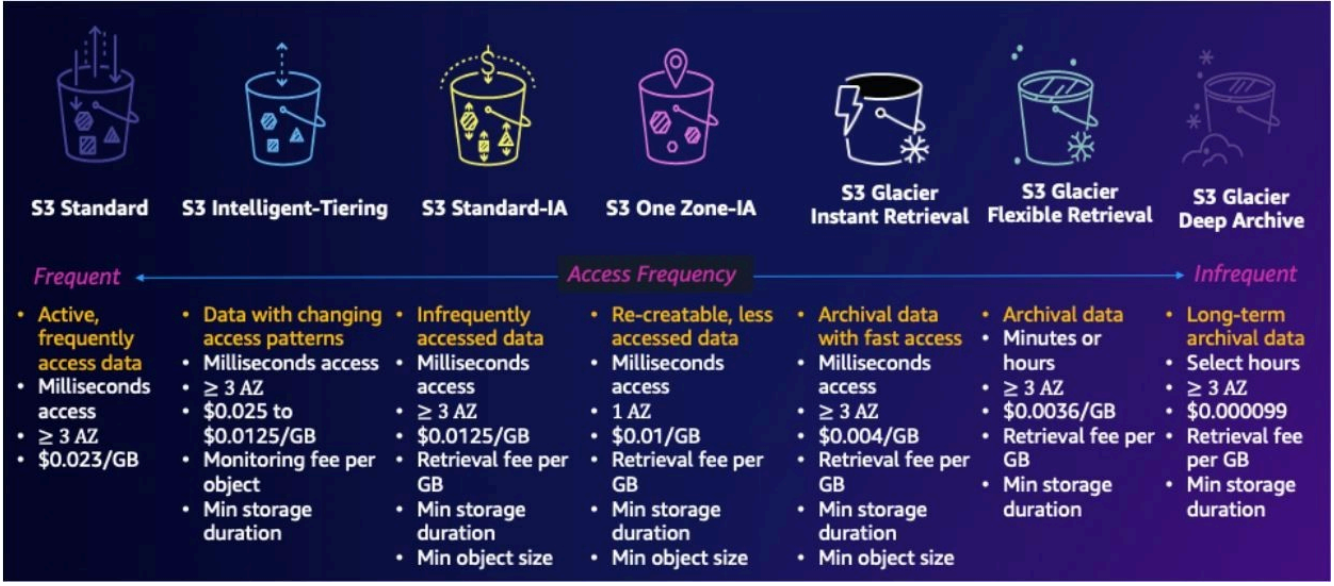
✪ Example: S3 standard has 99.99% availability = not available 53 minutes a year

---

## S3 Standard General Purpose

Designed for frequently accessed data and have **a**vailability that requires high throughput and low latency. This can sustain 2 concurrent facility failures.

**Use Cases:** Big Data analytics, mobile & gaming applications, content distribution…



S3 Standard General Purpose

## S3 Storage Classes - Infrequent Access

For data that is less frequently accessed, but requires rapid access when needed. Lower cost than S3 Standard.

### S3 Standard Infrequent Access (S3 Standard-IA)

99.9% Availability

**Use cases:** Disaster Recovery, backups

### S3 One Zone Infrequent Access (S3 One Zone-IA)

- High durability (99.999999999%) in a single AZ; data lost when AZ is destroyed
- 99.5% Availability
- Use Cases: Storing secondary backup copies of on-premise data, or data you can recreate

## Amazon S3 Glacier Storage Classes

- Low-cost object storage meant for archiving / backup
- Pricing: price for storage + object retrieval cost

Amazon S3 Glacier Storage Classes

### Amazon S3 Glacier Instant Retrieval

- Millisecond retrieval, great for data accessed once a quarter
- Minimum storage duration of 90 days

### Amazon S3 Glacier Flexible Retrieval (formerly Amazon S3 Glacier)

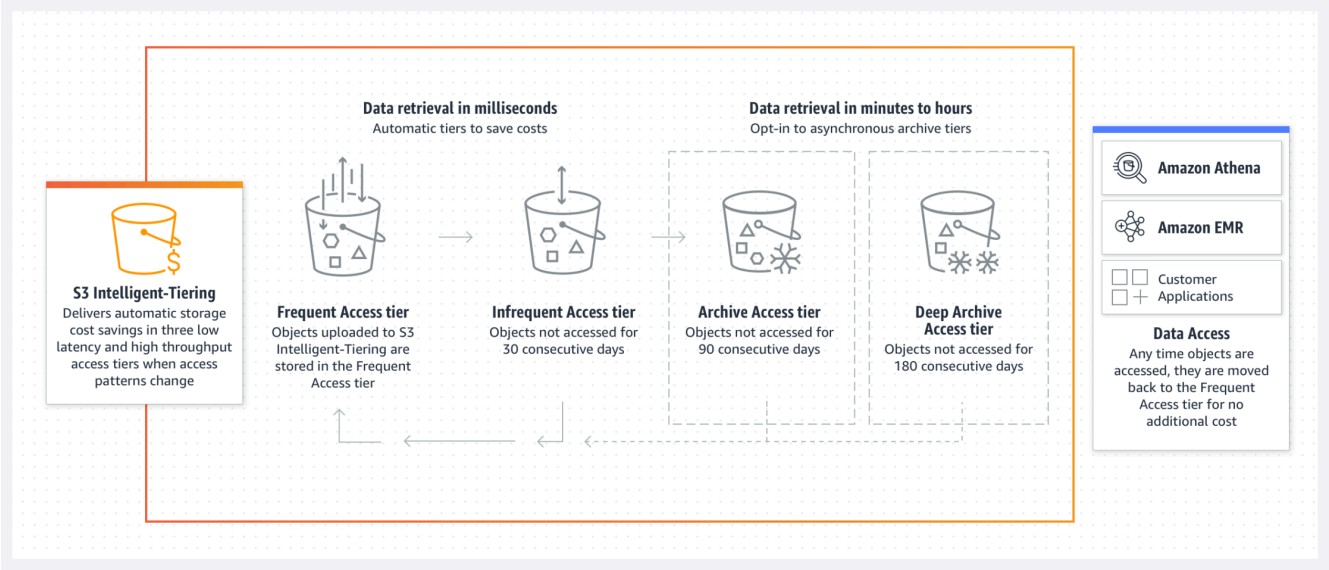
- Expedited (1 to 5 minutes), Standard (3 to 5 hours), Bulk (5 to 12 hours) – free
- Minimum storage duration of 90 days

### Amazon S3 Glacier Deep Archive - for long term storage

- Standard (12 hours), Bulk (48 hours)
- Minimum storage duration of 180 days

## S3 Intelligent-Tiering

This storage class automatically moves data between two access tiers (frequent and infrequent) to optimize costs based on changing access patterns.




S3 Intelligent-Tiering

- Small monthly monitoring and auto-tiering fee
- Moves objects automatically between Access Tiers based on usage
- There are no retrieval charges in S3 Intelligent-Tiering
- **Frequent Access tier (automatic):** default tier
- **Infrequent Access tier (automatic):** objects not accessed for 30 days
- **Archive Instant Access tier (automatic):** objects not accessed for 90 days
- **Archive Access tier (optional):** configurable from 90 days to 700+ days
- **Deep Archive Access tier (optional):** config from 180 days to 700+ days

## S3 Object Lock & Glacier Vault Lock

- **S3 Object Lock**: Prevents objects from being deleted or overwritten for a specified retention period.

- **Glacier Vault Lock**: Enforces compliance controls on individual S3 Glacier vaults. Adopt a WORM (Write Once Read Many) model. This Locks the policy for future edits (can no longer be changed).

## Shared Responsibility Model for S3

AWS and customers share responsibility for security and compliance:

- **AWS**: Responsible for the infrastructure security of the cloud.

- **Customers**: Responsible for managing their data, identity, and access management within S3.

| AWS | YOU |
| --- | --- |
| Infrastructure (global security, durability, availability, sustain concurrent loss of data in two facilities) | S3 Versioning, S3 Bucket Policies, S3 Replication Setup |
| Configuration and vulnerability analysis | Logging and Monitoring, S3 Storage Classes |
| Compliance validation | Data encryption at rest and in transit |

Shared Responsibility Model for S3

---

## AWS Snow Family

- Highly-secure, portable devices to collect and process data at the edge, and migrate data into and out of AWS

- **Data migration:** ☯ Snowcone, ☯ Snowball Edge, ☯ Snowmobile

- **Edge computing:** ☯ Snowcone, ☯ Snowball Edge

### Data Migrations with AWS Snow Family

- **AWS Snow Family:** Offline devices to perform data migrations. If it takes more than a week to transfer over the network, use Snowball devices!

- Challenges:

☼ Limited connectivity

☼ Limited bandwidth

☼ High network cost

☼ Shared bandwidth (can’t maximize the line)

☼ Connection stability

| Data | 100 Mbps | 1Gbps | 10Gbps |
|---|---|---|---|
| 10 TB | 12 days | 30 hours | 3 hours |
| 100 TB | 124 days | 12 days | 30 hours |
| 1 PB | 3 years | 124 days | 12 days |

Data Migrations with AWS Snow Family - Time to Transfer

## Data migration

### ☯ AWS Snowcone ☯

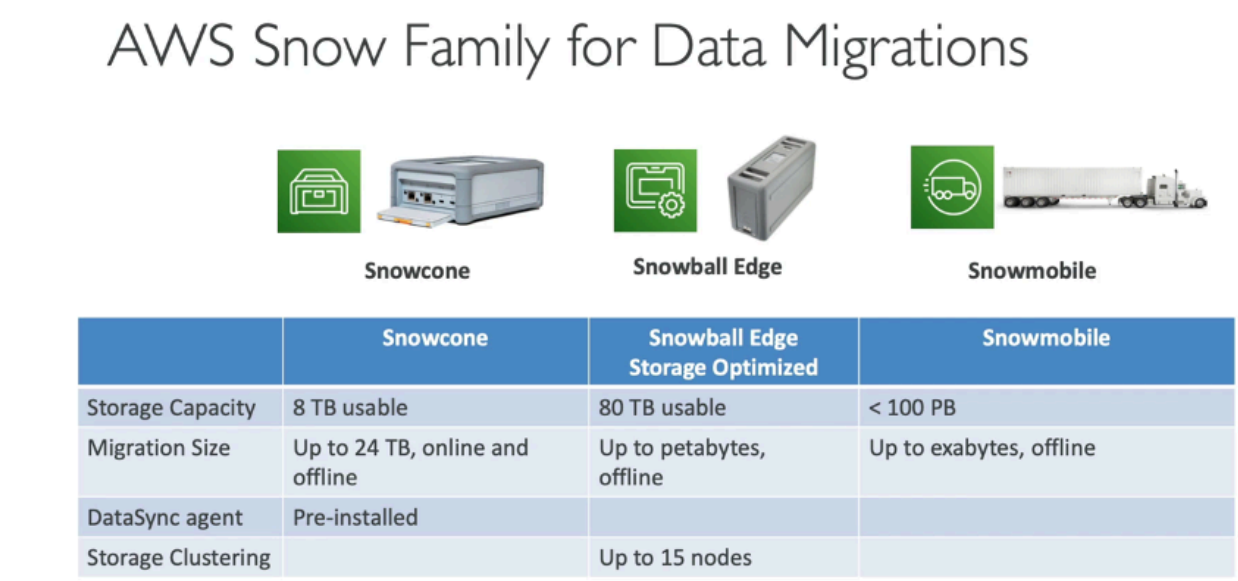
- Small, portable computing, anywhere, rugged & secure, withstands harsh environments
- Light (4.5 pounds, 2.1 kg)
- Device used for edge computing, storage, and data transfer
- **8 TBs of usable storage**
- Use Snowcone where Snowball does not fit (space-constrained environment)
- Must provide your own battery / cables
- Can be sent back to AWS offline, or connect it to internet and use **AWS DataSync** to send data

### ☯ Snowball Edge (for data transfers) ☯

- **Physical data transport solution:** move TBs or PBs of data in or out of AWS
- Alternative to moving data over the network (and paying network fees)
- Pay per data transfer job
- Provide block storage and Amazon S3-compatible object storage
- **Snowball Edge Storage Optimized** - 80 TB of HDD capacity for block volume and S3 compatible object storage
- **Snowball Edge Compute Optimized** - 42 TB of HDD capacity for block volume and S3 compatible object storage
- **Use cases:** large data cloud migrations, DC decommission, disaster recovery

## ☯ AWS Snowmobile ☯

- Transfer exabytes of data (1 EB = 1,000 PB = 1,000,000 TBs)
- Each Snowmobile has 100 PB of capacity (use multiple in parallel)
- High security: temperature controlled, GPS, 24/7 video surveillance
- ***Better than Snowball if you transfer more than 10 PB***




| | Snowcone | Snowball Edge Storage Optimized | Snowmobile |
|---|---|---|---|
| Storage Capacity | 8 TB usable | 80 TB usable | < 100 PB |
| Migration Size | Up to 24 TB, online and offline | Up to petabytes, offline | Up to exabytes, offline |
| DataSync agent | Pre-installed | | |
| Storage Clustering | | Up to 15 nodes | |

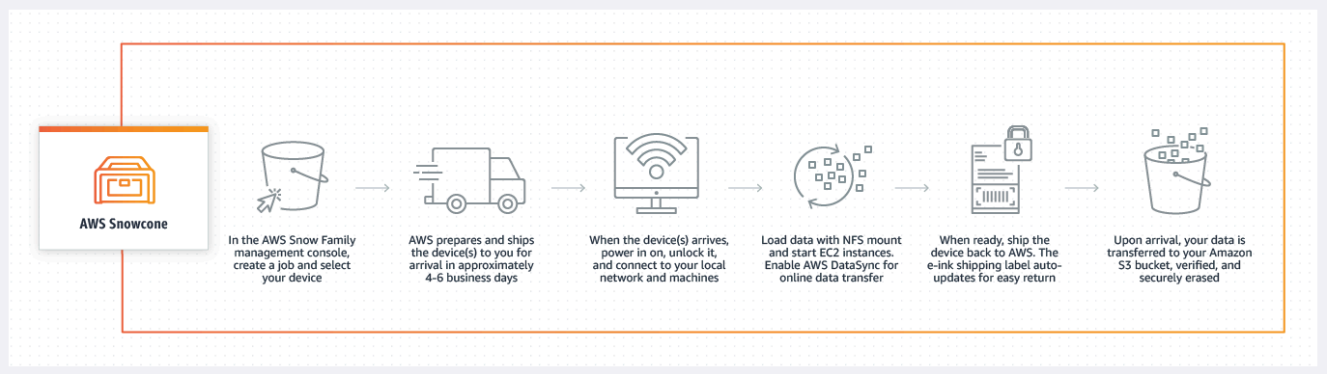
Snow Family Data Migrations

## Snow Family - Usage Process

1. Request Snowball devices from the AWS console for delivery
2. Install the snowball client / AWS OpsHub on your servers
3. Connect the snowball to your servers and copy files using the client
4. Ship back the device when you’re done (goes to the right AWS facility)

5. Data will be loaded into an S3 bucket

6. Snowball is completely wiped




Snow Family - Usage Process

## What is Edge Computing?

- Process data while it's being created on an edge location : A truck on the road, a ship on the sea, a mining station underground…
- These locations may have Limited / no internet access or Limited / no easy access to computing power
- We setup a **Snowball Edge / Snowcone** device to do edge computing
- Use cases of Edge Computing: Preprocess data, Machine learning at the edge, Transcoding media streams
- Eventually (if need be) we can ship back the device to AWS (for transferring data for example)

### Snow Family - Edge Computing

- **Snowcone (smaller)** 2 CPUs, 4 GB of memory, wired or wireless access, USB-C power using a cord or the optional battery

- **Snowball Edge – Compute Optimized** - 52 vCPUs, 208 GiB of RAM, Optional GPU (useful for video processing or machine learning), 42 TB usable storage

- **Snowball Edge – Storage Optimized** - Up to 40 vCPUs, 80 GiB of RAM, Object storage clustering available

- All: Can run EC2 Instances & AWS Lambda functions (using AWS IoT Greengrass)

- **Long-term deployment options:** 1 and 3 years discounted pricing

---

## AWS OpsHub

- **Historically**, to use Snow Family devices, you needed a CLI (Command Line Interface tool)

- Today, you can use **AWS OpsHub** (a software you install on your computer / laptop) to manage your Snow Family Device. Unlocking and configuring single or clustered devices. Transferring files. Launching and managing instances running on Snow Family Devices. Monitor device metrics (storage capacity, active instances on your device). Launch compatible AWS services on your devices (ex: Amazon EC2 instances, AWS DataSync, Network File System (NFS)).

## Hybrid Cloud for Storage

- AWS is pushing for ”hybrid cloud”Part of your infrastructure is on-premisesPart of your infrastructure is on the cloud

- This can be due toLong cloud migrationsSecurity requirementsCompliance requirementsIT strategy

- S3 is a proprietary storage technology (unlike EFS / NFS), so how do you expose the S3 data on-premise?
- AWS Storage Gateway!

## AWS Storage Gateway

- Bridge between on-premise data and cloud data in S3
- Hybrid storage service to allow on- premises to seamlessly use the AWS Cloud
- Use cases: disaster recovery, backup & restore, tiered storage
- Types of Storage Gateway: ✓ File Gateway ✓ Volume Gateway ✓ Tape Gateway

## Amazon S3 - Summary

- Buckets vs Objects: global unique name, tied to a region
- S3 security: IAM policy, S3 Bucket Policy (public access), S3 Encryption
- S3 Websites: host a static website on Amazon S3
- S3 Versioning: multiple versions for files, prevent accidental deletes
- S3 Access Logs: log requests made within your S3 bucket
- S3 Replication: same-region or cross-region, must enable versioning
- S3 Storage Classes: Standard, IA, 1Z-IA, Intelligent, Glacier, Glacier Deep Archive
- S3 Lifecycle Rules: transition objects between classes

- S3 Glacier Vault Lock / S3 Object Lock: WORM (Write Once Read Many)
- Snow Family: import data onto S3 through a physical device, edge computing
- OpsHub: desktop application to manage Snow Family devices
- Storage Gateway: hybrid solution to extend on-premises storage to S3

---

*Happy Learning !*