

What's In This Module



Core Services - Storage:

- Part 1: Amazon Elastic Block Store (Amazon EBS)
- Part 2: Amazon Simple Storage Service (Amazon S3)
- Part 3: Amazon Elastic File System (Amazon EFS)
- Part 4: Amazon Glacier

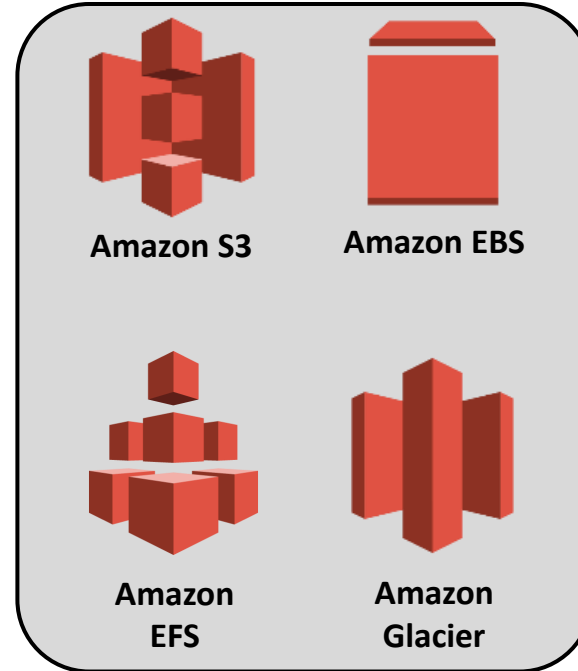
Core AWS Services



**Amazon
VPC**



Amazon EC2



Storage



Amazon RDS



**Amazon
DynamoDB**

Database



AWS IAM

AWS storage choices



Amazon EFS

File

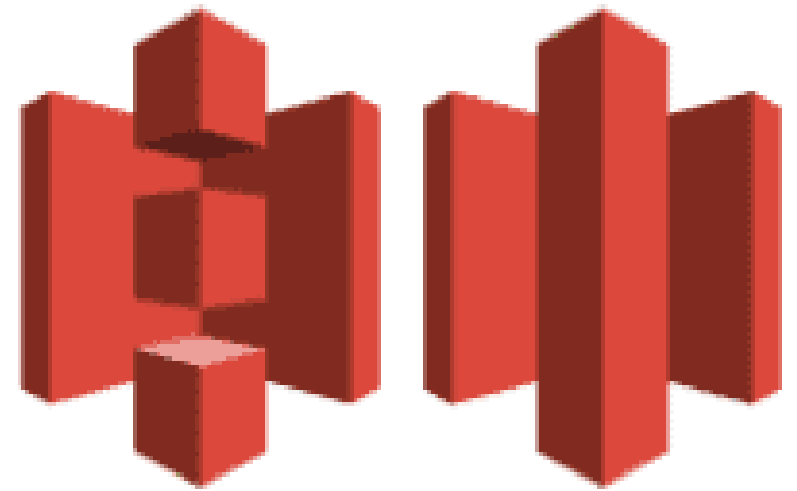


Amazon EBS



Amazon EC2
Instance Store

Block



Amazon S3

Amazon Glacier

Object

Block storage vs Object storage

Block Storage

- Suitable for transactional db, random read/write, and structured db storage
- Data is divided and stored in evenly sized blocks.
- Data blocks would not contain metadata.
- It only contains index of data block, doesn't care about the data in it.

Object Storage

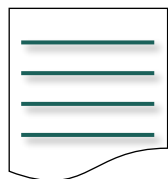
- Stores the files as a whole and doesn't divide them.
- An object has its data and metadata with a unique ID
- It can not be mounted as a drive
- Global unique ID is unique globally and it can be retrieved globally.

	Amazon EBS	Amazon EFS	Amazon EC2 Instance Store	Amazon S3	Amazon Glacier
Description	Persistent block storage volumes for use with Amazon EC2 instances in the AWS Cloud	A highly available and durable, multi-AZ, fully-managed file system	Block storage included at no additional charge with select Amazon EC2 instance types	Secure, durable, massively scalable object storage	A secure, durable, and extremely low-cost cloud storage service for data archiving and long-term backup
HPC Data Category	Hot	Hot	Hot	Warm	Cold
Attachment Method	Attached to EC2 instances	Attached from EC2 instances or from on-premises servers using Direct Connect	Attached to EC2 instances	API-based storage	API-based storage
Access Latency	Low	Medium	Low	Medium	High
Persistence	User-defined on startup	Available by default	Temporary	Available by default	Available by default
Estimated monthly bill for 100GB	\$15 for provisioned capacity	\$30 for used capacity	Included with select Amazon EC2 instance types	\$2.30 for used capacity	\$0.40 for used capacity
Concurrent access	Yes, via EC2-based shared file system	Yes, natively	Yes, via EC2-based shared file system	Yes, natively with appropriate IAM credentials	Not applicable
Recommended HPC Use	Use for high-IOPs and general purpose working storage; Lustre-compatible	Use as a shared file system for working storage	Use for read-often temporary working storage	Primary durable and scalable storage for HPC data	Use for long-term, lower-cost archival of HPC data

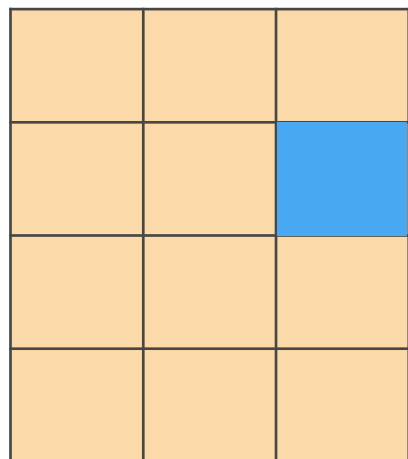


**Amazon Elastic Block Store
(Amazon EBS)**

AWS Storage Options: Block vs. Object Storage

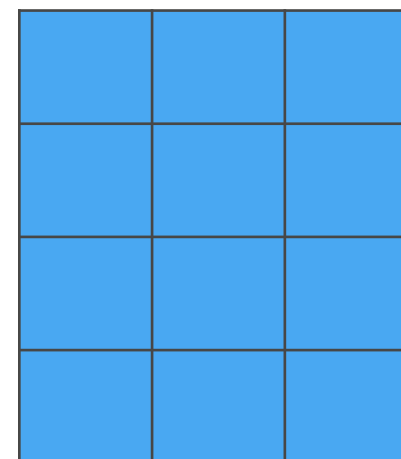


What if you want to **change** one character in a 1-GB file?



Block Storage

Change one block (piece of the file)
that contains the character



Object Storage

Entire file must be updated



Amazon EBS allows you to **create individual storage volumes** and **attach them** to an Amazon EC2 instance.

- Amazon EBS offers block-level storage.
- Volumes are automatically replicated within its Availability Zone.
- Can be backed up automatically to Amazon S3.
- Uses:
 - Boot volumes and storage for Amazon EC2 instances
 - Data storage with a file system
 - Database hosts
 - Enterprise applications



Amazon EBS Volume Types

	Solid-State Drives (SSD)		Hard Disk Drives (HDD)	
	General Purpose	Provisioned IOPS	Throughput-Optimized	Cold
Max volume size	16 TiB	16 TiB	16 TiB	16 TiB
Max IOPS/volume	16,000	64,000	500	250
Max throughput/volume	250 MiB/s	1,000 MiB/s	500 MiB/s	250 MiB/s



Amazon EBS Volume Types

Use Cases

Solid-State Drives (SSD)		Hard Disk Drives (HDD)	
General Purpose	Provisioned IOPS	Throughput-Optimized	Cold
<ul style="list-style-type: none">• Recommended for most workloads• System boot volumes• Virtual desktops• Low-latency interactive apps• Development and test environments	<ul style="list-style-type: none">• Critical business applications that require sustained IOPS• More than 16,000 IOPS or 250 MiB/s of throughput per volume• Large database workloads	<ul style="list-style-type: none">• Streaming workloads requiring consistent, fast throughput at a low price• Big data• Data warehouses• Log processing• Cannot be a boot volume	<ul style="list-style-type: none">• Throughput-oriented storage for large volumes of data that is infrequently accessed• Scenarios where the lowest storage cost is important• Cannot be a boot volume



Snapshots:

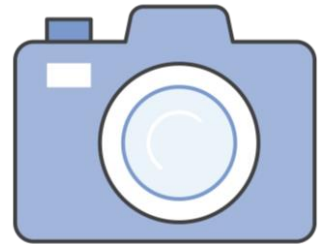
- Point-in-time snapshots
- Recreate a new volume at any time

Encryption:

- Encrypted Amazon EBS volumes
- No additional cost

Elasticity:

- Increase capacity
- Change to different types



1. Volumes:

- Amazon EBS volumes persist independently from the instance.
- All volume types are charged by the amount provisioned per month.

2. Input Output Operations per Second (IOPS):

- General Purpose (SSD)
 - Charged by the amount your provision in GB per month until storage is released
- Magnetic
 - Charged by the number of requests to volume
- Provisioned IOPS (SSD)
 - Charged by the amount you provision in IOPS (by % of day / month used)



3. Snapshots:

- Added cost of Amazon EBS snapshots to Amazon S3 is per GB-month of data stored.

4. Data Transfer:

- Inbound data transfer is free.
- Outbound data transfer charges are tiered.



Amazon EBS Features:

- Persistent and customizable block storage for Amazon EC2
- HDD and SSD types
- Replicated in the same Availability Zone
- Easy and transparent encryption
- Elastic volumes
- Back up using snapshots



An **object** storage service:



Amazon S3

- It stores massive (unlimited) amounts of unstructured data
- Data files are stored as objects in a **bucket** that you define
- 5 TB is the maximum file size of a single object
- All objects have a globally unique URL (universal namespace)
- All objects have a **key**, **version ID**, **value**, **metadata**, and **subresources**

Amazon S3 benefits



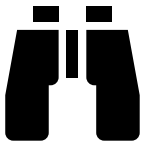
Durability

- It ensures data is not lost
- S3 Standard storage provides 11 9s (or 99.999999999%) of durability



Scalability

- It offers virtually unlimited capacity
- Any single object of 5 TB or less



Availability

- You can access your data when needed
- S3 Standard storage class is designed for four 9s (or 99.99%) availability



Security

- It offers fine-grained access control



Performance

- It is supported by many design patterns

Amazon S3 common usage patterns



Amazon S3



What problems can you solve by using Amazon S3?
You will now consider some **use cases**.

Amazon S3 use case 1:

Store and distribute web content and media

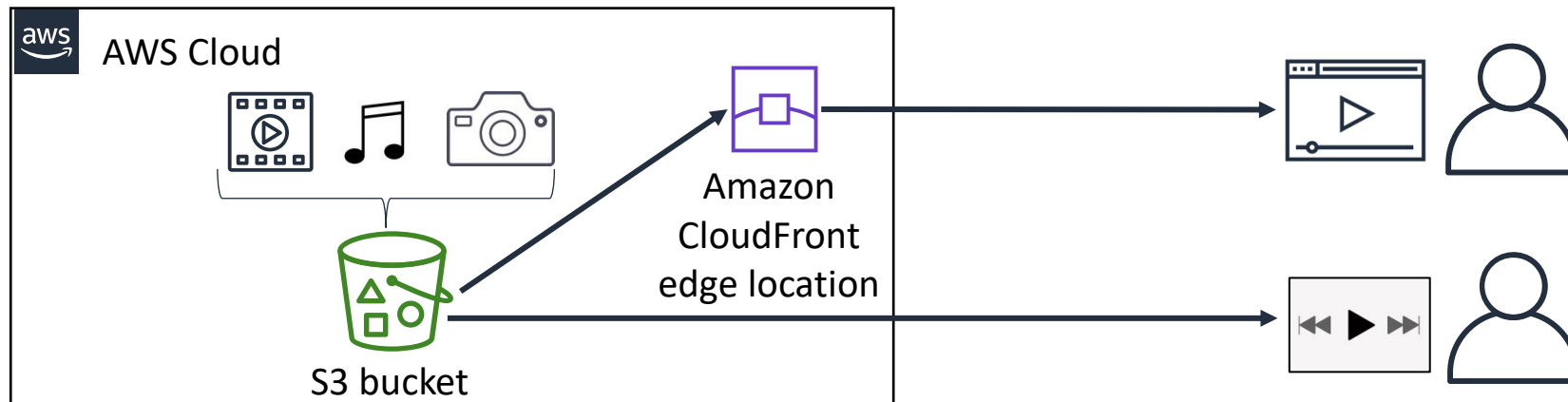
Build a redundant, scalable, and highly available infrastructure that hosts video, photo, or music uploads and downloads.



`https://<bucket-name>.s3.amazonaws.com`



`https://<bucket-name>.s3.amazonaws.com/video.mp4`



Securing Amazon S3 buckets and objects

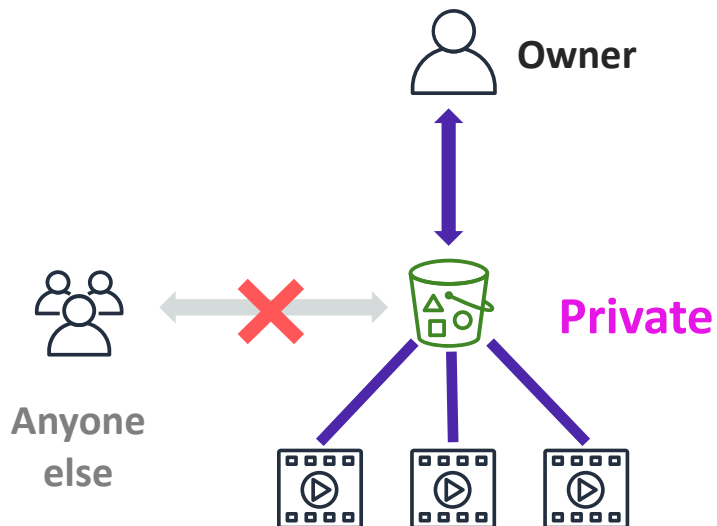
- Newly created S3 buckets and objects are **private** and **protected** by default
- When use cases must share Amazon S3 data –
 - Manage and control the data access
 - Follow the **principle of least privilege**
- Tools and options for controlling access to Amazon S3 data –
 - [Block Public Access](#) feature: It is enabled on new buckets by default, simple to manage
 - [IAM policies](#): A good option when the user can authenticate using IAM
 - [Bucket policies](#): You can define access to a specific object or bucket
 - [Access control lists](#) (ACLs): A legacy access control mechanism
 - [S3 Access Points](#): You can configure access with names and permissions specific to each application
 - [Presigned URLs](#): You can grant time-limited access to others with temporary URLs
 - [AWS Trusted Advisor](#) bucket permission check: A free feature



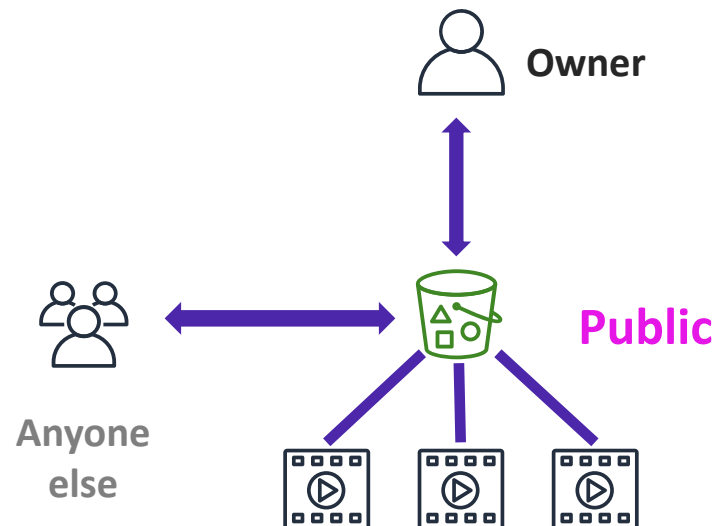
Three general approaches to configuring access

Configure the appropriate security settings for your use case on the bucket and objects.

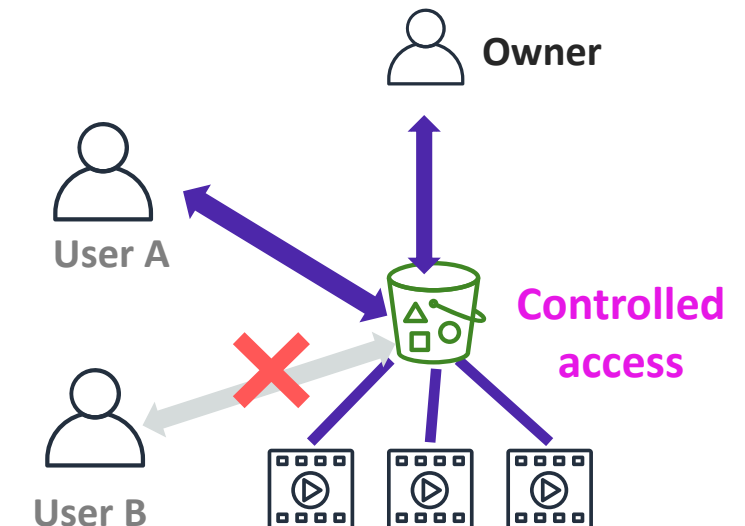
Default Amazon S3 security settings



Public access Amazon S3 security settings



Access policy applied to Amazon S3 security settings



Consider encrypting objects in Amazon S3

- **Encryption** encodes data with a **secret key**, which makes it unreadable

- Only users who have the secret key can decode the data
- Optionally, use AWS Key Management Service (AWS KMS) to manage secret keys



- *Server-side* encryption

- On the bucket, enable this feature by selecting the Default encryption option
- Amazon S3 encrypts objects before it saves the objects to disk, and decrypts the objects when you download them

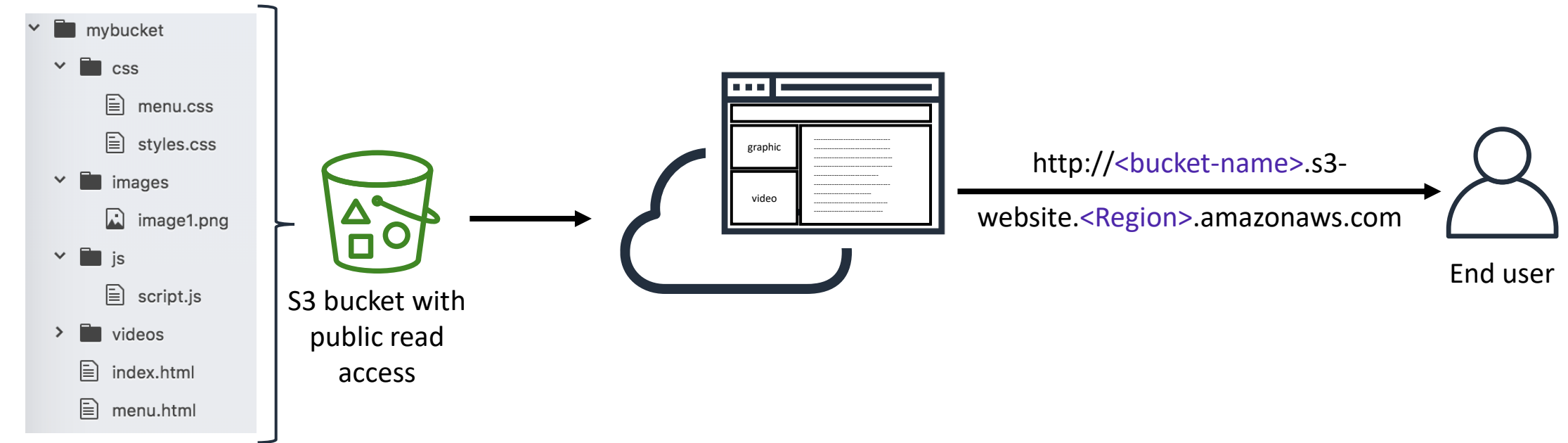


- *Client-side* encryption

- Encrypt data on the client side and upload the encrypted data to Amazon S3
- In this case, you manage the encryption process



Amazon S3 use case 2: Host static websites

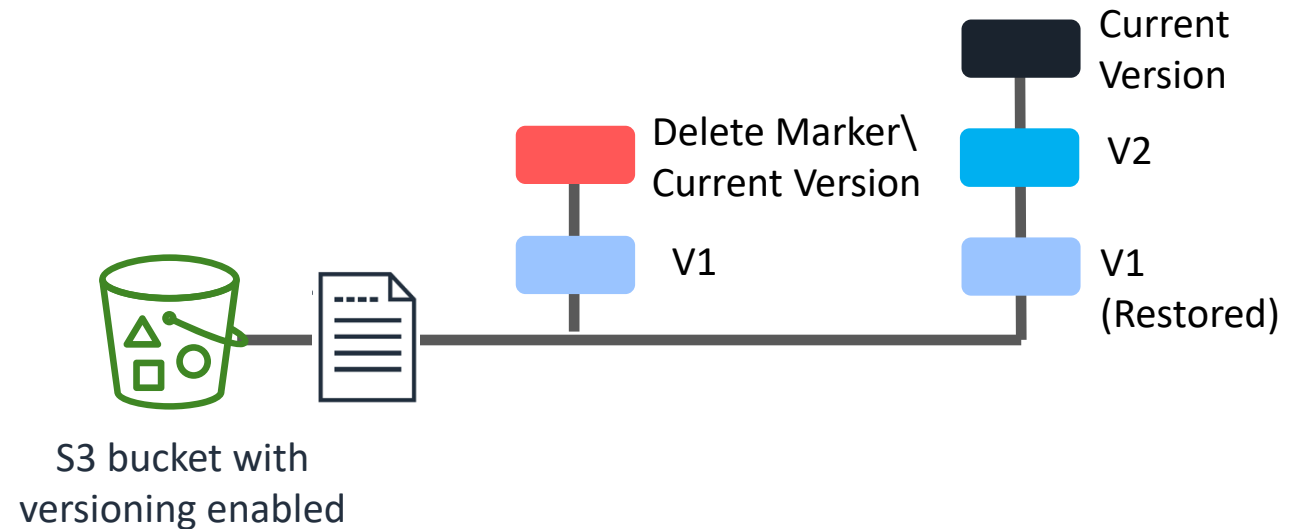


Example objects stored in the bucket configured for website hosting

Supports static content including HTML files, images, videos, and client-side scripts.

Amazon S3 best practice: Versioning

- Protects against accidental overwrites and deletes with no performance penalty
- Generates a new version with every upload
- Enables easy retrieval of deleted objects or rollback to previous versions
- Three possible states of an S3 bucket –
 1. *Default: Versioning not enabled*
 2. Versioning-enabled
 3. Versioning-suspended



S3 bucket versioning

- Can be used for data retention and archive
- Maintains all the versions of bucket
- Even after deleting the object, you can see the object and delete marker
- Deleted objects can be restored after deleting the delete marker
- Charged is applied to all S3 bucket versions

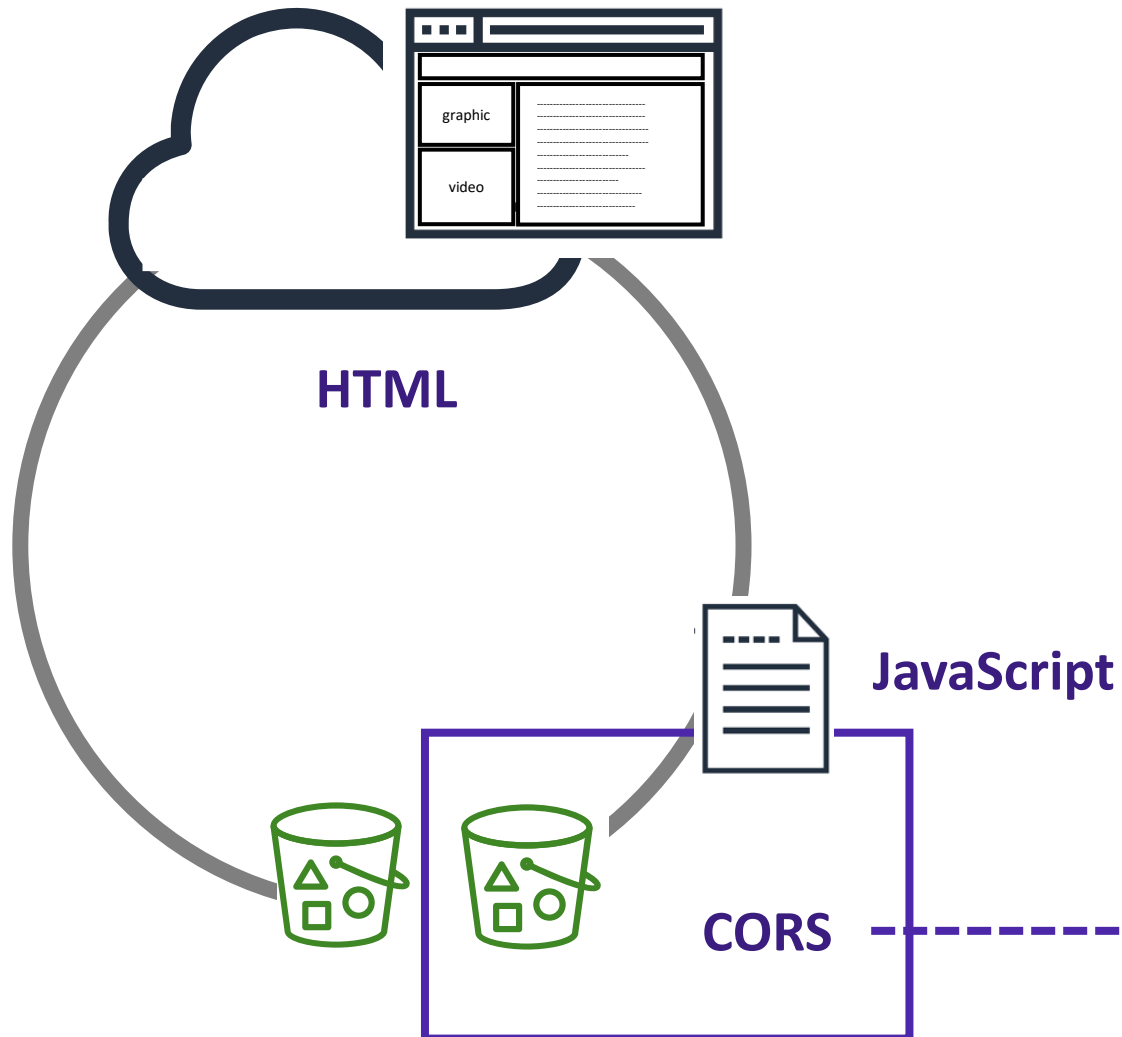


S3 bucket versioning

- Older version can be removed or sent to cheaper storage class using bucket lifecycle policies.
- Versioning applied to all objects inside the bucket
- Object existing before versioning enabled will have a NULL version ID
- Versions will not be updated while the versioning is suspended
- New uploaded object after the suspension with some existing bucket name will overwrite the previous content
- An object deletion in a suspended state will only delete the object with ID NULL.



Support for Cross-Origin Resource Sharing (CORS)



Example CORS configuration

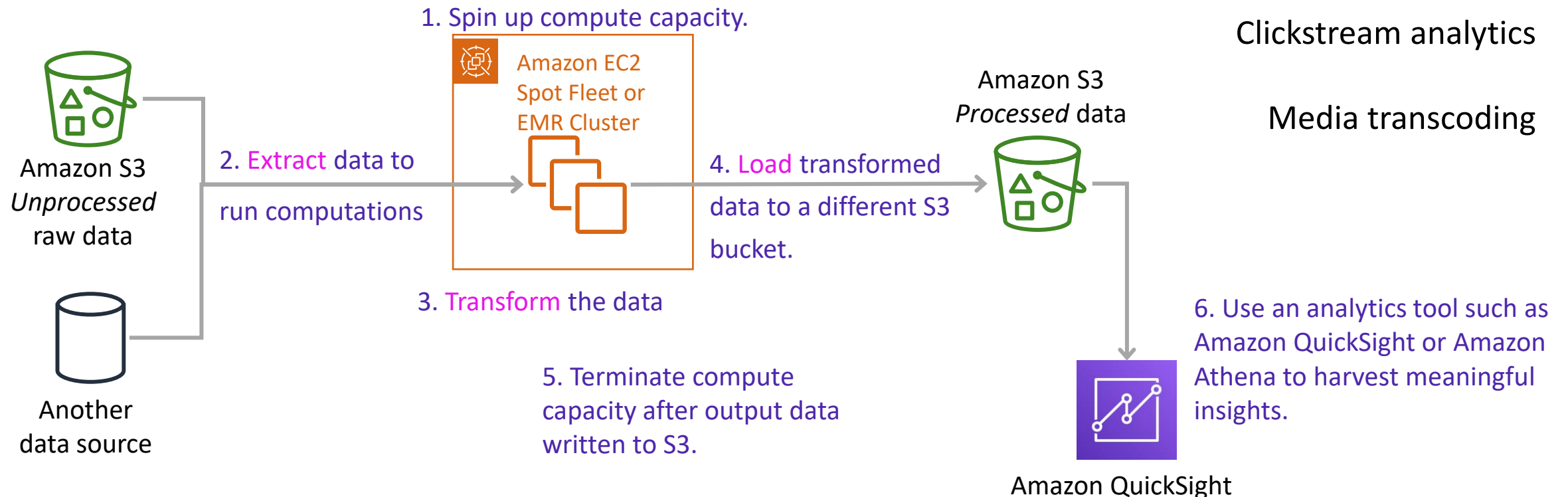
```
<CORSConfiguration>
  <CORSRule>

    <AllowedOrigin>http://www.example.com</AllowedOrigin>
    <AllowedMethod>PUT</AllowedMethod>
    <AllowedMethod>POST</AllowedMethod>
    <AllowedMethod>DELETE</AllowedMethod>
    <AllowedHeader>*</AllowedHeader>
    <MaxAgeSeconds>3000</MaxAgeSeconds>
    <ExposeHeader>x-amz-server-side-
encryption</ExposeHeader>
    <ExposeHeader>x-amz-request-id</ExposeHeader>
    <ExposeHeader>x-amz-id-2</ExposeHeader>
  </CORSRule>
</CORSConfiguration>
```

Amazon S3 use case 3: Data store for computation and analytics

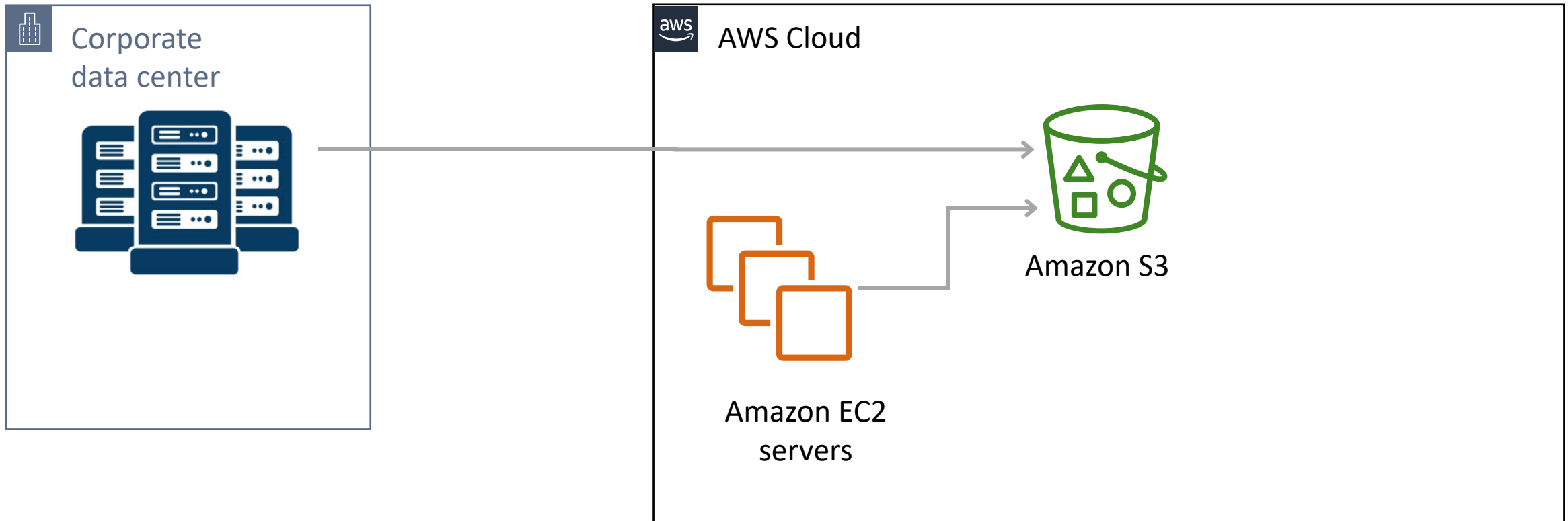
Data store for computation and large-scale analytics

Example data integration and preparation pattern



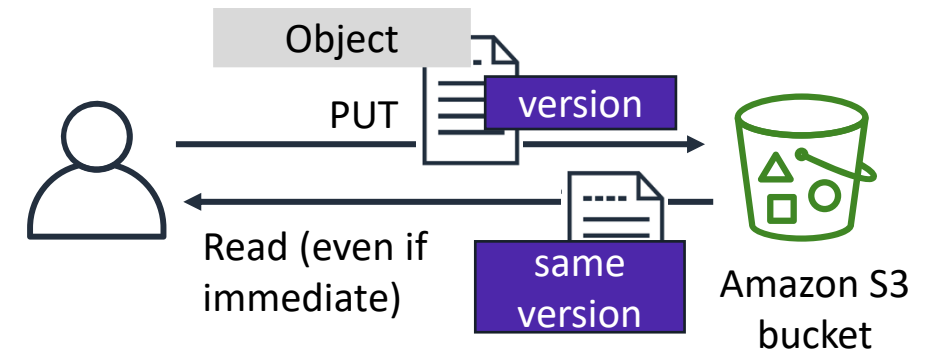
Amazon S3 use case 4: Back up and archive critical data

Amazon S3 as a data backup solution



Amazon S3 data consistency model

- Amazon S3 is **strongly consistent** for all new and existing *objects* in all Regions
 - Provides **read-after-write consistency** for all **GET**, **LIST**, and **PUT** operations on objects in S3 buckets
 - The consistency model offers an advantage for big data workloads
- Bucket configurations have an eventually consistent model

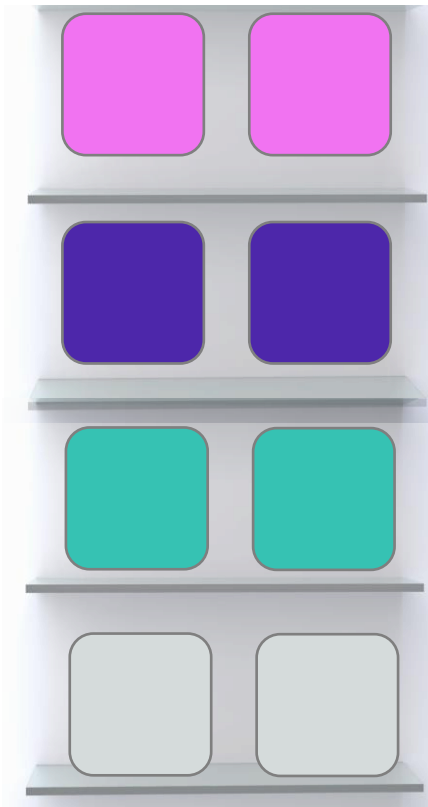






Key takeaways



- Buckets must have a **globally unique name** and are defined at the Region level
- Buckets are **private** and protected by default
- Amazon S3 security can be configured with IAM policies, bucket policies, access control lists, S3 access points, and presigned URLs
- Amazon S3 is **strongly consistent** for all new and existing objects in all Regions
- **5 TB** is the maximum size of a single object
- Amazon S3 is often used as a data store for computation and analytics, and as a backup and archive service for critical data

Amazon S3 and Amazon S3 Glacier storage classes



-  **S3 Standard:**
Frequently accessed data
-  **S3 Standard IA:**
Long-lived, infrequently accessed data
-  **S3 One Zone IA:**
Long-lived, infrequently accessed, non-critical data
-  **Amazon S3 Glacier or Deep Archive:**
Archiving rarely accessed data

Amazon S3 Intelligent Tiering

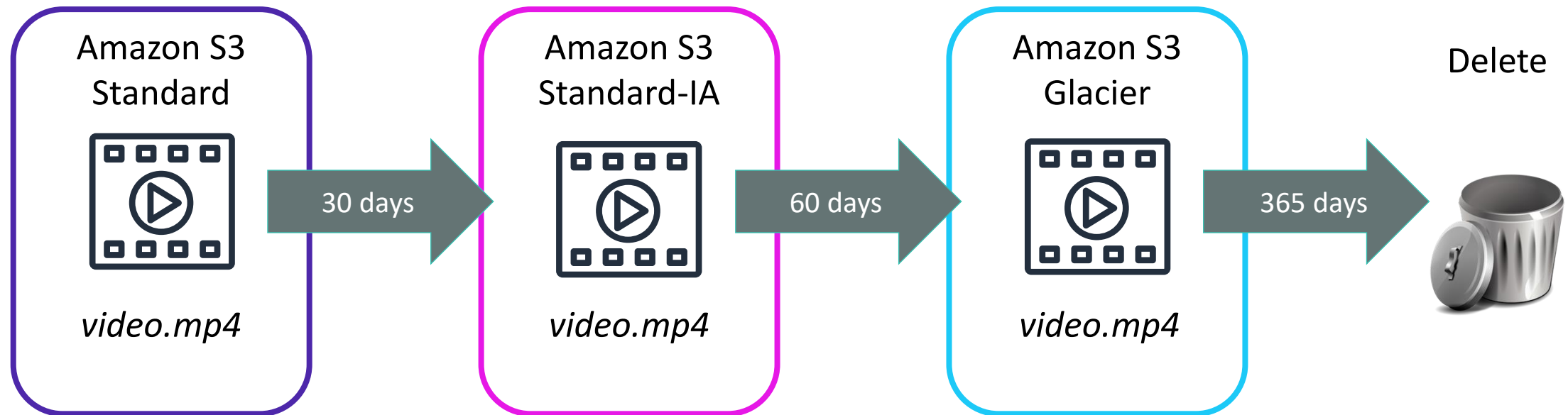
Automatically moves your objects between storage classes based on data access patterns.



 [Amazon S3 Storage Class](#) details

Amazon S3 lifecycle policies

Configure an **Amazon S3 Lifecycle Policy** to delete or move objects based on age.



Amazon S3 Storage Class: Standard

Offers high durability, availability, and performance

Object level storage for frequently accessed data

Durability - 99.999999999 %

Availability - 99.99 % per year

Ensure security of the data in transit as well as in rest.

Storage in charge is high but storage out charge is very less.

Largest size of the object that can be PUT is 5 GB.



Amazon S3 Storage Class: Standard IA

Stores infrequent and rapid access data

If data is deleted within 30 days of upload, you need to pay the cost for 30 days duration.

Backed with SLA

Durability - 99.999999999 %

Availability - 99.9 % per year

Ensure security of the data in transit as well as in rest.

Storage in charge is less but storage out charge is more than standard.



Amazon S3 Storage Class: Intelligent tiering

Optimize cost by moving data to the most cost-effective access tier

It stores data in two tiers

Durability
99.999999999%

Object less than 128 KB cannot be moved to IT

No retrieval and transition fee

Low latency and high performance

If an object is accessed, then automatic transition from infrequent to frequent access



Amazon S3 Storage Class: One zone IA

Less frequently used but rapid access is required

Data stored in single AZ

Lower cost option for IA data

Good choice for secondary backup data for on-premises data.

S3 lifecycle policy can be applied

Availability 99.5%

Data will be lost if AZ is crashed



Amazon S3 Storage Class: Glacier

Low-cost option
for data archiving

Three retrieval
options to keep
cost low

Object can be
uploaded directly
or via lifecycle
policy

Durability- 11 9's

Support security
for data in transit
and in rest

10 GB data can be
freely retrieved

Data will be
resilient if one AZ
is crashed



Amazon S3 Storage Class: Glacier deep archive

Cheapest
storage option

can save data
upto 10 years

Durability - 11
9's

Ideal alternative
to manage tape
libraries

Retrieval time <
12 hours

Storage cost is
upto 75% less
than Glacier

Availability -
99.9%



S3 bucket

Data is stored in buckets in S3

Maximum capacity of a bucket is 5 TB

Bucket ownership is non-transferrable

You can have 100 buckets per account

S3 bucket is region specific

Nested bucket can not be created

By default, only owner can access the bucket



S3 bucket naming convention

Should be globally unique

Cannot be changed once they are created

Once bucket is deleted, the name can be used again

Length should be at least 3 and no more than 63 character long

Should be the part of URL used to access bucket

Should not be an IP address

Consists of lowercase, number and hyphen

Multiple labels can be provided to the name

Consists of two parts: Region's endpoint/ bucket name



Amazon S3 costs



Pay only for use, including:

GBs of objects stored (per month). Different pricing per *Region* and per *storage class*.

Transfer OUT to other Regions or the internet.

PUT, COPY, POST, LIST, GET, SELECT, lifecycle transition, data retrieval requests.

No charge for:

Data transfers IN from the internet to Amazon S3.

Transfers between S3 buckets or from Amazon S3 to any services in the same AWS Region.

Transfer OUT to Amazon CloudFront.

DELETE and CANCEL requests.

Amazon S3: Storage Pricing

To estimate Amazon S3 costs, consider the following:

1. Types of storage classes:

- Standard Storage
 - 99.999999999% durability
 - 99.99% availability
- Standard-Infrequent Access (SIA)
 - 99.999999999% durability
 - 99.9% availability

2. Amount of storage:

- The number and size of objects
- Type of storage



3. Requests:

- The number of requests (**GET, PUT, COPY**):
- Type of requests
 - Different rates for GET requests than other requests.

4. Data Transfer:

- Pricing based on the amount of data transferred out of the Amazon S3 region
 - Data transfer in is free, but charges for data transfer out.



Key takeaways



- Amazon S3 storage classes include –
 - S3 Standard
 - S3 Standard-IA
 - S3 One Zone-IA
 - S3 Intelligent-Tiering
 - S3 Glacier
 - S3 Glacier Deep Archive
- An Amazon S3 lifecycle policy can delete or move objects to less expensive storage classes based on age
- Transferring data in from the internet to Amazon S3 is free, but transferring out to other Regions or to the internet incurs a fee

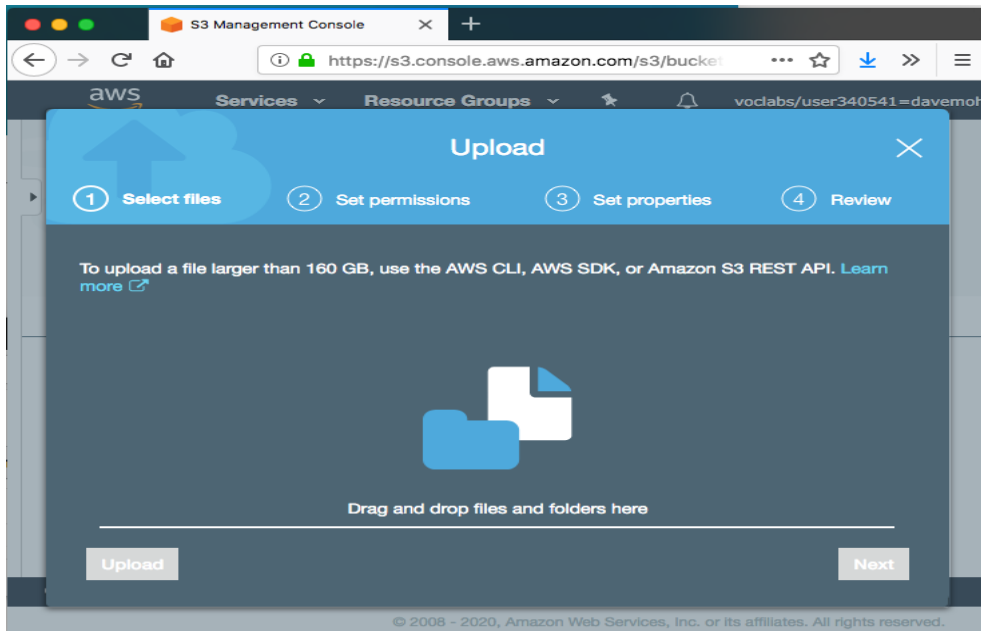
Moving data to and from Amazon S3

Moving objects to Amazon S3



AWS Management Console

Upload or download by using a browser.



AWS Command Line Interface

Upload or download from a terminal command prompt or in a call from a script.

- Example upload command:

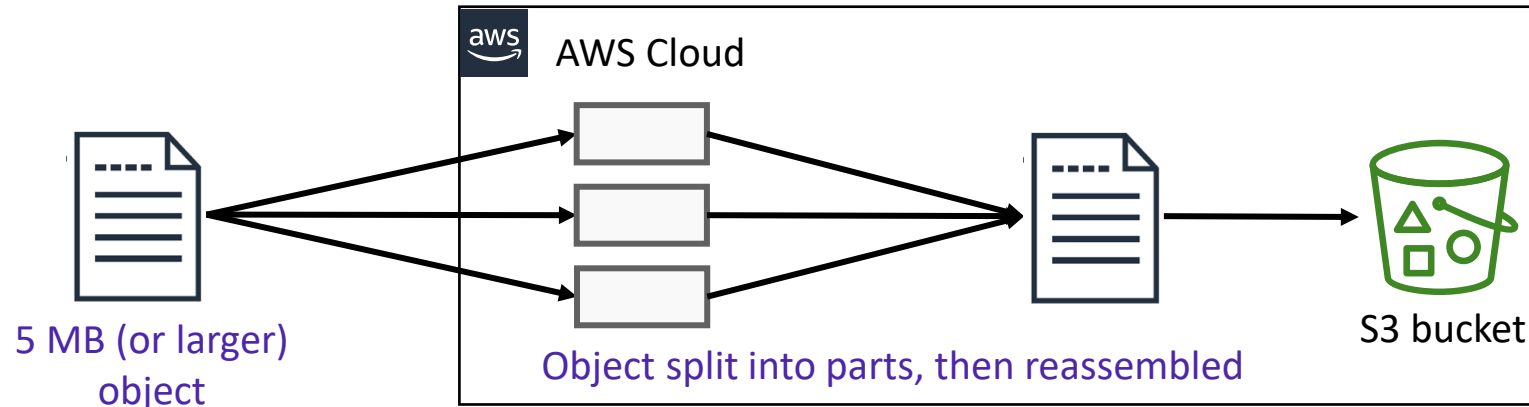
```
$ aws s3 cp test.txt \s3://AWSDOC-EXAMPLE-BUCKET/test.txt
```



AWS Tools and SDKs

Move objects programmatically by using AWS tools or SDKs.

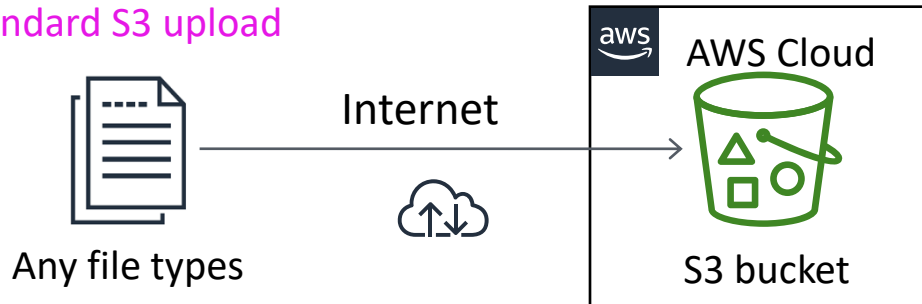
Multipart upload



- Files can be uploaded by using the Multipart Upload API
 - You can upload a single object as a set of parts
 - Each part is a contiguous portion of the object's data
 - After all parts of your object are uploaded, Amazon S3 assembles these parts and creates the object
- Typically only used for files larger than 100 MB
- Advantages –
 - Quick recovery from network issues: If transmission of any part fails, only need to retransmit that part
 - Ability to pause and resume object uploads
 - Improved throughput: Upload parts in parallel to improve throughput

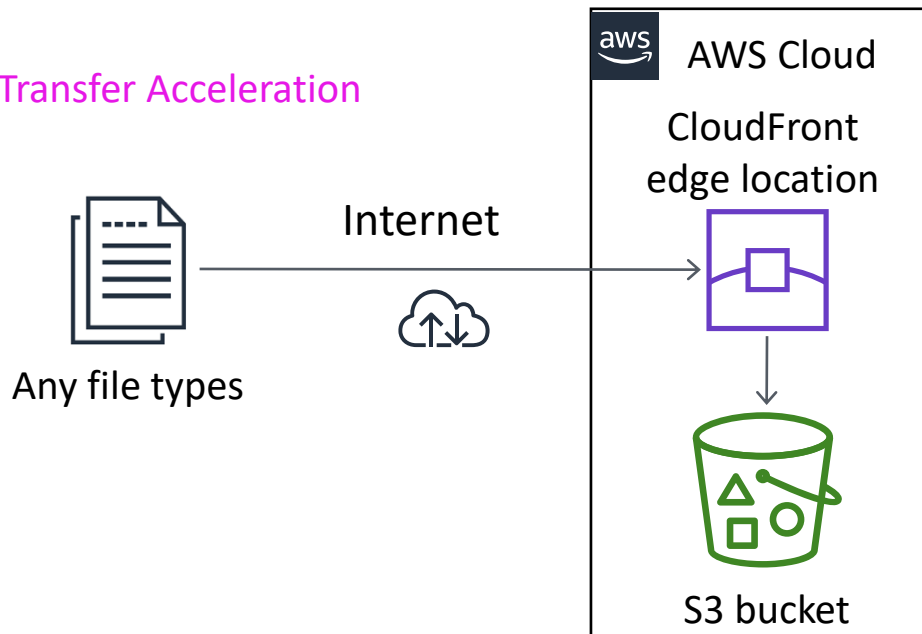
Amazon S3 Transfer Acceleration

Standard S3 upload



versus

S3 Transfer Acceleration



- Accelerates Amazon S3 data transfers
- Uses optimized network protocols and the AWS edge infrastructure
- Typical speed improvement:
 - 50–500% for cross-country transfer of larger objects
 - Can go even higher under certain conditions
- [Amazon S3 Transfer Acceleration Speed Comparison Tool](#)
 - Shows speed advantage gained (by Region)

Moving large amounts of data into Amazon S3: AWS Snowball



AWS Snowball

Petabyte-scale data transport



AWS
Snowball

- Can transport multiple terabytes of data into or out of Amazon S3
 - Multiple devices can be used to transfer petabytes
- Addresses concerns of large data transfers (network costs, transfer times, security)
 - *Example:* To transfer 10 petabytes (10 million GB) over the internet with a 10 Gbps upload speed would take over 100 days
- To use –
 - Create a job in the AWS Management Console and a Snowball will be shipped to you.
 - Attach to your local network, then download and run the Snowball Client
 - Select the file directories to transfer (encrypted) to the device
 - Ship the device back and track the status

Moving large amounts of data into Amazon S3: AWS Snowmobile



AWS Snowmobile
Exabyte-scale data transport



**AWS
Snowmobile**

- A 45-foot-long (13.7 meters) shipping container, pulled by a semi-trailer truck
- Can transfer up to 100 PB per Snowmobile
- Offers multiple layers of security –
 - Dedicated security personnel
 - GPS tracking, alarm monitoring, 24/7 video surveillance
 - Optional escort security vehicle while in transit
 - Data encrypted with 256-bit encryption keys

Key takeaways



- The **S3 multipart upload** option is a good option for files larger than 100 MB and in situations where network connectivity might be inconsistent
- **Amazon S3 Transfer Acceleration** uses edge locations and can significantly increase the speed of uploads
- **AWS Snowball** provides a way to transfer *petabytes* of data, and **AWS Snowmobile** provides a way to transfer *exabytes* of data to AWS



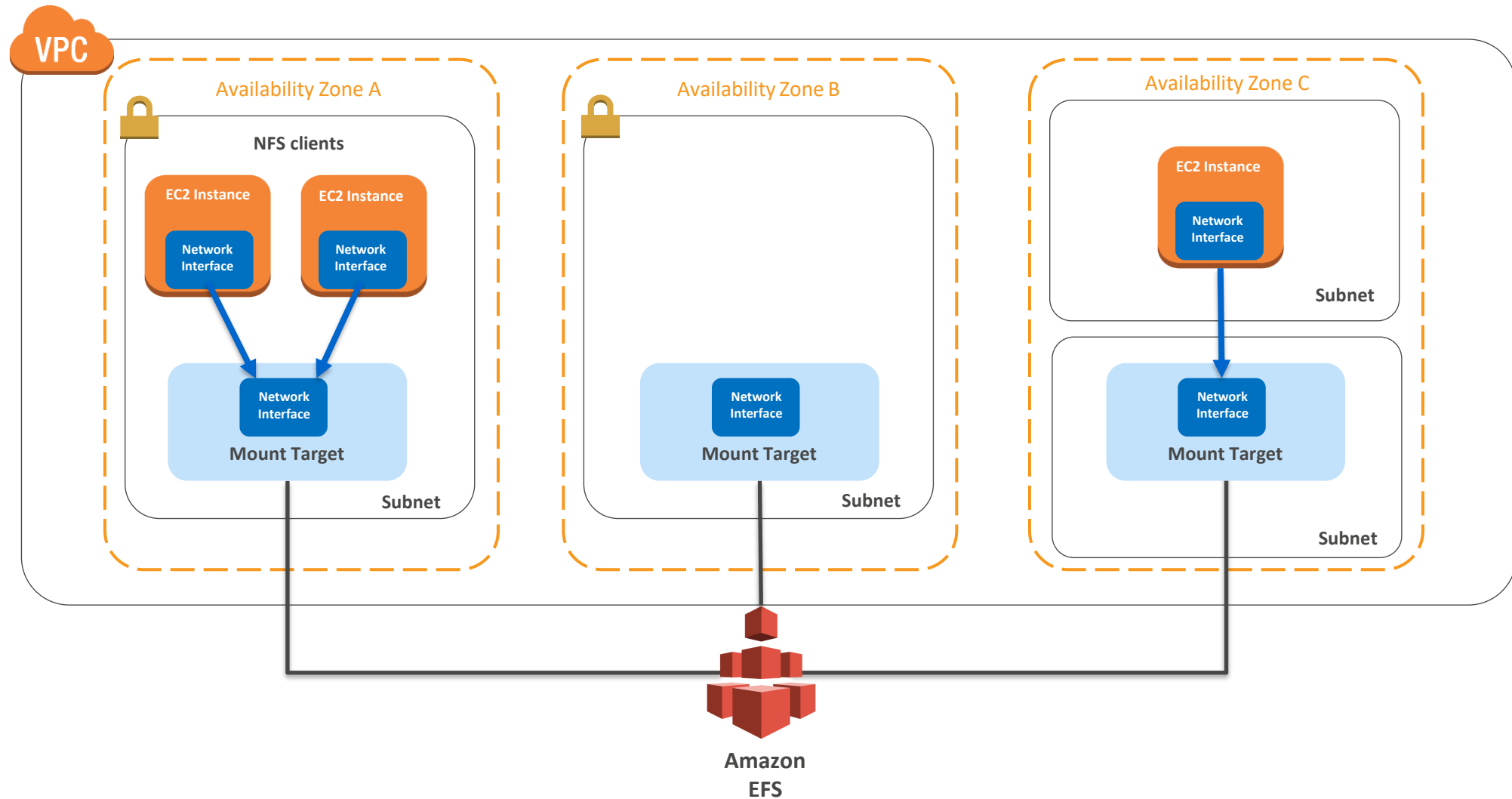
Amazon Elastic File System (Amazon EFS)

Amazon EFS Features

- File storage in the AWS cloud
- Perfect for big data and analytics, media processing workflows, content management, web serving and home directories
- Petabyte-scale, low latency file system
- Shared storage
- Elastic capacity
- Supports the Network File System versions 4.0 and 4.1 (NFSv4) protocol
- Compatible with all Linux-based AMIs for Amazon EC2



Amazon EFS Architecture



Amazon EFS Implementation

- ① Create your Amazon EC2 resources and launch your Amazon EC2 instance.
- ② Create your Amazon EFS file system.
- ③ Create your target mounts in appropriate subnets.
- ④ Connect your Amazon EC2 instances to target mounts.
- ⑤ Clean up resources and protect your AWS account.



File system

- Mount target
 - Subnet ID
 - Security groups
 - One or more per file system
 - Create in a VPC subnet
 - One per Availability Zone
 - Must be in the same VPC
- Tags
 - Key-value pairs



- Amazon EFS provides file storage over a network.
- Perfect for big data and analytics, media processing workflows, content management, web serving and home directories.
- Fully managed service that eliminates storage administration tasks.
- Accessible from the console, an API, or the CLI.
- Scales up or down as files are added or removed and you pay for what you use.





Amazon Glacier

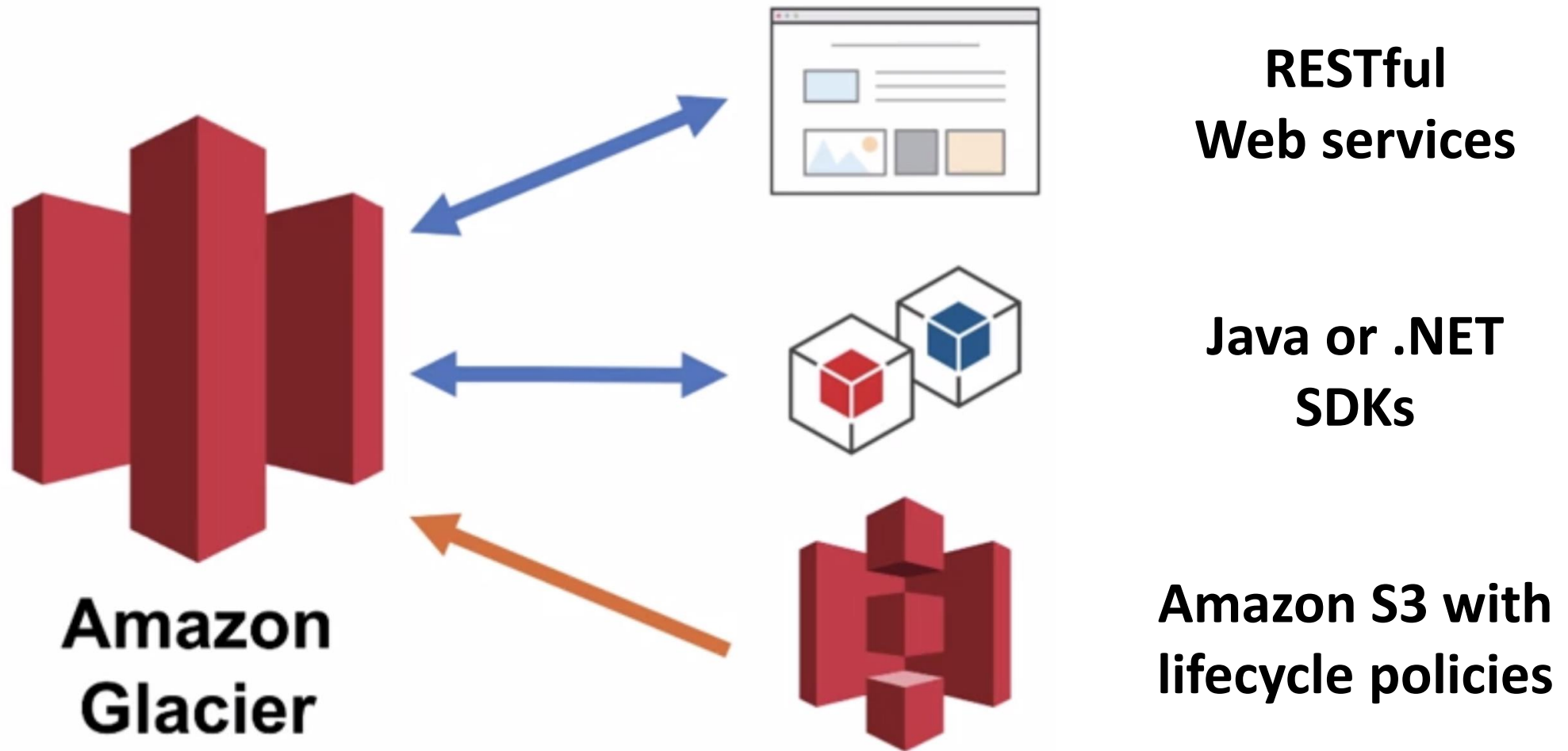
Amazon Glacier Review

Amazon Glacier is a **data archiving service** designed for **security**, **durability**, and an **extremely low cost**.

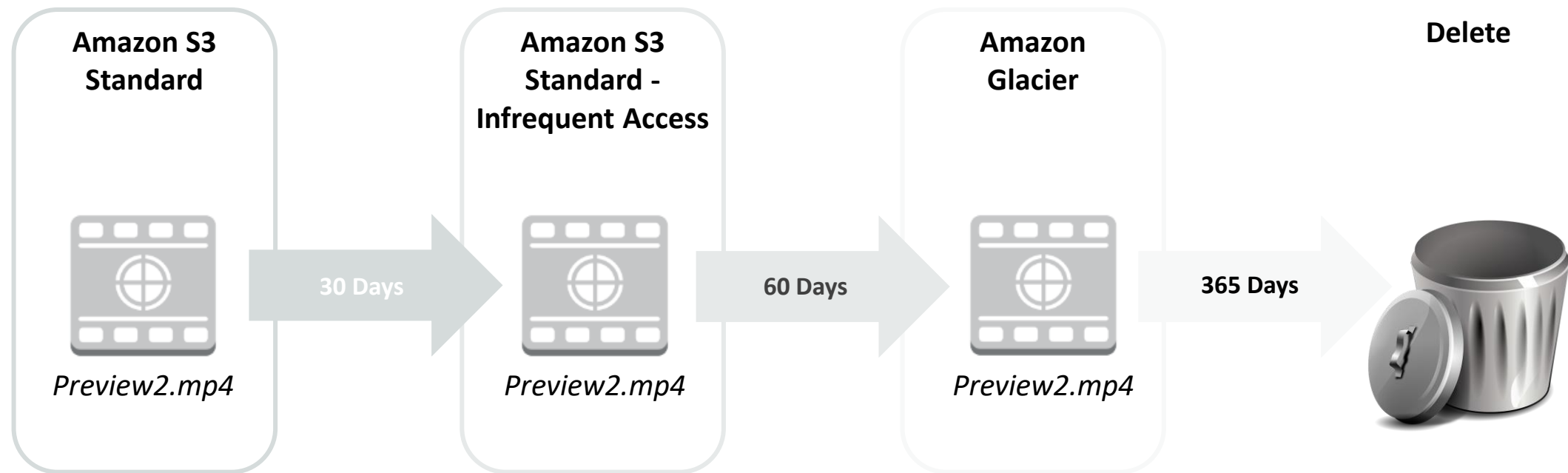
- Designed for durability of 99.999999999% of objects.
- Supports SSL/TLS encryption of data in transit and at rest.
- The Vault Lock feature enforces compliance via a lockable policy.
- Extremely low-cost design is ideal for long-term archiving.
 - Provides three options for access to archives (Expedited, Standard, and Bulk) from a few minutes to several hours.



Using Amazon Glacier



Amazon S3 lifecycle policies allow you to delete or move objects based on age.

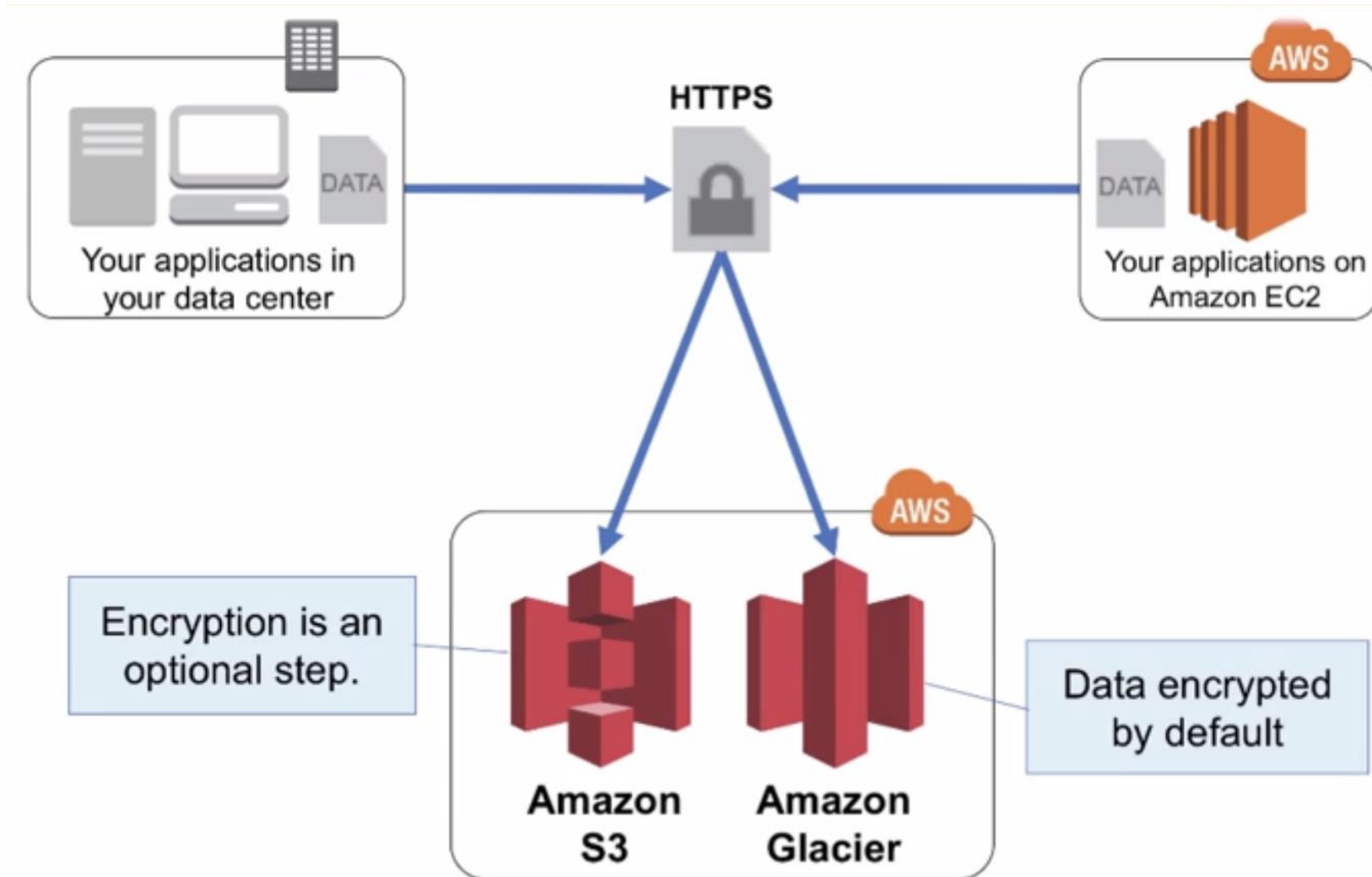


Storage Comparison

	Amazon S3	Amazon Glacier
Data Volume	No limit	No limit
Average Latency	ms	min/hrs
Item Size	5 TB max	40 TB max
Cost/GB Per Month	¢¢	¢
Billed Requests	PUT, COPY, POST, LIST, and GET	UPLOAD and retrieval
Retrieval Pricing	¢ Per request	¢¢ Per request and per GB



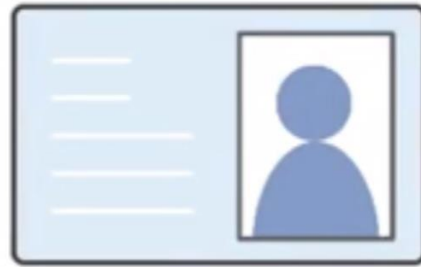
Server-Side Encryption



Security with Amazon Glacier



**Amazon
Glacier**



**Control access with
AWS IAM**



**Amazon Glacier encrypts
your data with AES-256**



**Amazon Glacier manages
your keys for you**



- Amazon Glacier is a data archiving service designed for security, durability, and an extremely low cost.
- Amazon Glacier pricing is region-based.
- Extremely low-cost design is ideal for long-term archiving.
- The service is designed for durability of 99.999999999% of objects.



Additional resources

- [Amazon S3 Developers Guide](#)
- [Amazon S3 FAQs](#)
- [Amazon S3 Common Use Scenarios](#)
- [AWS Storage Services Whitepaper](#)
- [Amazon S3 Storage Classes Comparison](#)
- [Amazon S3 Block Public Access](#)