Adding a Storage Layer with Amazon S3

AWS Academy Cloud Architecting

Introduction

Adding a Storage Layer with Amazon S3

Module objectives

This module prepares you to do the following:

- Define Amazon Simple Storage Service (Amazon S3) and how it works.
- Recognize the problems that Amazon S3 can solve.
- Describe how to move data to and from Amazon S3.
- Manage the storage of content efficiently by using Amazon S3.
- Recommend the appropriate use of Amazon S3 based on requirements.
- Configure a static website on Amazon S3.
- Use the AWS Well-Architected Framework principles when designing a storage layer with Amazon S3.

Module overview

Presentation sections

- Defining Amazon S3
- Using Amazon S3
- Moving data to and from Amazon S3
- Storing content with Amazon S3
- Designing with Amazon S3
- Applying the AWS Well-Architected Framework principles to storage

Demos

- Amazon S3 Transfer Acceleration
- Managing Lifecycles in Amazon S3
- Amazon S3 Versioning

Activity

- Designing with Amazon S3

Knowledge checks

- 10-question knowledge check
- Sample exam question

Hands-on lab in this module

Challenge (Café) lab

- Creating a Static Website for the Café

As a cloud architect working with Amazon S3:

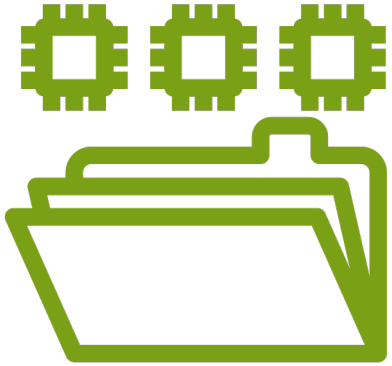
- I need to consider access patterns and use cases of the business so that I can choose Amazon S3 configuration options that optimize cost while supporting performance and compliance requirements.
- I need to apply security best practices so that the storage layer is protected against unwanted access and accidental data loss.

Defining Amazon S3

Adding a Storage Layer with Amazon S3

Types of storage

Block storage	File storage	Object storage
Data is stored on a device in fixed-sized blocks.	Data is stored in a hierarchical structure.	Data is stored as objects based on attributes and metadata.



Amazon S3

Amazon S3

- Amazon S3 stores massive (unlimited) amounts of unstructured data.
- Amazon S3 stores data files as objects in a bucket that you define.
- Five TB is the maximum file size of a single object.
- Objects have a globally unique URL (universal namespace).
- Objects have a key, version ID, value, metadata, and sub-resources.

Amazon S3 components

Bucket

A bucket is a container for objects.
Each bucket has a unique name and a Region-specific endpoint in this format.

https://s3-<*aws-region*>.amazonaws.com/<*bucket-name*>/ <*object-key* >

Object

An object is a file and the metadata that describes that file.
Metadata is a set of name-value pairs that are assigned when the object is uploaded.
An object includes a key name (object-key) that uniquely identifies the object.

Use prefixes to imply a folder structure in an S3 bucket

The following objects are in a bucket named **graphics-bucket**.

```
photos/2022/catpiano.jpg
photos/2022/catonphone.jpg
photos/2022/ninepuppies.png
photos/2021/lakefront.png
photos/2021/coveredbridge.png
photos/2021/openairmarket.jpg
video-source/9984.mp4
video-source/9918.mp4
video-source/18446.mp4
```

A **GET** query with the prefix **photos/2022** returns the following objects:

```
graphics-bucket/photos/2022/catpiano.jpg
graphics-bucket/photos/2022/catonphone.jpg
graphics-bucket/photos/2022/ninepuppies.png
```

Amazon S3 benefits

Benefit	Description
 Durability	• Helps ensure that data is not lost • Provides S3 Standard storage with 11 nines (or 99.999999999 percent) of durability
 Availability	• Provides access to data when needed • Includes unlimited capacity to store data • Provides S3 Standard storage with 4 nines (or 99.99 percent) of availability
 High performance	• Achieves thousands of transactions each second when uploading and retrieving storage • Automatically scales to high request rates

Key takeaways: Defining Amazon S3

- Storage comes in three basic types: block storage, file storage, and object storage.
- Amazon S3 is an object storage service.
- Amazon S3 stores massive amounts of unstructured data.
- A bucket is a container for objects that are stored in Amazon S3.
- An object is the fundamental entity that is stored in Amazon S3.
- The key benefits of Amazon S3 include durability, availability, and high performance.

Using Amazon S3

Adding a Storage Layer with Amazon S3

How customers use Amazon S3

Spikes in demand

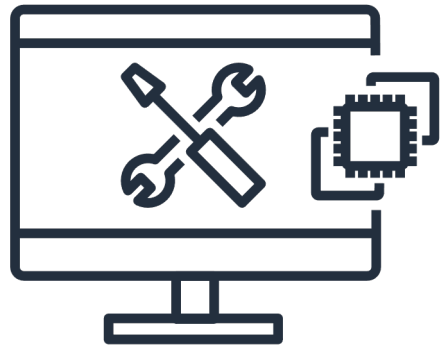
Host web content that needs bandwidth to address extreme spikes in demand.

Static site

Host a static site that consists of HTML files, images, and videos.

Financial analysis

Store data that other services can use for analysis.

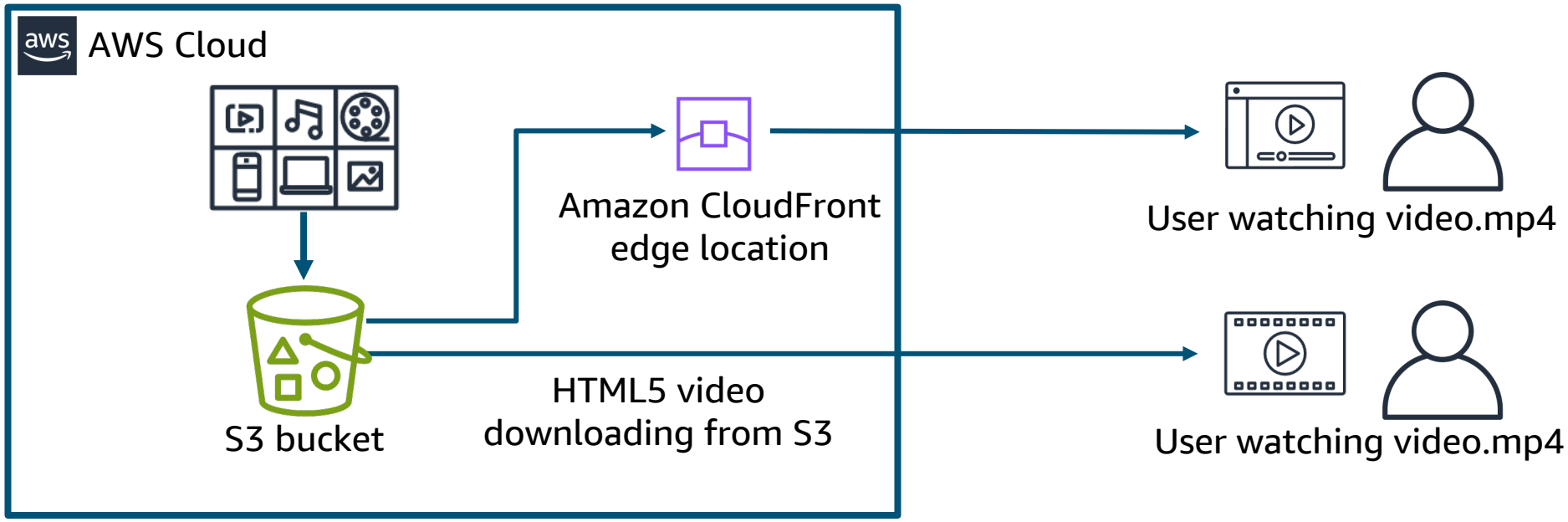


Disaster recovery

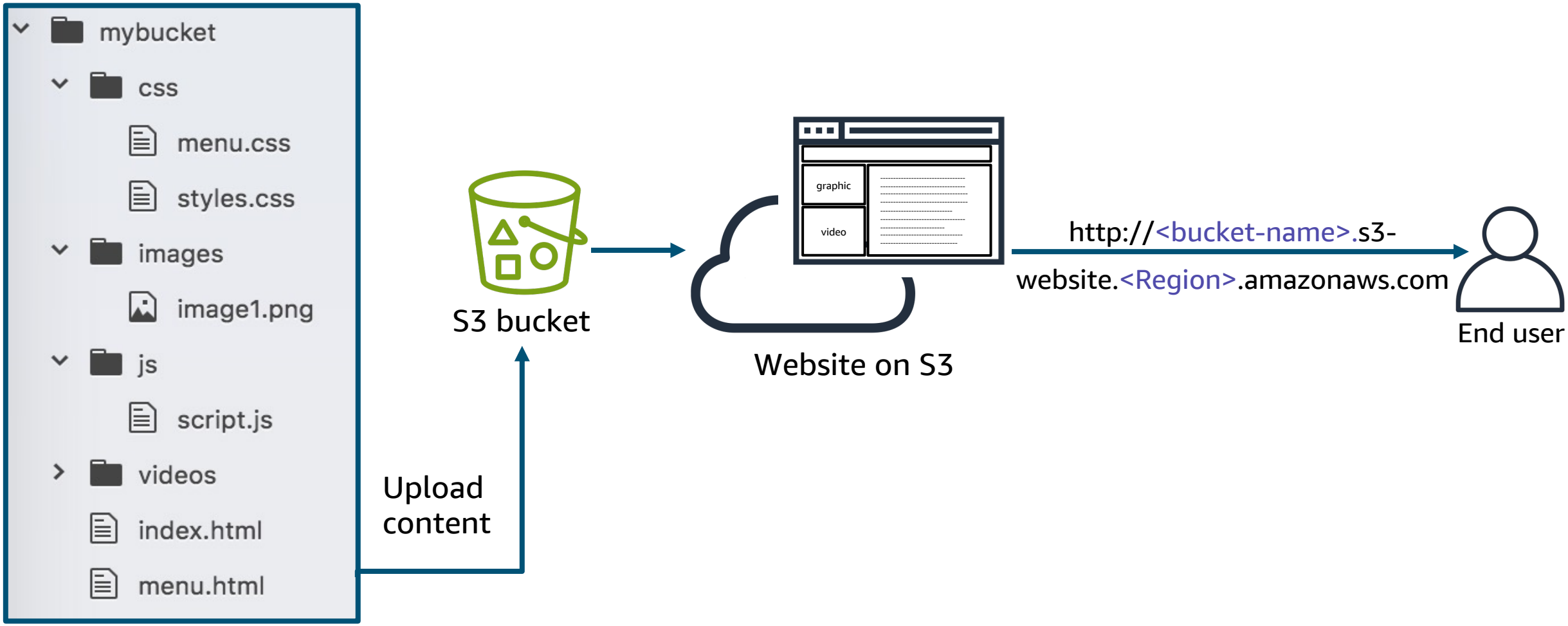
Support disaster recovery or data backup solutions.

Use case: Media hosting

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

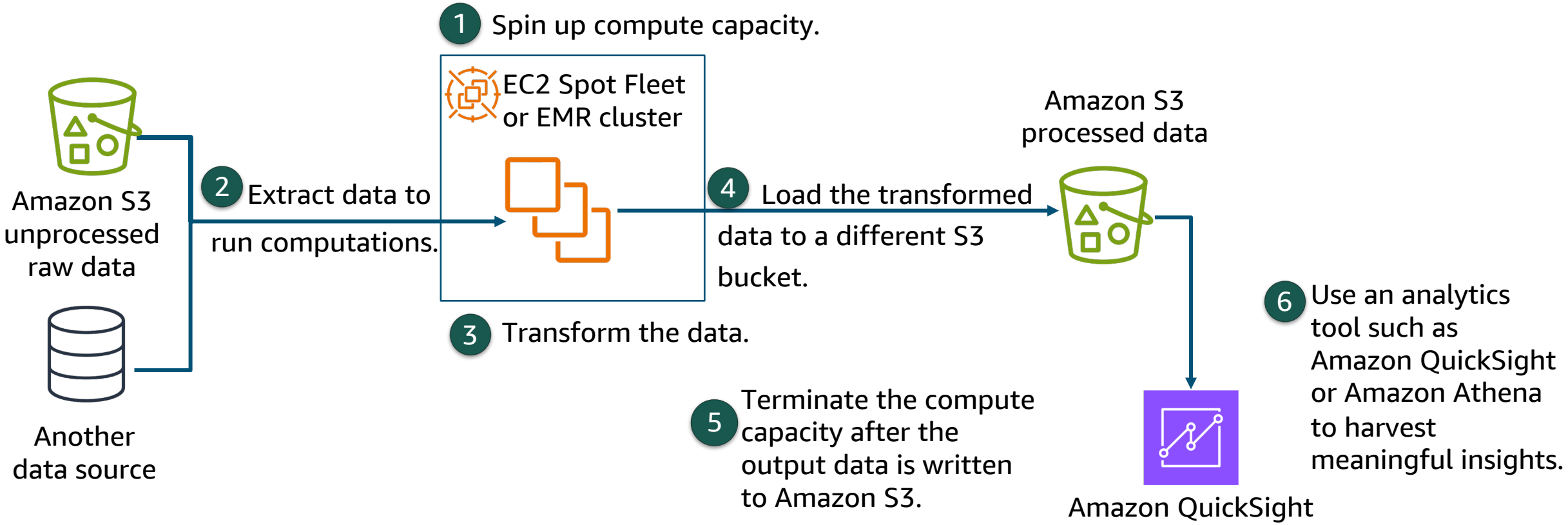


Use case: Host static websites

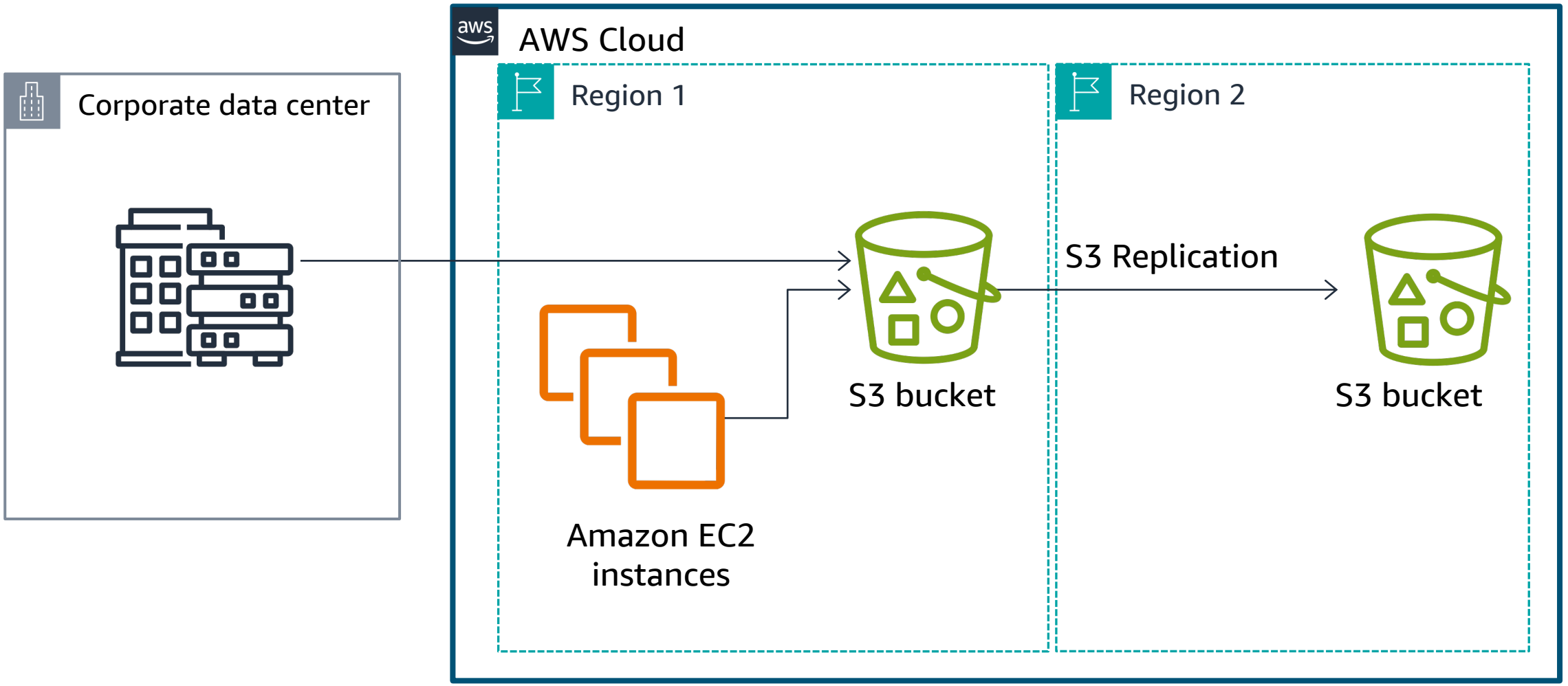


Use case: Data store for computation and analytics

Example data integration and preparation pattern



Use case: Back up and archive critical data



Key takeaways: Using Amazon S3

Amazon S3 is often used to do the following:

- Store and distribute videos, photos, music files, and other media.
- Support static content, including HTML files, images, videos.
- Store data for computation and large-scale analytics.
- Provide a data backup solution.

Moving data to and from Amazon S3

Adding a Storage Layer with Amazon S3

Storing objects in Amazon S3

- There is no limit to the number of objects in a bucket.
- Uploading an object requires write permission to the bucket.
- Objects are encrypted by default.
  - During upload, objects are automatically encrypted by using server-side encryption.
  - During download, objects are decrypted.

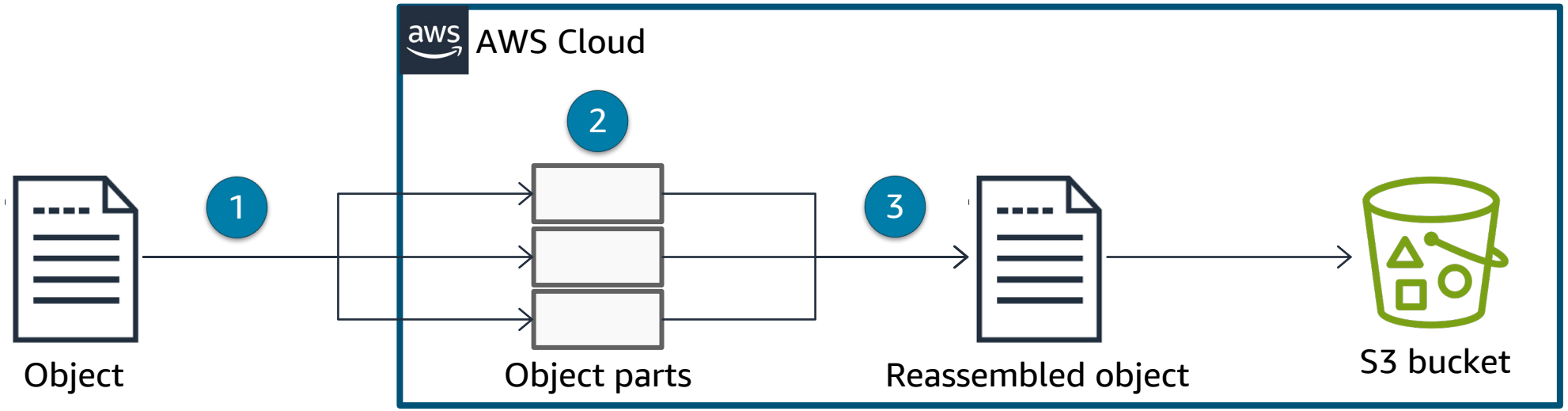
Options for uploading objects to Amazon S3

Uploading Amazon S3	Description
AWS Management Console	Use a wizard-based approach to move data into or out of Amazon S3, including the option to drag and drop files. (maximum size is 160 GB).
AWS Command Line Interface (AWS CLI)	Upload or download from a terminal command prompt or in a call from a script.
AWS SDKs	Use AWS SDKs to upload objects programmatically.
Amazon S3 REST API	Send a PUT request to upload data in a single operation.

Amazon S3 feature: Multipart upload

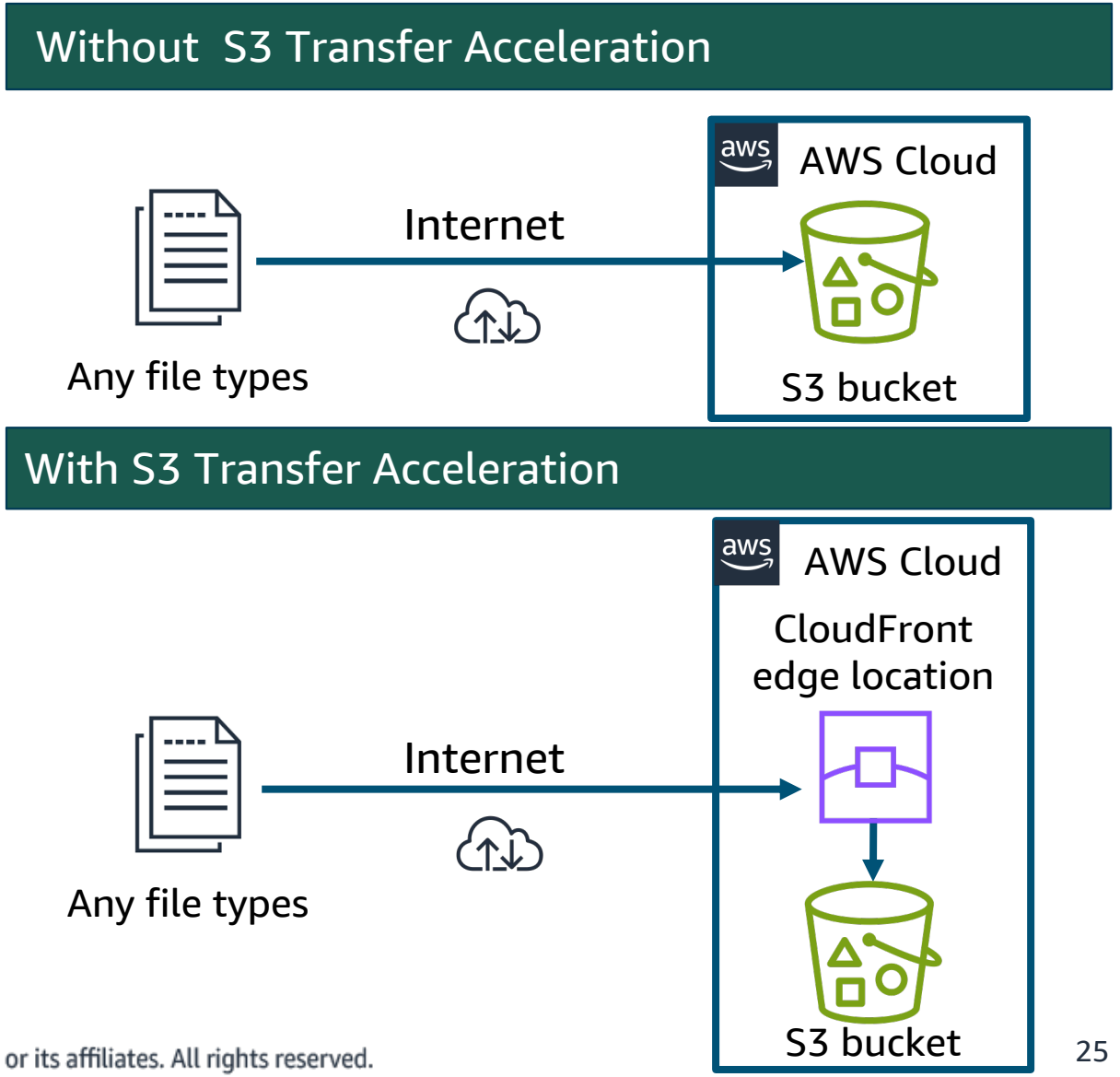
Multipart uploads have the following advantages:

- Improve throughput.
- Recover quickly from any network issues.
- Pause and resume object uploads.
- Begin an upload before you know the final object size.

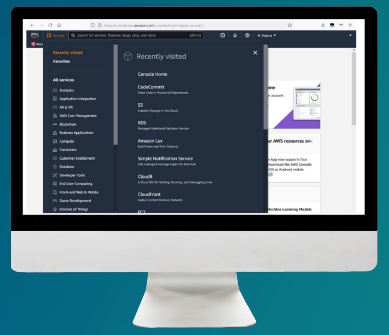


S3 Transfer Acceleration

- Provides fast and secure transfers of files over long distances.
- Optimizes transfer speeds from across the world into S3 buckets.
- Uses globally distributed edge locations in CloudFront.
- Improves speed by 50–500 percent on average for cross-country transfer of larger objects.

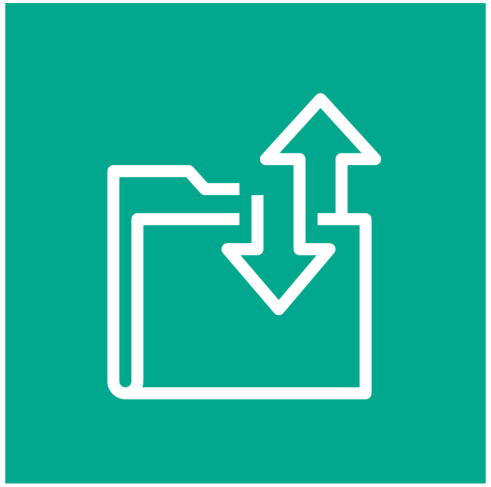


Demo: Amazon S3 Transfer Acceleration



- This demo uses Amazon S3 Transfer Acceleration.
- In this demonstration, you will see how to enable transfer acceleration for an S3 bucket and access accelerated data transfers.

AWS Transfer Family



AWS Transfer Family

- Is a fully managed AWS service
- Is used to transfer files into and out of Amazon S3 storage or Amazon Elastic File System (Amazon EFS) file systems over the following protocols:
  - Secure Shell (SSH) File Transfer Protocol (SFTP) version 3
  - File Transfer Protocol Secure (FTPS)
  - File Transfer Protocol (FTP)
  - Applicability Statement 2 (AS2)

Transfer Family benefits

- Transfer Family is a managed service that scales in real time to meet your needs.
- You don't need to modify your applications or run any file transfer protocol infrastructure.
- With Transfer Family, you use native AWS services for processing, analytics, reporting, auditing, and archival functions with your data in durable Amazon S3 storage.
- Transfer Family is a managed elastic file system (with Amazon EFS) for use with AWS Cloud services and on-premises resources.
- Transfer Family is a managed, serverless file transfer workflow service that you can use to set up, run, automate, and monitor file uploads.
- You pay for only the use of the service, and there are no upfront costs.

Use cases for Transfer Family

Amazon S3

- Data lakes in AWS for uploads from third parties
- Subscription-based data distribution with customers
- Internal transfers within your organization

Amazon EFS

- Data distribution
- Supply chain
- Content management
- Web-serving applications

Key takeaways: Moving data to and from Amazon S3

- Upload objects to Amazon S3 by using the AWS Management Console, AWS CLI, AWS SDKs, or Amazon S3 REST API.
- Use multipart upload to upload a single object as a set of parts when the object size reaches 100 MB or greater.
- Use S3 Transfer Acceleration on an S3 bucket to provide fast and secure transfers of files over long distances between your client and an S3 bucket.
- Use Transfer Family to provide a secure transfer of files into and out of AWS storage services.

Storing content with Amazon S3

Adding a Storage Layer with Amazon S3

Object storage classes

General purpose	Intelligent tiering	Infrequent access	Archive
S3 Standard	S3 Intelligent-Tiering	S3 Standard-IA	S3 Glacier Instant Retrieval
		S3 One Zone-IA	S3 Glacier Flexible Retrieval
			S3 Glacier Deep Archive
			S3 on Outposts

S3 storage classes breakdown

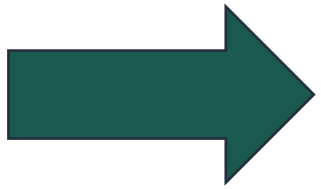
	S3 Standard	S3 Intelligent-Tiering	S3 Standard-IA	S3 One Zone-IA	S3 Glacier Instant Retrieval	S3 Glacier Flexible Retrieval	S3 Glacier Deep Archive
Availability Zones	≥3	≥3	≥3	1	≥3	≥3	≥3
Minimum capacity charge for each object	N/A	N/A	128 KB	128 KB	128 KB	N/A	N/A
Minimum storage duration charge	N/A	N/A	30 days	30 days	90 days	90 days	180 days
Retrieval charge	N/A	N/A	Per GB retrieved	Per GB retrieved	Per GB retrieved	Per GB retrieved	Per GB retrieved

Configuring an Amazon S3 Lifecycle

Amazon S3 lifecycle configurations are a set of rules that define actions that Amazon S3 applies to a group of objects.

- Transition actions transition to another storage class.
- Expiration actions define when objects expire.

Set an S3 lifecycle policy.



Data will automatically transfer to a different storage class without any changes to your application.

Amazon S3 lifecycle examples

1

S3 Standard

Required for 30 days

Delete

2

Amazon S3 Glacier

Archived for 10 years for – regulatory compliance

3

S3 Standard

Frequently accessed for 60 days

S3 Standard-IA

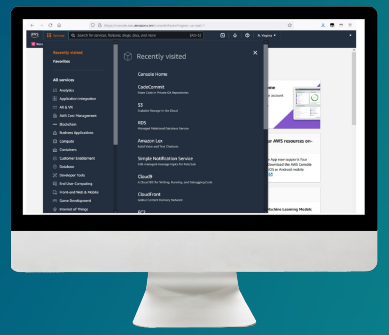
Infrequently accessed for 1 year

Amazon S3 Glacier

Archived for 7 years

Delete

Demo: Managing Lifecycles in Amazon S3



- This demo uses Amazon S3.
- In this demonstration, you will see how to create a lifecycle rule for Amazon S3 and apply transition actions to objects.

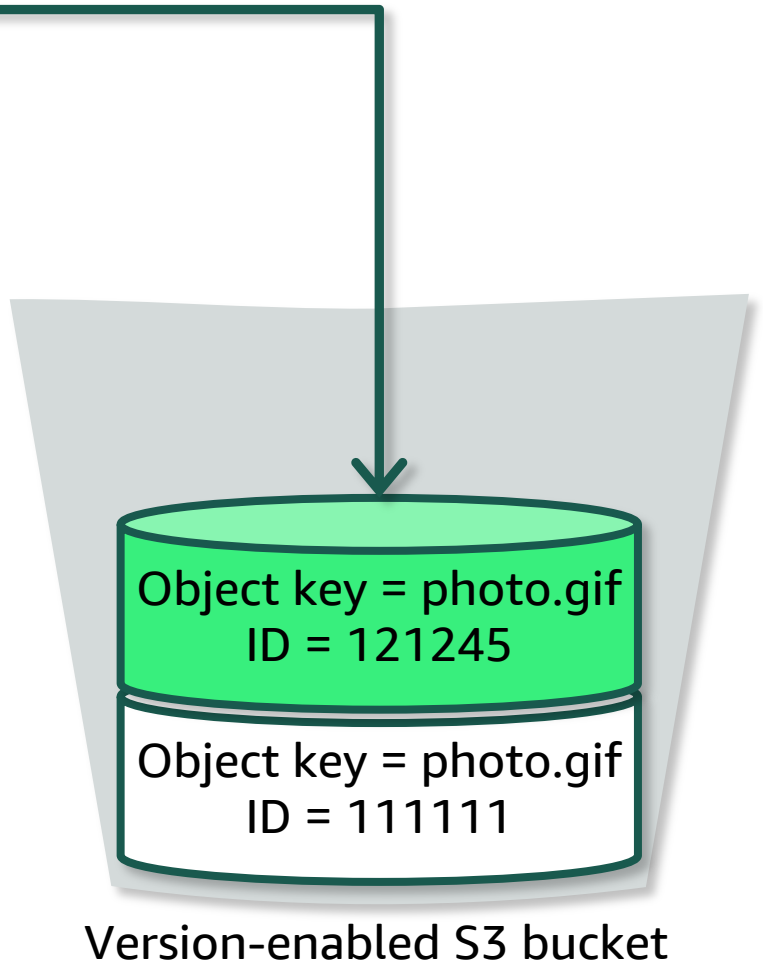
Amazon S3 versioning

Amazon S3 versioning protects objects from accidental overwrites and deletes.

Action	Versioning Enabled	Versioning Disabled or Versioning Suspended
Upload an object with the same key	Creates a new object with a different version ID, and both are retrievable by the version ID.	Overwrites the original object, and the previous object is no longer retrievable.
Delete	Adds a delete marker, but the object is still retrievable by the version ID.	Deletes the object, and it is no longer retrievable.

Adding an object in a versioning-enabled bucket

PUT
Object key = photo.gif

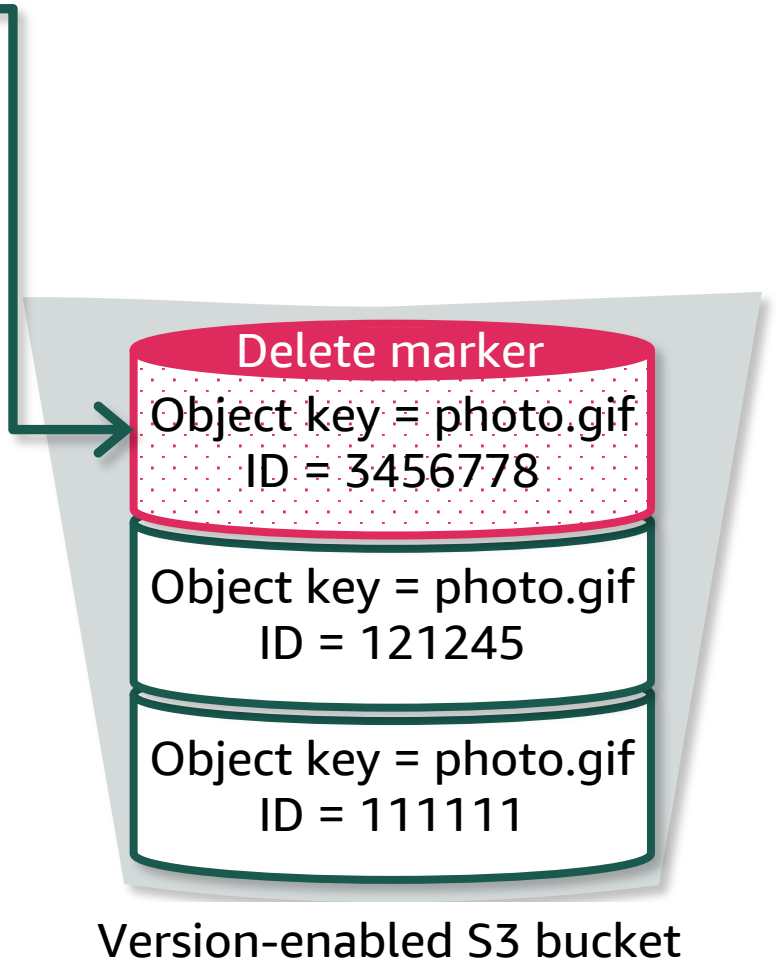


Version-enabled S3 bucket

- Amazon S3 generates a new version ID and adds this newer version of the object to the bucket.
- The original version remains in the bucket.

Deleting an object in a version-enabled bucket

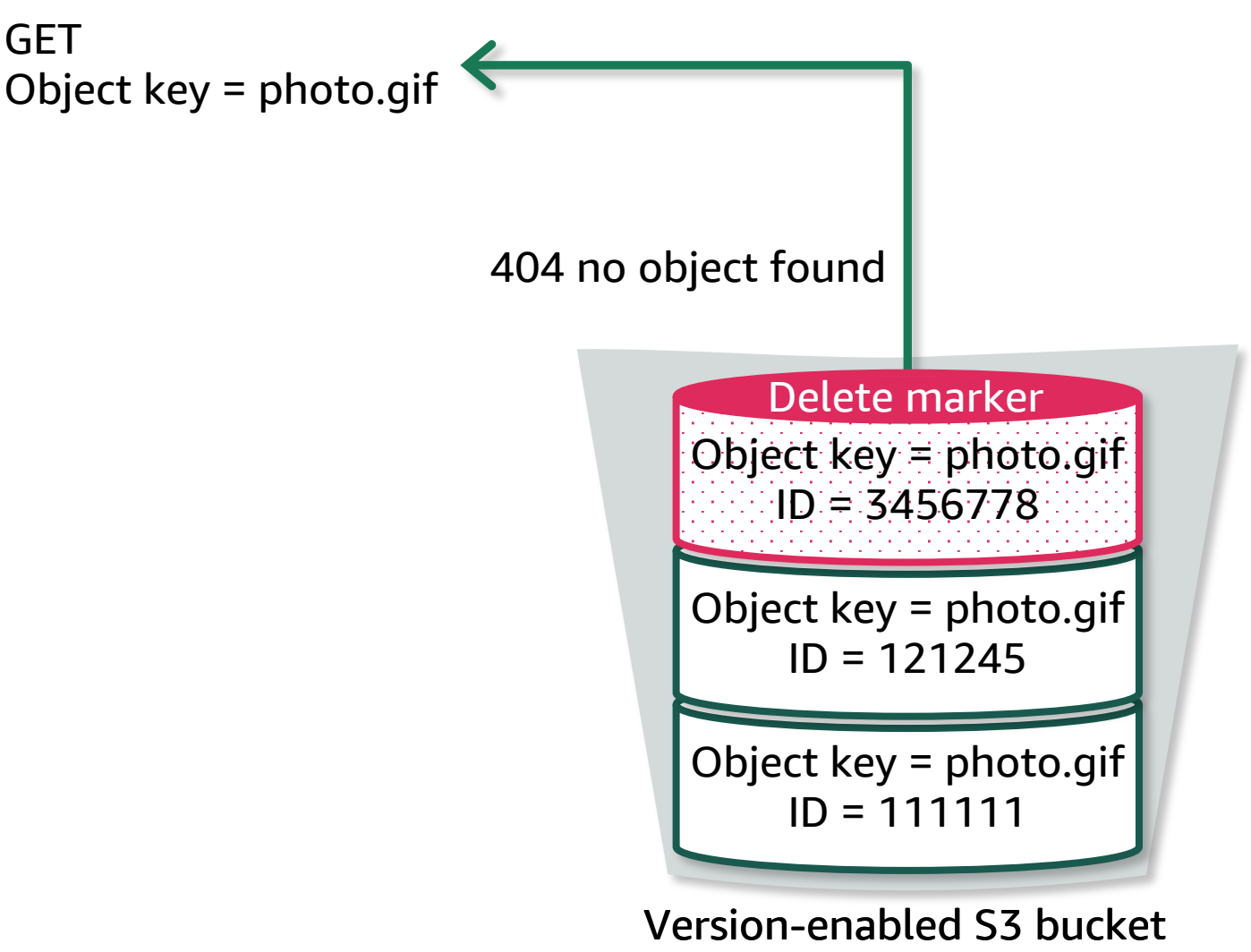
DELETE
Object key = photo.gif



Version-enabled S3 bucket

When a request is made to delete an object in a version-enabled bucket, all versions remain in the bucket, but Amazon S3 inserts a delete marker.

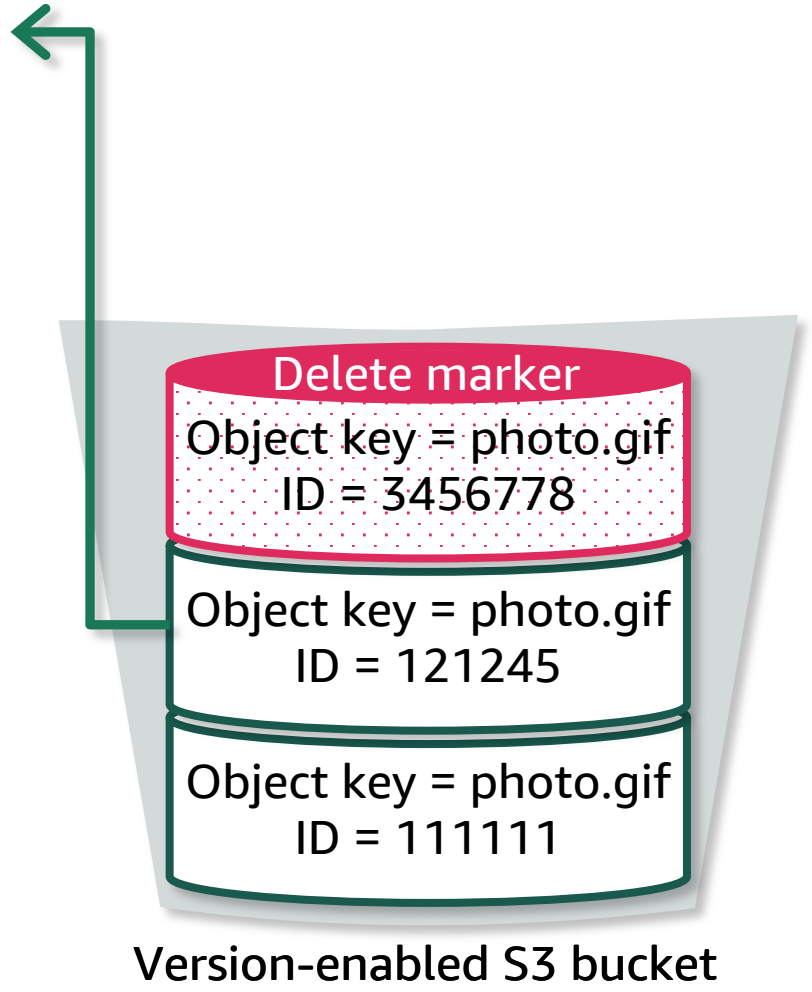
Retrieving the most recently stored version



- Requests for an object key return the most recent version.
- If the most recent version is a delete marker, the request is not successful.

Retrieving an object with its specific ID

GET
Object key = photo.gif
with version ID = 121245

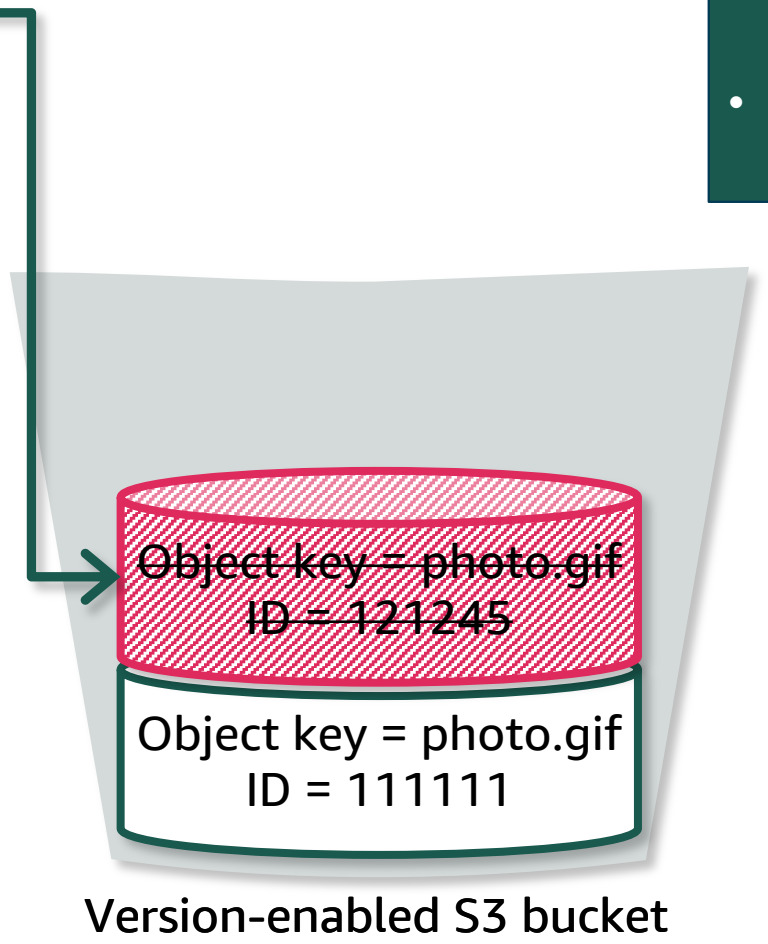


Version-enabled S3 bucket

Requests for an object with its version ID will successfully return that version of the object.

Permanently delete an object

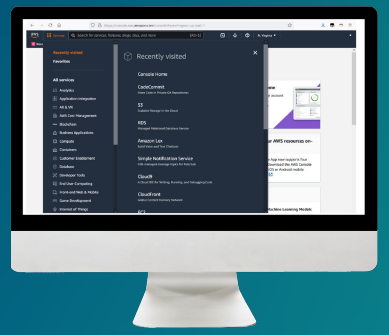
DELETE
Object key = photo.gif
With version id = 121245



Version-enabled S3 bucket

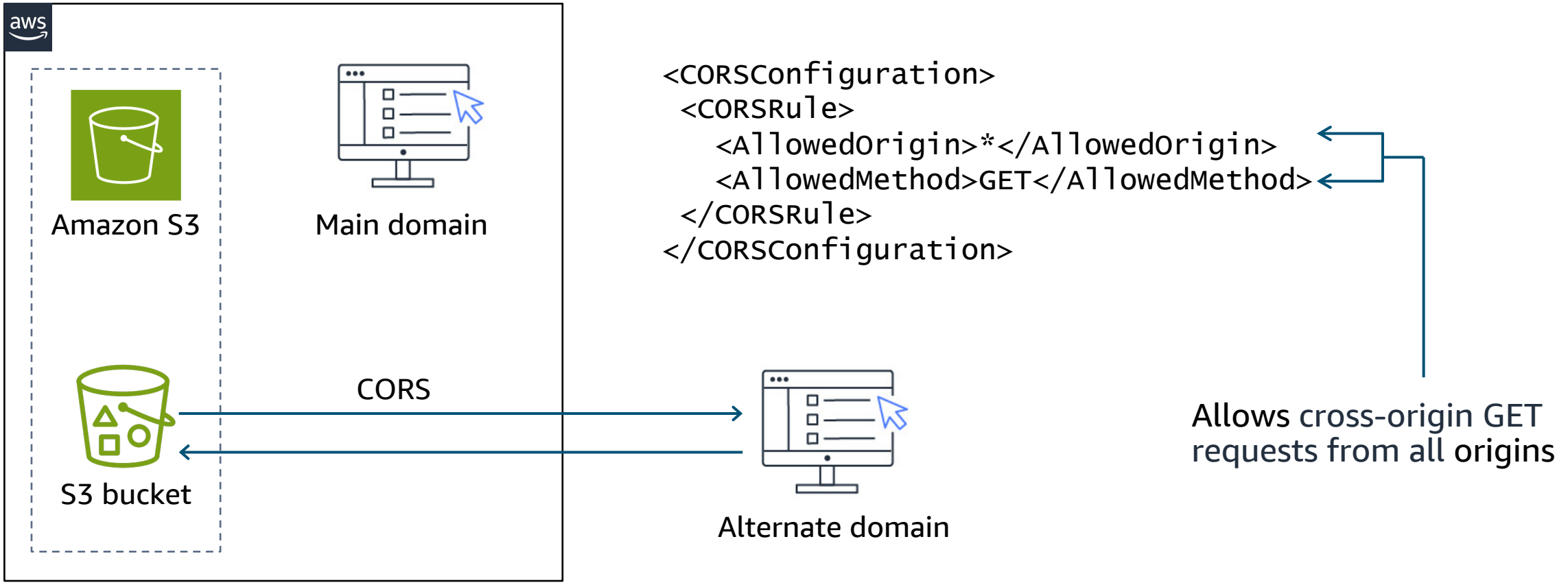
- Owners of the bucket can permanently delete an object by using delete with the version ID.
- In this case, no delete marker is added, and the specified version is not recoverable.

Demo: Amazon S3 Versioning



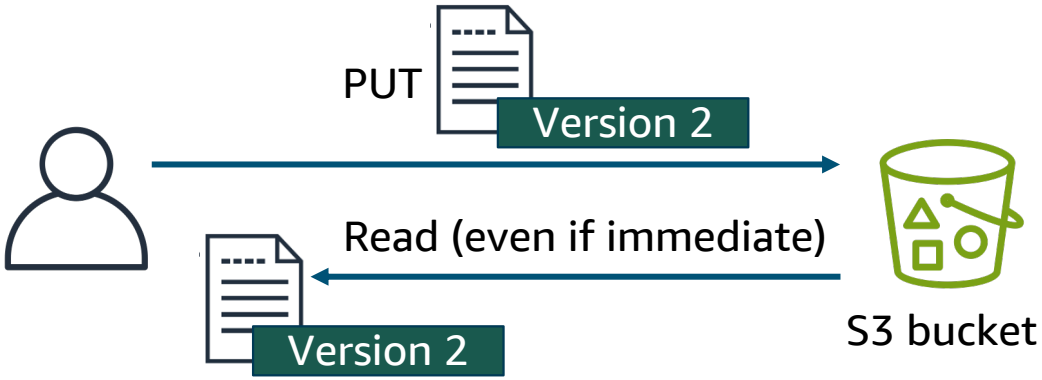
- This demo uses Amazon S3.
- In this demonstration, you will see how to enable Amazon S3 versioning on an S3 bucket.

Support for cross-origin resource sharing (CORS)



Amazon S3 data consistency model

- Is consistent for all new and existing objects in all Regions
- Provides read-after-write consistency for all GET, LIST, PUT, and DELETE operations on objects in S3 buckets
- Offers an advantage for big data workloads
- Simplifies the migration of on-premises analytics workloads



Key takeaways: Storing content with Amazon S3

- S3 Standard storage is appropriate for cloud applications, dynamic websites, content distribution, mobile and gaming applications, and big data analytics.
- Set an S3 lifecycle policy and your data will automatically transfer to a different storage class without any changes to your application.
- Recover from both unintended user actions and application failures by using versioning.
- CORS defines a way for client web applications that are loaded in one domain to interact with resources in a different domain.
- The Amazon S3 data consistency model simplifies the migration of on-premises analytics workloads by removing the need to make changes to support applications.

Designing with Amazon S3

Adding a Storage Layer with Amazon S3

Amazon S3 default security configurations

- S3 buckets and objects created are private and protected by default.
- S3 buckets have encryption configured by default.
- Server-side encryption with Amazon S3 managed keys (SSE-S3) is the default encryption.

When use cases must share Amazon S3 data, do the following:

- Manage and control the data access.
- Follow the principle of least privilege.

Encrypting objects in Amazon S3

Encryption encodes data with a secret key, which makes it unreadable without a key.

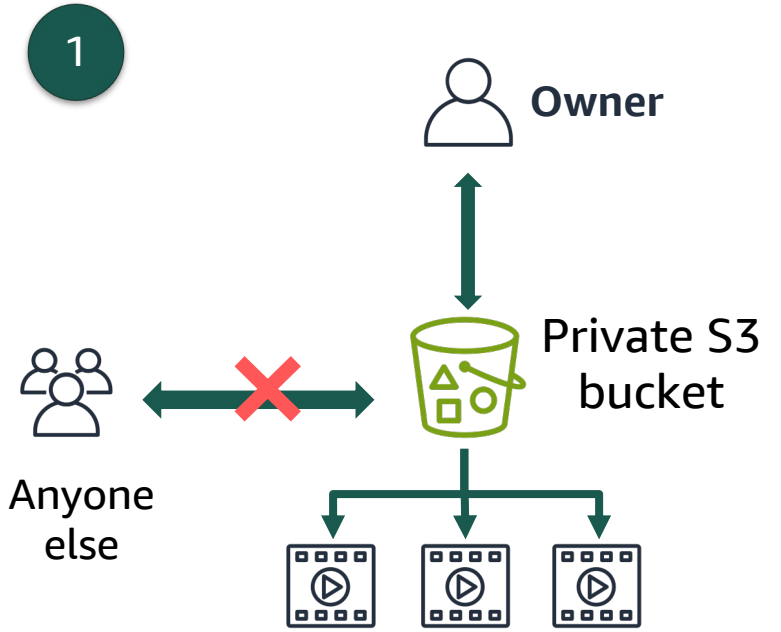
- Server-side encryption
  - Amazon S3 encrypts objects before it saves the objects to disk and decrypts the objects when you download them.
  - Enable this feature by selecting the default encryption option on the bucket.
- Client-side encryption
  - Encrypt data on the client side and upload the encrypted data to Amazon S3.
  - In this case, you manage the encryption process.

Amazon S3 tools for protecting buckets and objects

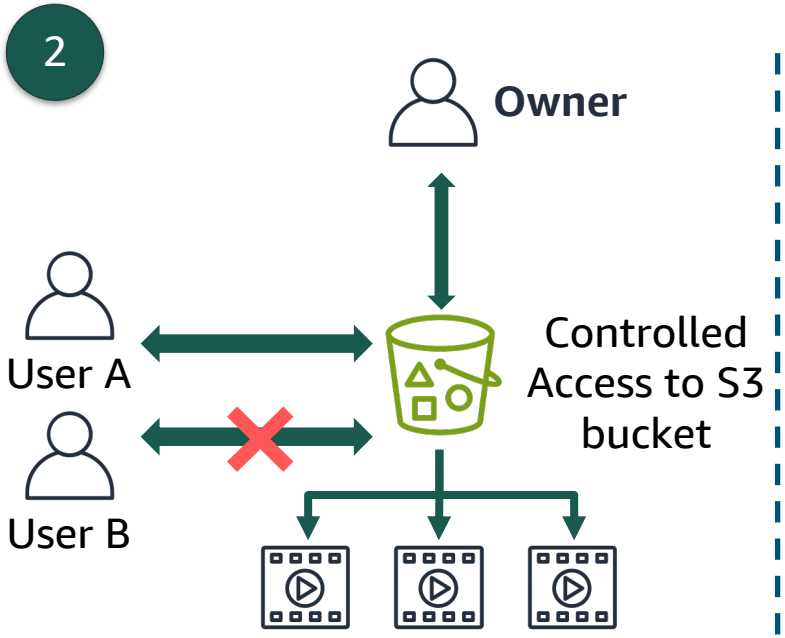
Tool	Description
Block Public Access feature	Makes buckets inaccessible to the public
AWS Identity and Access Management (IAM) policies	Authenticates users by using IAM
Bucket policies	Defines access based on specific written rules
Access control lists (ACLs)	Sets rules for access to buckets and objects (bucket policies are the preferred method for controlling bucket access)
Amazon S3 access points	Configures access with names and permissions specific to each application
Preassigned URLs	Grants time-limited access to others with temporary URLs
AWS Trusted Advisor	Provides a bucket permission check

Three general approaches to configuring access

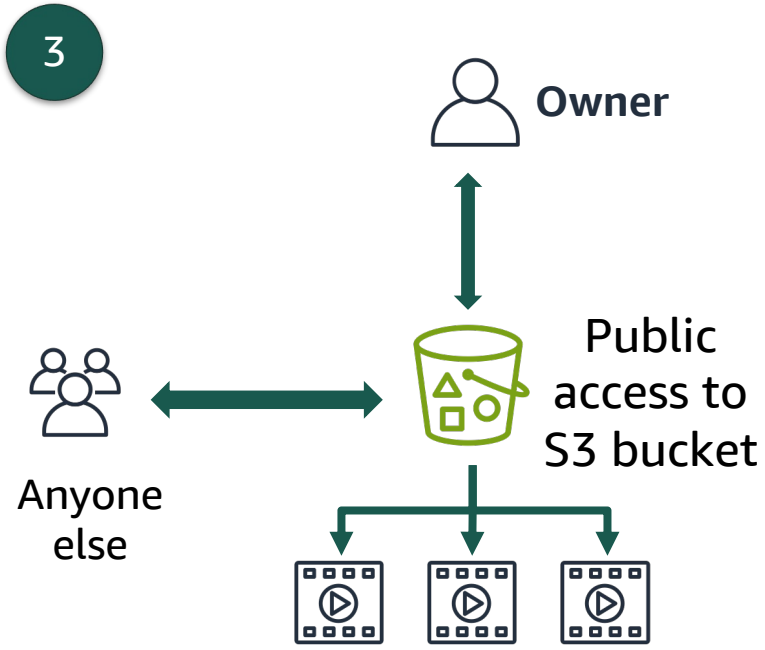
Default



Access policy applied



Public Access



Not recommended

Considerations when choosing a Region

Considerations	Details
Data privacy laws and regulatory compliance	• Are there relevant Region data privacy laws? • Can customer data be stored outside the country? • Can you meet your governance obligation?
Proximity of users to data	• Small differences in latency can impact customer experience. • Choose the Region closest to your users.
Availability service and feature	• Not all AWS services are available in all Regions. • Services expand to new Regions regularly. • Use some services cross-Region but at increased latency.
Cost-effectiveness	• Costs vary by Region. • Some services such as Amazon S3 have costs for transferring data out. • Consider the cost-effectiveness of replicating the entire environment in another Region.

Amazon S3 Inventory

- Use Amazon S3 Inventory to help manage your storage.
- Use it to audit and report on the replication and encryption status of your objects for business, compliance, and regulatory needs.
- Speed up business workflows and big data jobs by using Amazon S3 Inventory.
- Provide a scheduled alternative to the Amazon S3 synchronous List API operations.

Amazon S3 costs

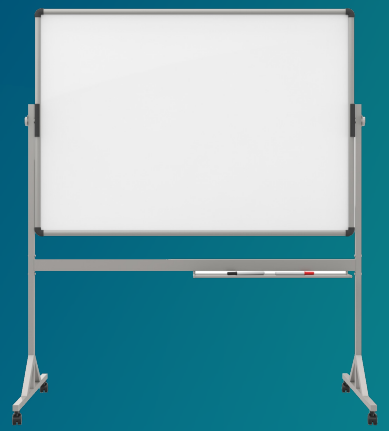
Pay for only what you use:

- Gigabytes of objects stored (per month) with different pricing for each Region and each storage class
- PUT, COPY, POST, LIST or lifecycle transition to move data into any Amazon S3 storage class

No charge for data transferred:

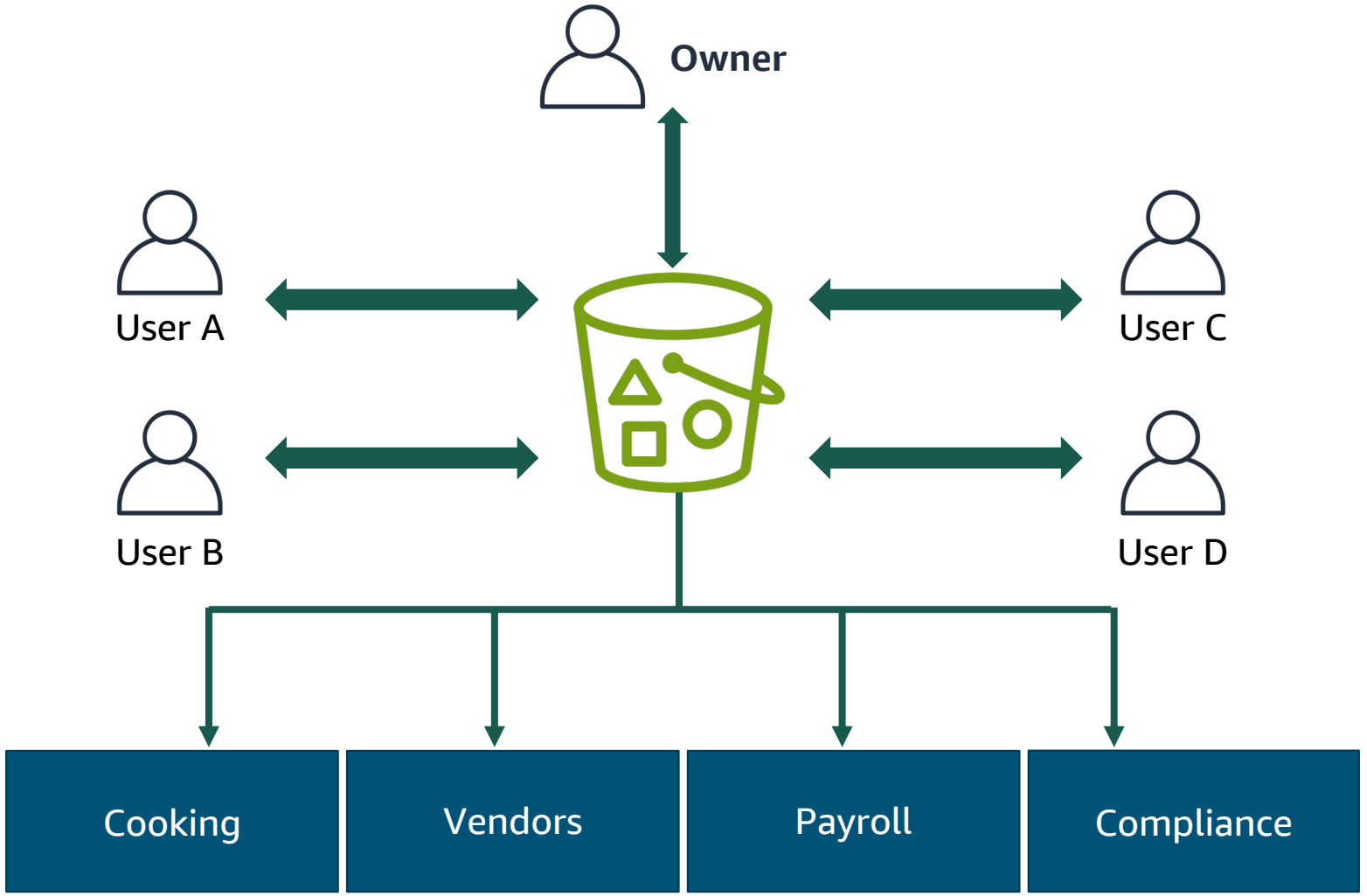
- Out to the internet for the first 100 GB per month
- In from the internet
- Between S3 buckets or to any service in the same AWS Region
- From an S3 bucket to any AWS service or services within the same AWS Region as the S3 bucket
- Out to CloudFront

Activity: Designing with Amazon S3



- The café owner has successfully created a static website but now wants to make documents more accessible to employees. They would like to house content (cooking instructions, vendor information, payroll, and compliance training) that only employees can access.
- How could you set this up to meet the owner’s needs?

Activity solution: Created controlled access for each user



Key takeaways: Designing with Amazon S3

- S3 buckets are private and can be accessed only by users who are explicitly granted access.
- SSE-S3 is the default encryption configuration for every bucket in Amazon S3.
- Considerations when choosing a Region include data privacy laws and regulatory compliance, proximity of users to data, availability service and features, and cost-effectiveness.
- Amazon costs are based on your objects' size, the amount of time you stored the objects during the month, and the storage class.

Café lab: Creating a Static Website for the Café (Café Website Lab)

The evolving café architecture: version 1

Architecture Version	Business Reason for the Update	Technical Requirements and Architecture Update
V1	Create a static website for a small business.	Host the website on Amazon S3.
V2	Add online ordering.	Deploy the web application and database on Amazon EC2.
V3	Reduce the effort to maintain the database and secure its data.	Separate web and database layers. Migrate the database to Amazon Relational Database Service (Amazon RDS) on a private subnet.
V4	Enhance the security of the web application.	Use Amazon Virtual Private Cloud (Amazon VPC) features to configure and secure public and private subnets.
V5	Create separate access mechanisms based on role.	Add IAM groups and attach resource polices to application resources. Add IAM users to groups based on role.
V6	Help ensure that the website can handle an expected increase in traffic.	Add a load balancer, implement auto scaling on the EC2 instances, and distribute compute and database instances across two Availability Zones.

Café Website Lab tasks

- In this lab, you will do the following:
  - Create an S3 bucket to host your static website and upload content to your S3 bucket.
  - Create a bucket policy to grant public read access.
  - Enable versioning on the S3 bucket.
  - Set lifecycle policies.
  - Enable cross-Region replication.

Debrief: Café Website Lab

- How did you protect the café website from accidental overwrite and deletion?
- Which strategy did you use to help save on costs as the size of the website grows through the use of versioning?

Applying the AWS Well-Architected Framework principles to storage

Adding a Storage Layer with Amazon S3

Well-Architected Framework best practices for storage

Securi ty

Reliabili ty

Performan ce Efficien cy

Cost Optimization

Best practice approach: Data protection – protecting data at rest

Best practices

Enforce encryption at rest.

Enforce access control.

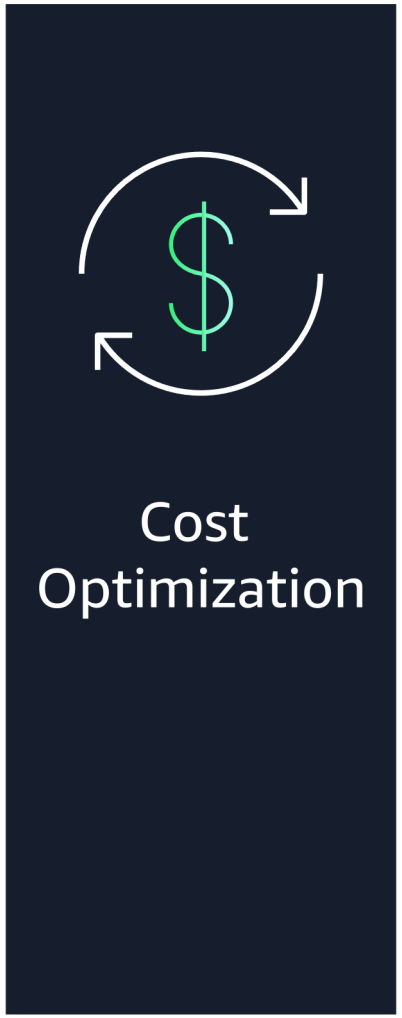
Best practice approach: Architecture selection

Best practices

Learn about and understand available cloud services and features

Factor cost into architectural decisions

Best practice approach: Cost-effective resources – evaluate cost when selecting services



Best practice

Perform cost analysis for different usage over time

Best practice approach: Failure management - Use fault isolation to protect your workload

Best practice

Select the appropriate locations for your multi-location deployment

Key takeaways: Applying the AWS Well-Architected Framework principles to storage

- Protecting data is a security best practice that Amazon S3 supports through these default configurations: encrypting objects, making objects private, blocking public access.
- You can protect data in Amazon S3 by limiting access through IAM policies and enabling versioning.
- Selecting an architecture is a performance efficiency best practice that Amazon S3 supports through its ability to store massive amounts of unstructured data.
- Amazon S3 includes performance-improving options such as S3 Transfer Acceleration and multipart upload.
- Selecting cost-effective resources is a cost-optimization best practice that Amazon S3 supports through features such as lifecycle policies, intelligent tiering, and Amazon S3 Inventory.
- Failure management is a reliability best practice that Amazon S3 has been designed for through its durability and availability features.
- You can use Amazon S3 for backing up data to improve failure management of your applications and data.

Module wrap-up

Adding a Storage Layer with Amazon S3

Module summary

This module prepared you to do the following:

- Define Amazon S3 and how it works.
- Recognize the problems that Amazon S3 can solve.
- Describe how to move data to and from Amazon S3.
- Manage the storage of content efficiently by using Amazon S3.
- Recommend the appropriate use of Amazon S3 based on requirements.
- Configure a static website on Amazon S3.
- Use the Well-Architected Framework principles when designing a storage layer with Amazon S3.

Considerations for the café

- Discuss how the café lab in this module addressed the cloud architect's key concerns presented at the start of this module.

Module knowledge check

- The knowledge check is delivered online within your course.
- The knowledge check includes 10 questions based on material presented on the slides and in the slide notes.
- You can retake the knowledge check as many times as you like.

Sample exam question

Company salespeople upload their sales figures daily to Amazon S3, but their solutions architect is concerned that users might accidentally delete or overwrite important documents.

Which action will protect against unintended user actions?

Identify the key words and phrases before continuing.

The following are the key words and phrases:

- Amazon S3
- Accidentally delete or overwrite
- Protect against unintended

Sample exam question: Response choices

Company salespeople upload their sales figures daily to **Amazon S3**, but their solutions architect is concerned that users might **accidentally delete or overwrite** important documents.

Which action will **protect against unintended** user actions?

Choice	Response
A	Store data in two S3 buckets in different AWS Regions.
B	Enable versioning on the S3 bucket where files are stored.
C	Move uploaded data to an Amazon S3 Infrequent Access storage class at the end of each week.
D	Use Amazon S3 Inventory to audit the status of objects in the S3 bucket where files are stored.

Sample exam question: Answer

The answer is B.

Choice	Response
A	
B	Enable versioning on the S3 bucket where files are stored.
C	
D	

Thank you

Corrections, feedback, or other questions?
Contact us at https://support.aws.amazon.com/#/contacts/aws-academy.