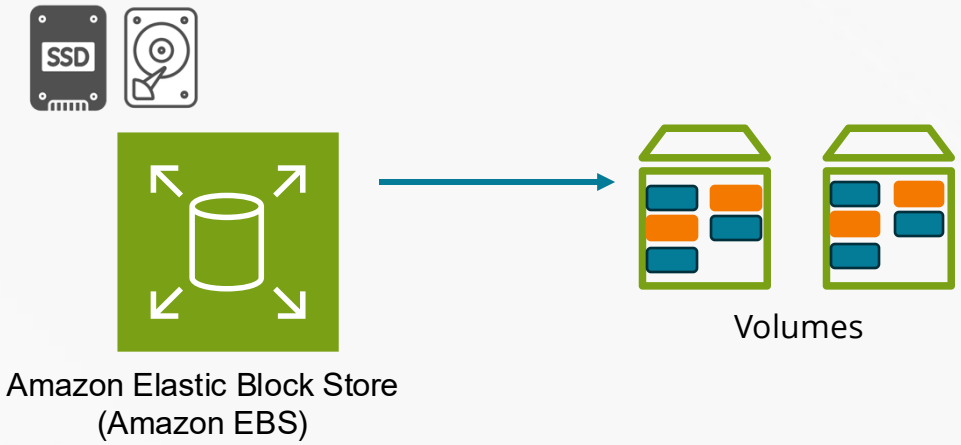
Introduction to Storage Options

Block, File and Object Storage

IAAS ACADEMY

Block Storage



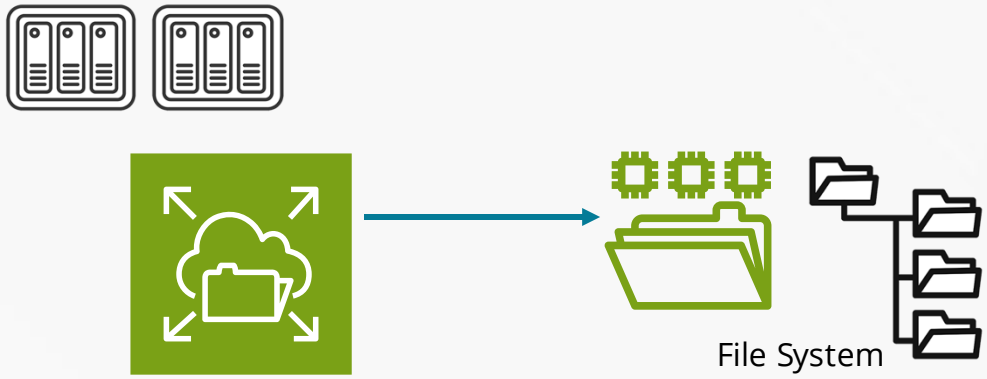
Block storage divides data into fixed-sized blocks and stores them as separate pieces

- Data divided into fixed-size blocks
- Unique identifier
- Retrieve data by its block ID

Low Latency: high-performance applications requiring low latency.
High Customizability: fine-tuned storage optimizations (e.g., databases, transactional applications).
Compatibility: high-performance applications like virtual machines, databases, and enterprise applications.

Complex Management: Block storage can be harder to manage, especially when scaling.
No Metadata: Blocks don't store metadata, so context for each block isn't preserved.

File Storage



Amazon Elastic File System
(Amazon EFS)

Amazon FSx for
Windows File Server

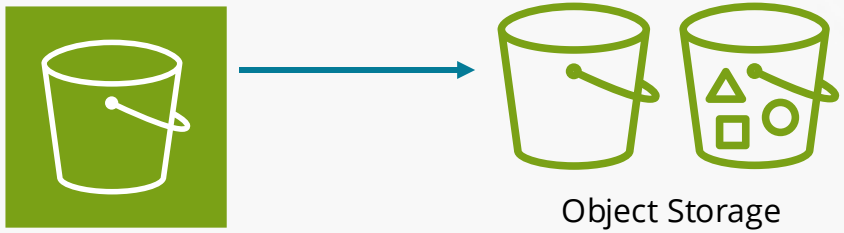
File storage organizes data in a hierarchical structure of files and folders. Data is stored and retrieved as complete files using a path-based system.

- Hierarchical file structure
- Can define tags and permissions on files
- Enables file sharing and collaboration

Familiar Structure: Simple, easy-to-use folder-based structure
Supports Metadata: Allows tagging and permissions on files, making data easier to manage and control.
Shared Access: Works well in environments where multiple users need to access files (e.g., NAS)

Scalability Limits: Less scalable than object storage due to limitations in the hierarchical structure
Lower Performance for Large Data Sets: Can be less efficient than block storage for high-performance needs.

Object Storage



Amazon Simple Storage Service (Amazon S3)

File Flat Namespace

Object storage manages data as individual objects with unique identifiers and metadata, storing them in a flat namespace in a repository. Each object includes data, metadata, and a unique ID.

- Data is stored with metadata and a unique ID
- Stored across a distributed network offering durability
- Erasure Coding
- Replication and horizontal scaling support

Scalability: Highly scalable, ideal for massive amounts of unstructured data
Metadata-Rich: Extensive metadata support, useful for organizing, indexing, and searching data
Durability and Redundancy: Often includes built-in replication and redundancy, improving data resilience.

Latency: Higher latency than block storage, making it less suited for high-performance applications.
Limited Compatibility: Not ideal for applications that require frequent updates to data, such as databases.

Introduction to Amazon S3

Buckets and Objects

IAAS ACADEMY

What is Amazon S3

- Global Object Storage Solution – *Accessible globally but hosted regionally.*
- Data is stored as **objects** (with metadata and a globally unique identifier).
- Public service with multi-connect access.
- Use cases include data lakes, websites, mobile apps, backups, archives, and big data analytics.
- Designed to offer **99.999999999 (11 9s)** of durability.
- From a single CSV file to a high-resolution video.

Object Storage – Not File or Block Storage

Amazon Simple Storage Service (Amazon S3)

Buckets are private by default

IAM Policies | Bucket Policies | ACLs

Access Methods

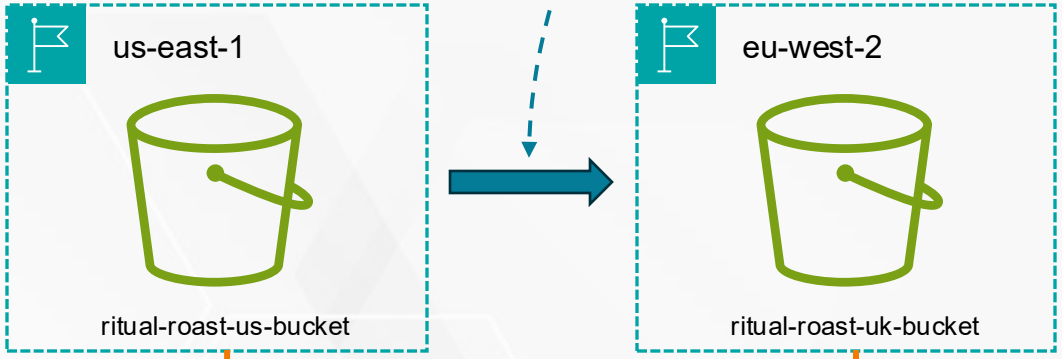
What is Amazon S3 – Buckets and Objects

AWS never replicates your data to other regions unless you configure it to do so.

Increases Durability & Availability

Compliance & Data Sovereignty

Ellie – CEO
Ritual Roast Limited



- Globally unique names across AWS accounts
- 3 to 63 characters in length
- Lowercase letters, numbers, dots (.) and hyphens (-)
- Cannot start with an IP address
- 100 soft limit and 1000 hard limit per account

Objects
- Size from 0 bytes to 5TB
- Key: blueberry-muffin.jpg
- Value:

Flat File Structure

- blueberry-muffic-recipe.doc
- chocolate-muffic-recipe.doc
- lemon-drizzle-muffic-recipe.doc
- /archive/walnut-cake-v2-recipe.doc
- /latest/walnut-cake-v3-recipe.doc

Delimiter (/)

Prefix

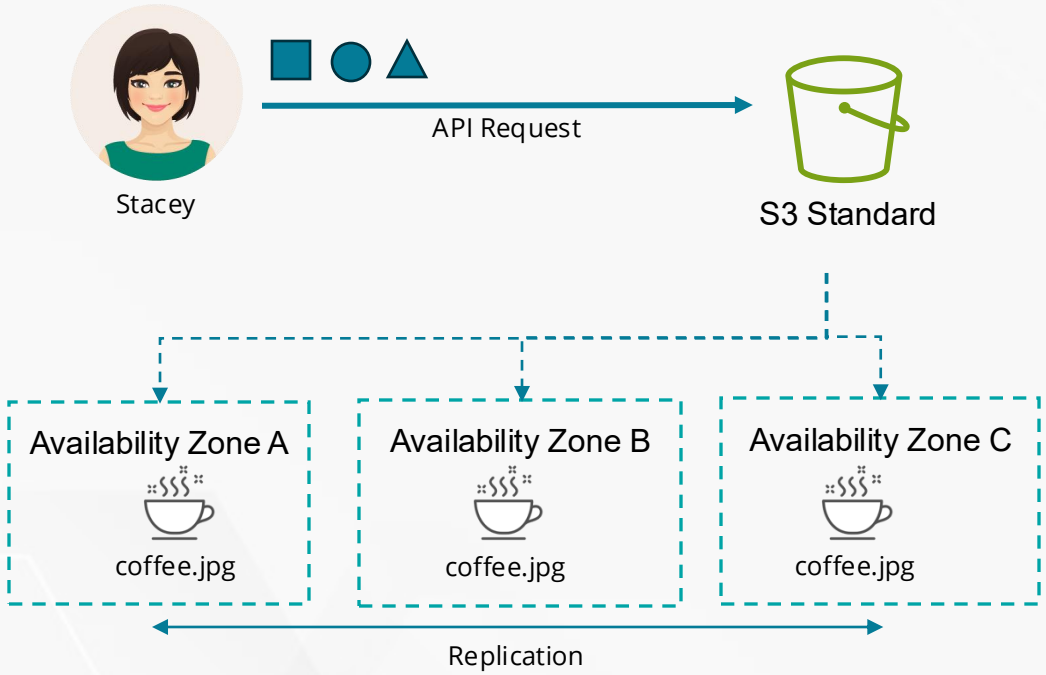
The prefix and delimiter parameters can be used to limit the results of a list operation

Amazon S3 Storage Classes

Choosing the best storage class for your use case

ACADEMY

Storage Class – S3 Standard



Default Storage Class

Instant Access, milliseconds first byte latency

Idea for frequently accessed data that is not replaceable

Buckets can be made publicly available

Using permissions and/or static website hosting

Durability

- Protection against data loss
- Prevent data corruption
- 99.999999999% (11 9s) durability

Content-MD5 Checksums and Cyclic Redundancy Checks (CRC) for data integrity

Tom

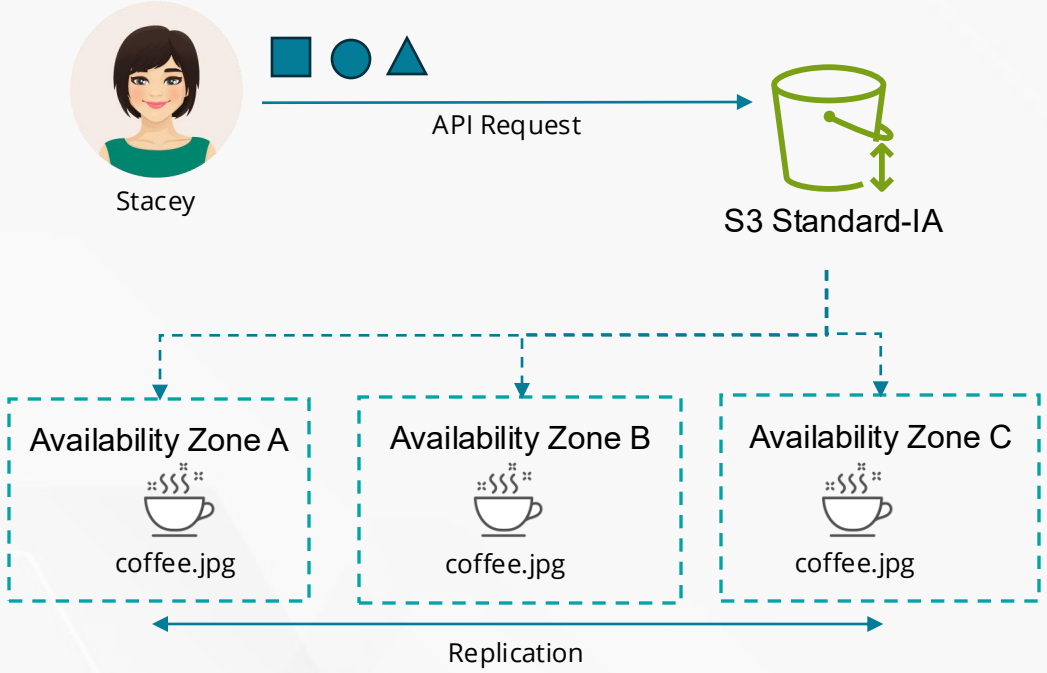
if you store 10,000,000 objects in Amazon S3, then on average, you can expect to experience the loss of a single object every 10,00 years

Lisa - CFO

Costs based on:

- Storage fee GB/month for data stored
- Transfer OUT fee
- Request fee per 1,000 requests

Storage Class – S3 Standard-IA (Infrequent Access)



Cheaper than S3 Standard

Instant Access: milliseconds first byte latency

Idea for infrequently accessed data that is not replaceable

Durability

- Protection against data loss
- Prevent data corruption
- 99.999999999% (11 9s) durability
- 99.9% availability

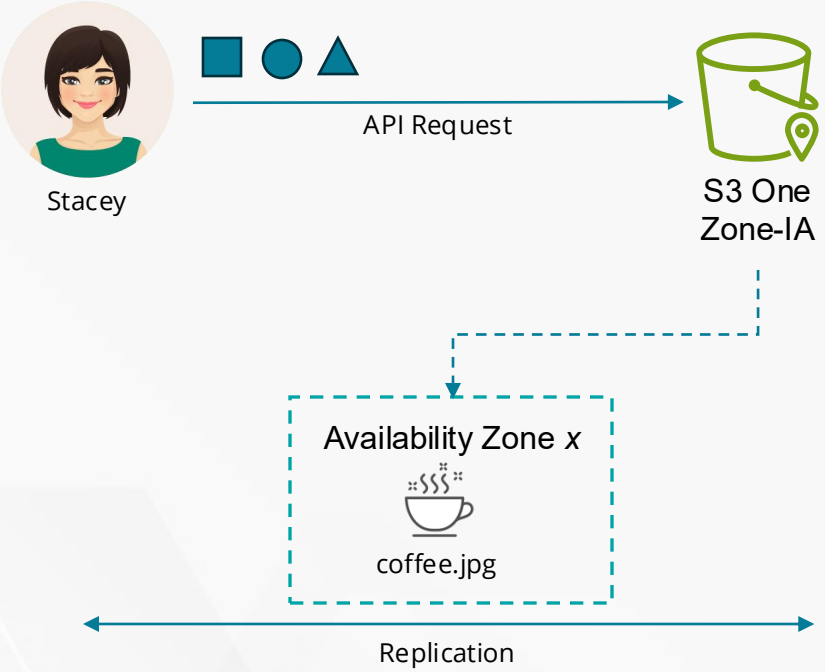
Content-MD5 Checksums and Cyclic Redundancy Checks (CRC) for data integrity

Lisa - CFO

Costs based on:

- Storage fee GB/month for data stored
- Transfer OUT fee
- Request fee per 1,000 requests
- **Minimum storage charge of 128KB**
- **Minimum duration charge of 30 days**
- **Per GB retrieval cost**

Storage Class – S3 One-Zone IA



Cheaper than S3 Standard and Standard IA

Instant Access: milliseconds first byte latency

Ideal for infrequently accessed data that **is** replaceable, e.g., secondary copies or replicas

Durability

- Protection against data loss
- Prevent data corruption
- 99.999999999% (11 9s) durability
- 99.5% availability

Content-MD5 Checksums and Cyclic Redundancy Checks (CRC) for data integrity

Data is stored in only one Availability Zone

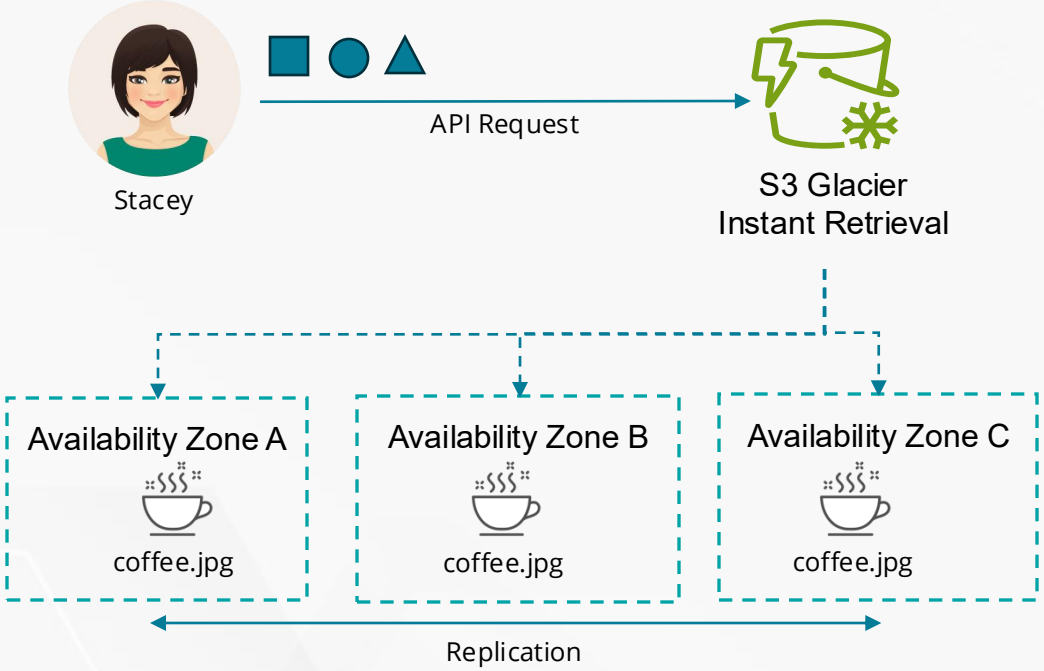
Do not use for critical data or data that cannot be replaced

Lisa - CFO

Costs based on:

- Storage fee GB/month for data stored
- Transfer OUT fee
- Request fee per 1,000 requests
- **Minimum storage charge of 128KB**
- **Minimum duration charge of 30 days**
- **Per GB retrieval cost**

Storage Class – S3 Glacier – Instant Retrieval



Designed for archival data
Cheaper than Standard Classes

Instant Access: milliseconds
first byte latency

Ideal for infrequently accessed data that is not replaceable, e.g. once per quarter

Durability

- Protection against data loss
- Prevent data corruption
- 99.999999999% (11 9s) durability
- 99.9% availability

Content-MD5 Checksums and Cyclic Redundancy Checks (CRC) for data integrity

Lisa - CFO

Costs based on:

- Storage fee GB/month for data stored
- Transfer OUT fee
- Request fee per 1,000 requests
- **Minimum storage charge of 128KB**
- **Minimum duration charge of 90 days**
- **Per GB retrieval cost**

Storage Class – S3 Glacier – Flexible Retrieval

Stacey

API Request

S3 Glacier
Flexible Retrieval

Availability Zone A
coffee.jpg

Availability Zone B
coffee.jpg

Availability Zone C
coffee.jpg

Replication

Designed for archival data
Cheaper than Standard Classes

Think: Cold Storage
Access requires a retrieval (restore) process.

Ideal for infrequently accessed archival data

Retrieval Process

- Initiate a Retrieval Job
- Data Retrieval - Tier
- Download Data

Durability

- Protection against data loss
- Prevent data corruption
- 99.999999999% (11 9s) durability
- 99.99% availability (after job restore operation)

Content-MD5 Checksums and Cyclic Redundancy Checks (CRC) for data integrity

No Public Access

Lisa - CFO

Costs based on:

- Storage fee GB/month for data stored
- Transfer OUT fee
- Request fee per 1,000 requests
- **Pay for the retrieval (restore) process**
- **Minimum duration charge of 90 days**
- **Per GB retrieval cost**

Retrieval Options – data is temporarily moved to Standard IA for download

Expedited Retrieval	Standard Retrieval	Bulk Retrieval
1 to 5 mins (ideal for subset of data)	3 to 5 hours	5 to 12 hours

Storage Class – S3 Glacier – Deep Archive

Stacey

API Request

S3 Glacier Deep Archive

Availability Zone A

coffee.jpg

Availability Zone B

coffee.jpg

Availability Zone C

coffee.jpg

Replication

Designed for archival data
Cheaper than Glacier Flexible Retrieval

Think: Cold Storage
Access requires a retrieval (restore) process.

Ideal for infrequently accessed archival data
Use to host data for very long term due to regulatory requirements

Retrieval Process

- Initiate a Retrieval Job
- Data Retrieval - Tier
- Download Data

Durability

- Protection against data loss
- Prevent data corruption
- 99.999999999% (11 9s) durability
- 99.99% availability (after job restore operation)

Content-MD5 Checksums and Cyclic Redundancy Checks (CRC) for data integrity

No Public Access

Lisa - CFO

Costs based on:

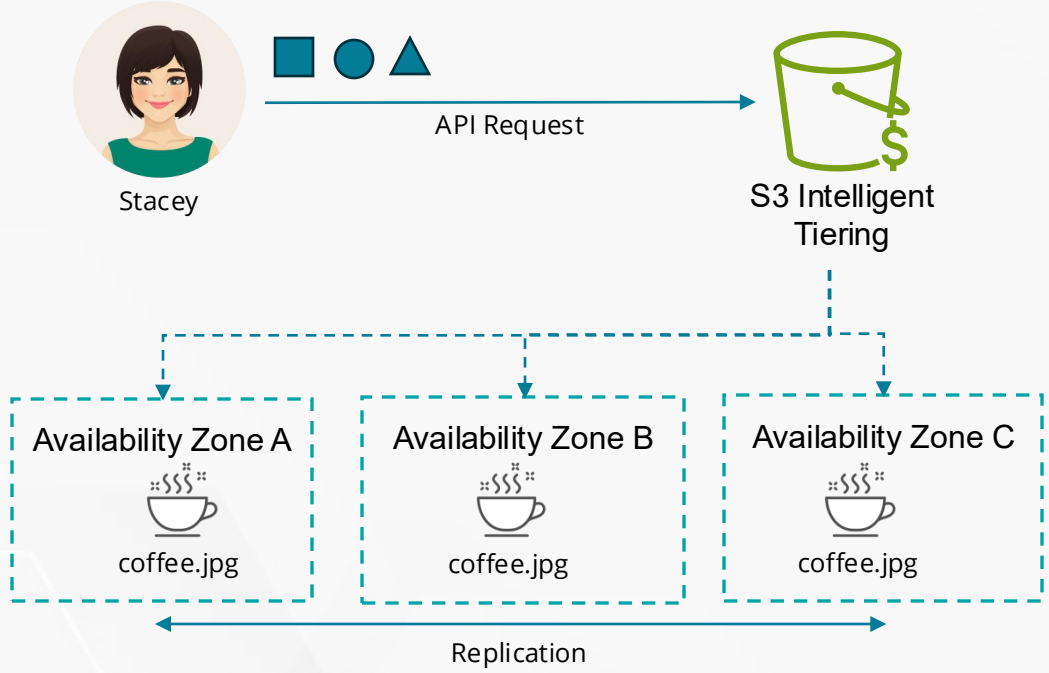
- Storage fee GB/month for data stored
- Transfer OUT fee
- Request fee per 1,000 requests
- **Pay for the retrieval (restore) process**
- **A minimum storage charge of 40 KB**
- **Minimum duration charge of 180 days**
- **Per GB retrieval cost**

Retrieval Options – data is temporarily moved to Standard IA for download

Standard Retrieval
Up to 12 hours

Bulk Retrieval
Within 48 hours

Storage Class – S3 Intelligent Tiering

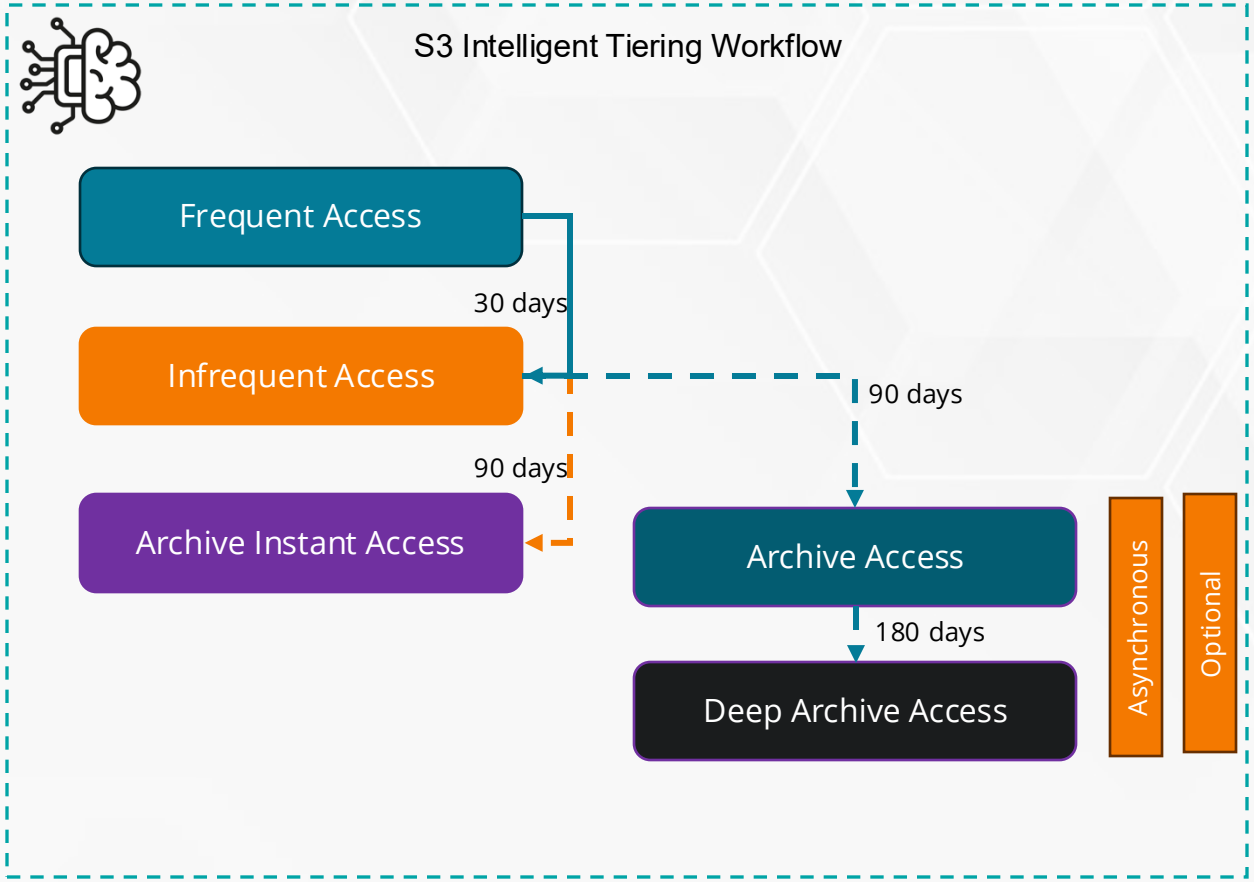


Lisa - CFO

Costs based on:

- Storage fee GB/month for data stored
- Transfer OUT fee
- Request fee per 1,000 requests
- **Monitoring and automation fee per object**
- **No retrieval costs**

Designed for data where the access pattern is unknown or can change quickly. Objects not accessed for 30 days are moved into low cost infrequent tier and further into archive instant access, archive access and deep archive access. If the object is accessed again it is moved into Standard tier.



Storage Class – S3 Express One-Zone

Stacey

API Request

S3 Express One Zone

Choose to co-locate your storage and compute resources in same AZ

Availability Zone *x*

S3 directory bucket

Single-Availability Zone storage class

Frequent access to single-digit millisecond latency-sensitive apps

10x faster & 50% lower than Standard

Use cases:

- Generative AI/ML model training
- Transactional tables/data lakes
- Video content creation
- Financial model simulations
- Genomics processing
- Image rendering

Directory buckets organize data into directory. No prefix limits and horizontally scalable. Can support hundreds of thousands of transactions per second (TPS)

Lisa - CFO

Costs based on:

- Storage fee GB/month for total storage used per hour
- Charged based on request type – PUTs and GETs
- Additional per GB fee for the portion of request size exceeding 512KB

Create an Amazon S3 Bucket

Features and Configuration

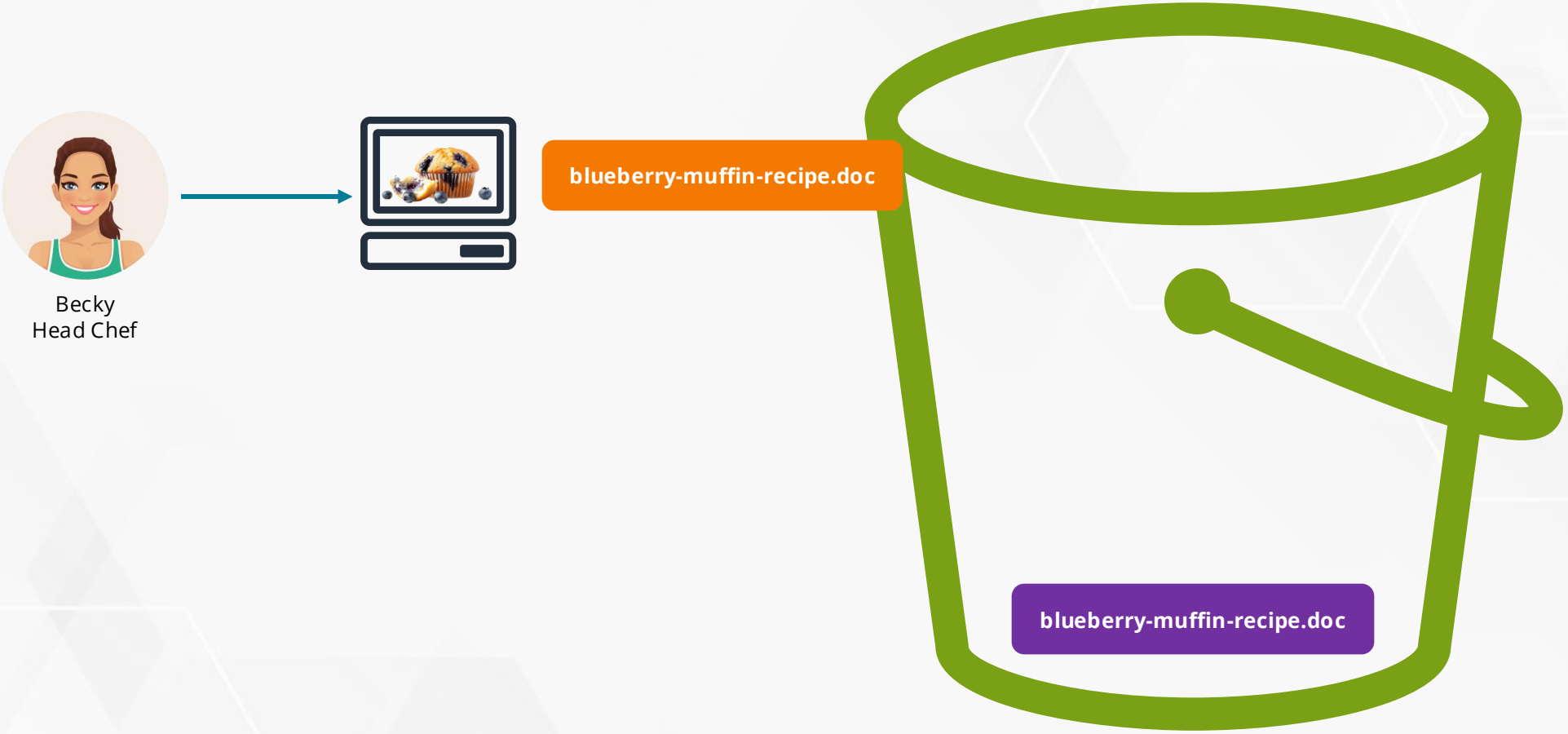
IAAS ACADEMY

Amazon S3 Object Versioning

Prevent accidental deletions and overwrites

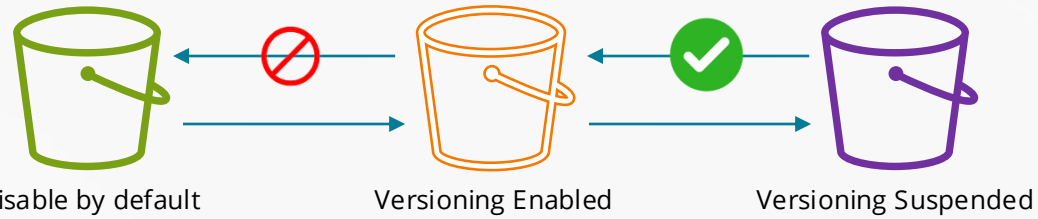
IAAS ACADEMY

Object Versioning



Amazon S3 Buckets
Default: Versioning Disabled

Bucket Versioning States



Becky
Head Chef

Delete

Delete

Delete ID = 111111

Versioning Disabled

blueberry-muffin-recipe.doc
ID = Null

Versioning Enabled

blueberry-muffin-recipe.doc
Delete Marker

blueberry-muffin-recipe.doc
ID = 222222

blueberry-muffin-recipe.doc
ID = 111111

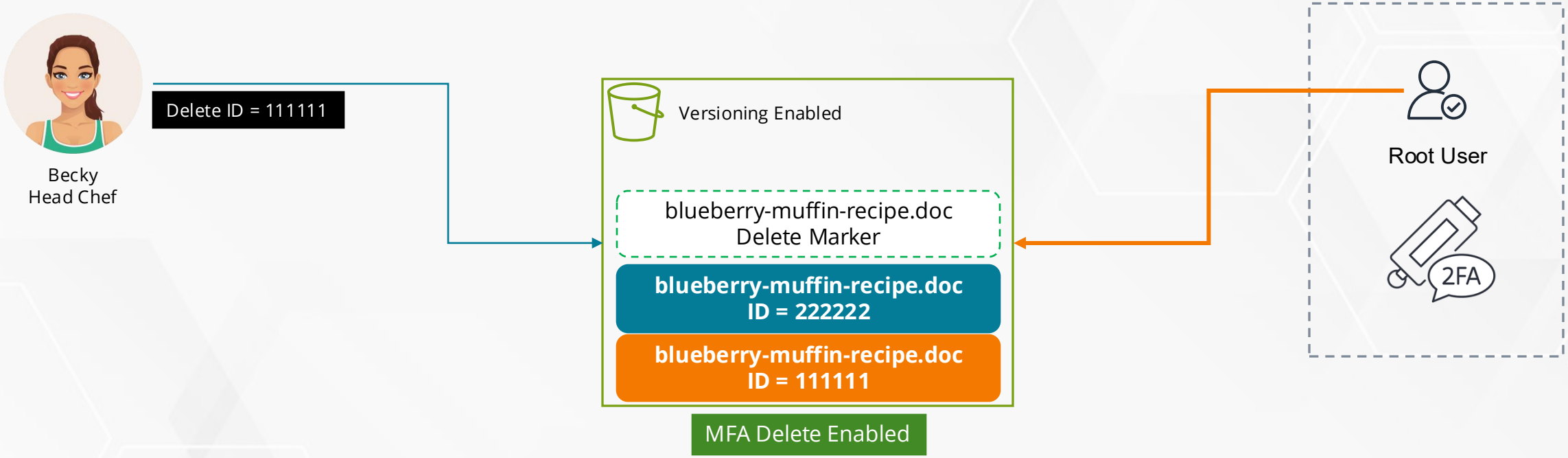
Delete the Delete Marker

Amazon S3 MFA Delete

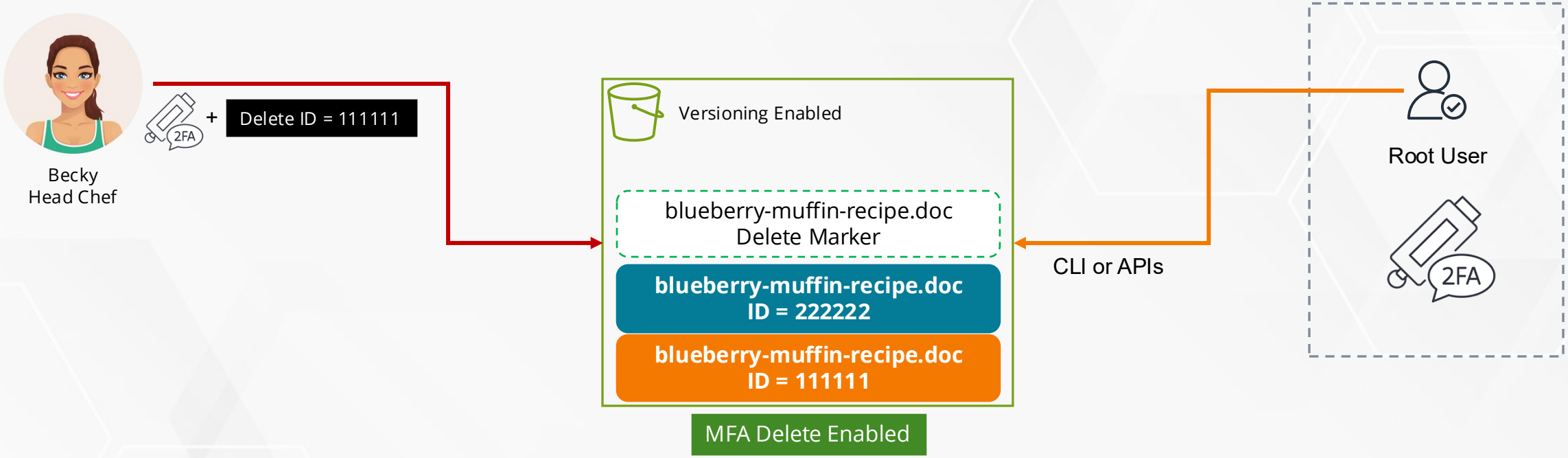
Using MFA Delete to prevent against accidental deletions

IAAS ACADEMY

Bucket Versioning States – Enabling MFA Delete



Bucket Versioning States – Enabling MFA Delete

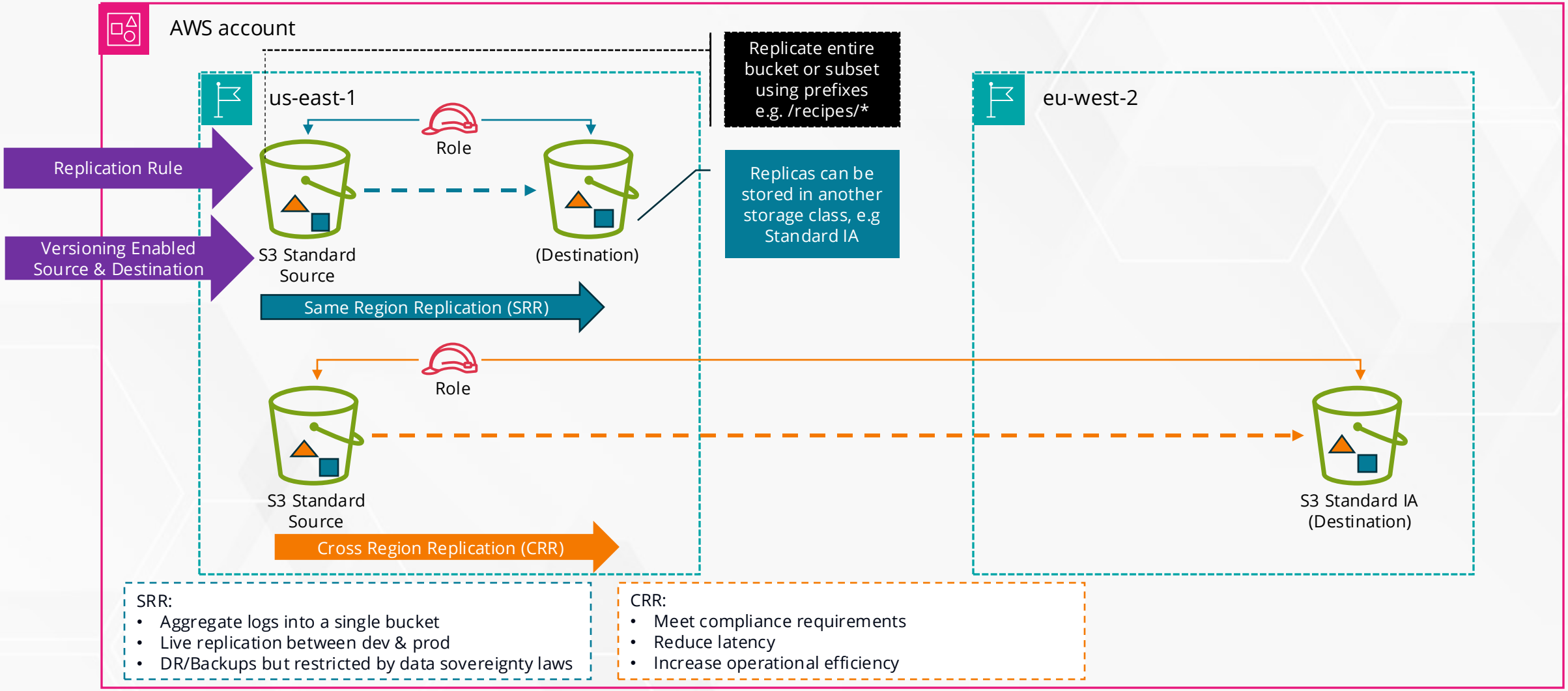


Amazon S3 Bucket Replication

Cross-Region and Same Region Replication

IAAS ACADEMY

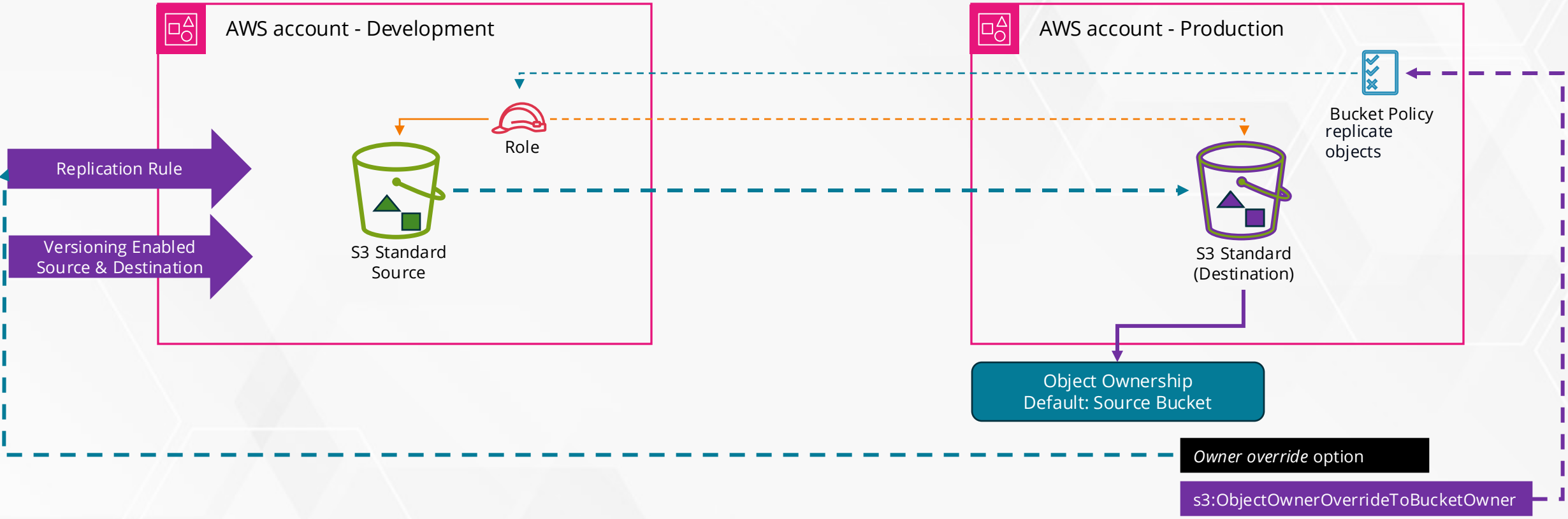
Amazon S3 Bucket Replication



SRR:
- Aggregate logs into a single bucket
- Live replication between dev & prod
- DR/Backups but restricted by data sovereignty laws

CRR:
- Meet compliance requirements
- Reduce latency
- Increase operational efficiency

Amazon S3 Bucket Replication – Cross Account



Amazon S3 Replication – Additional Points

- Replication Time Control (**RTC**) – **15 Minutes** Guarantee Window
- Default Replication Type – Live Replication (Not Retroactive) - automatically replicates new and updated objects
- On-demand replication – To replicate existing objects using S3 Batch Replication
- Not bidirectional by default
- Only user events are replicated (not system events)
- **Deletes are not replicated by <u>default</u>** (**Delete Markers Not Replicated**)

Configure Amazon S3 Replication

Cross Account Replication Configuration

IAAS ACADEMY

Amazon S3 Performance

Scaling and Performance Strategies

IAS ACADEMY

Using Prefixes to Improve Performance

Amazon S3 scales well and applications can achieve thousands of requests per second when uploading and retrieving from Amazon S3.

- 3,500 PUT/COPY/POST/DELETE
- 5,500 GET/HEAD
- No limits to the number of prefixes in a bucket
- Increase performance by parallelizing transactions

per prefix in a bucket

Example

- customerrecipe/image-a-e/*file...file...file*
- customerrecipe/image-f-j/*file...file...file*
- customerrecipe/image-k-o/*file...file...file*
- custoemrrecipe/image-p-t/*file...file...file*
- customerrecipe/image-u-z/file...file...file

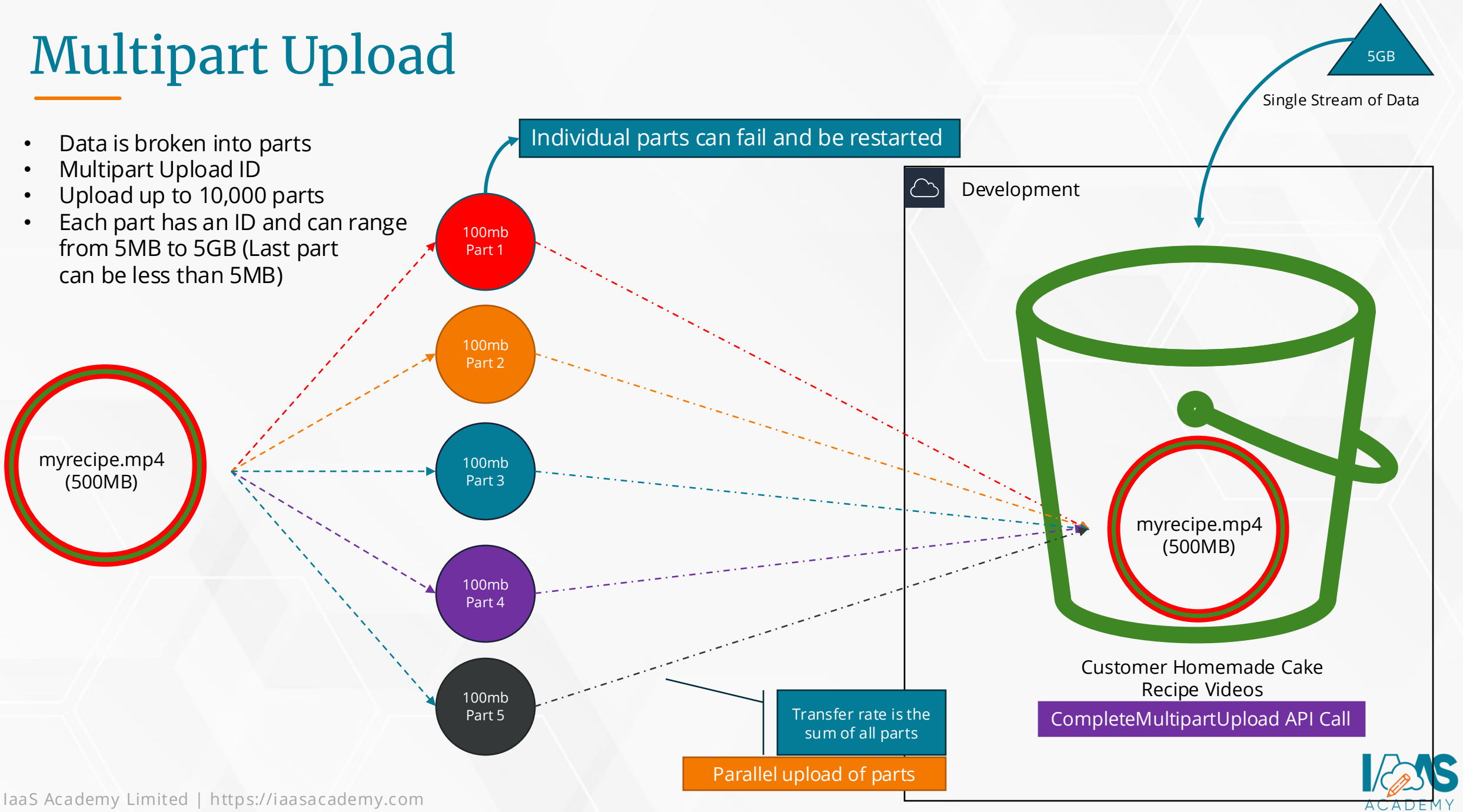
Spreading reads across all 5 prefixes, and parallelize those reads, gives you 27,500 read requests per second

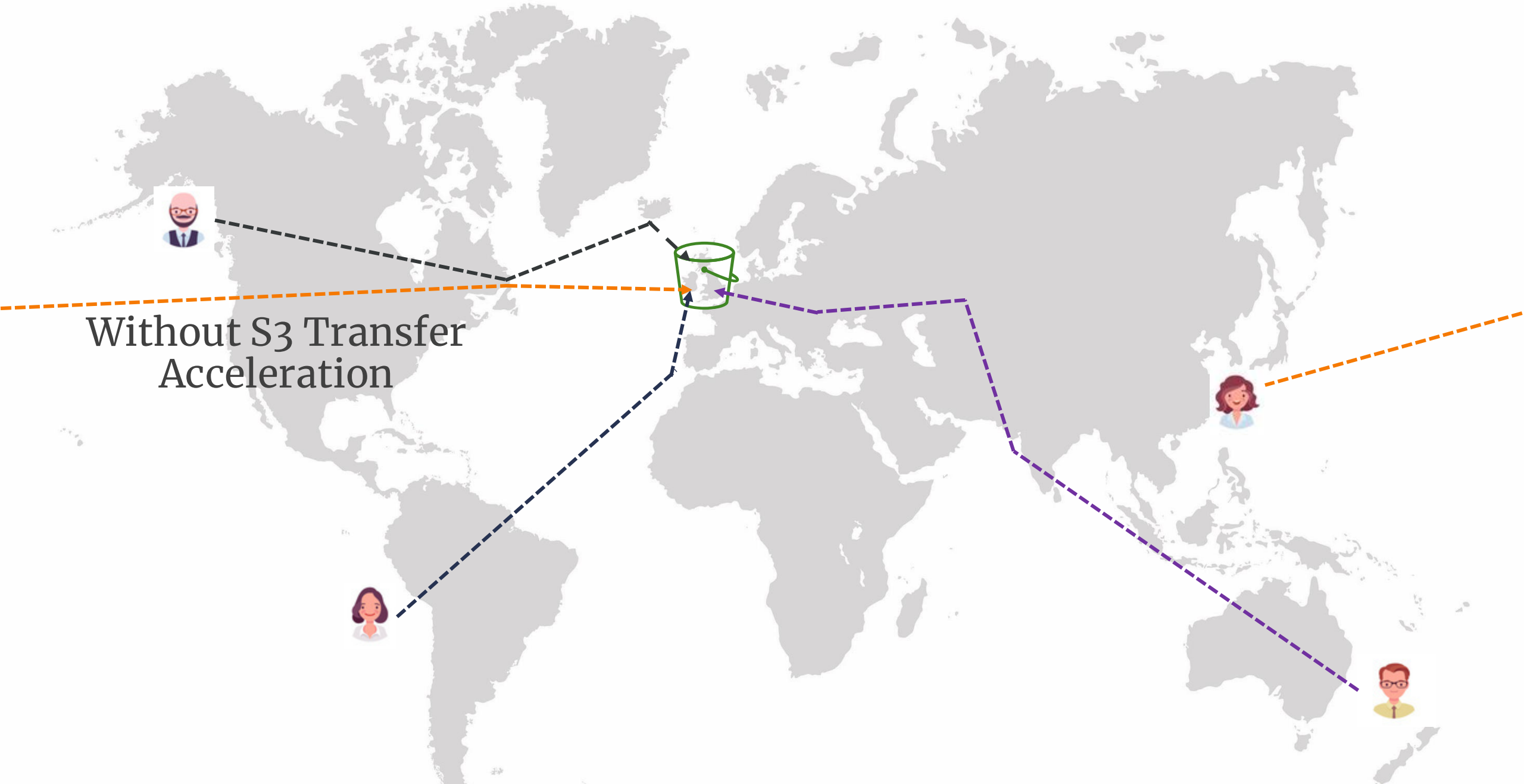
RITUAL ROAST

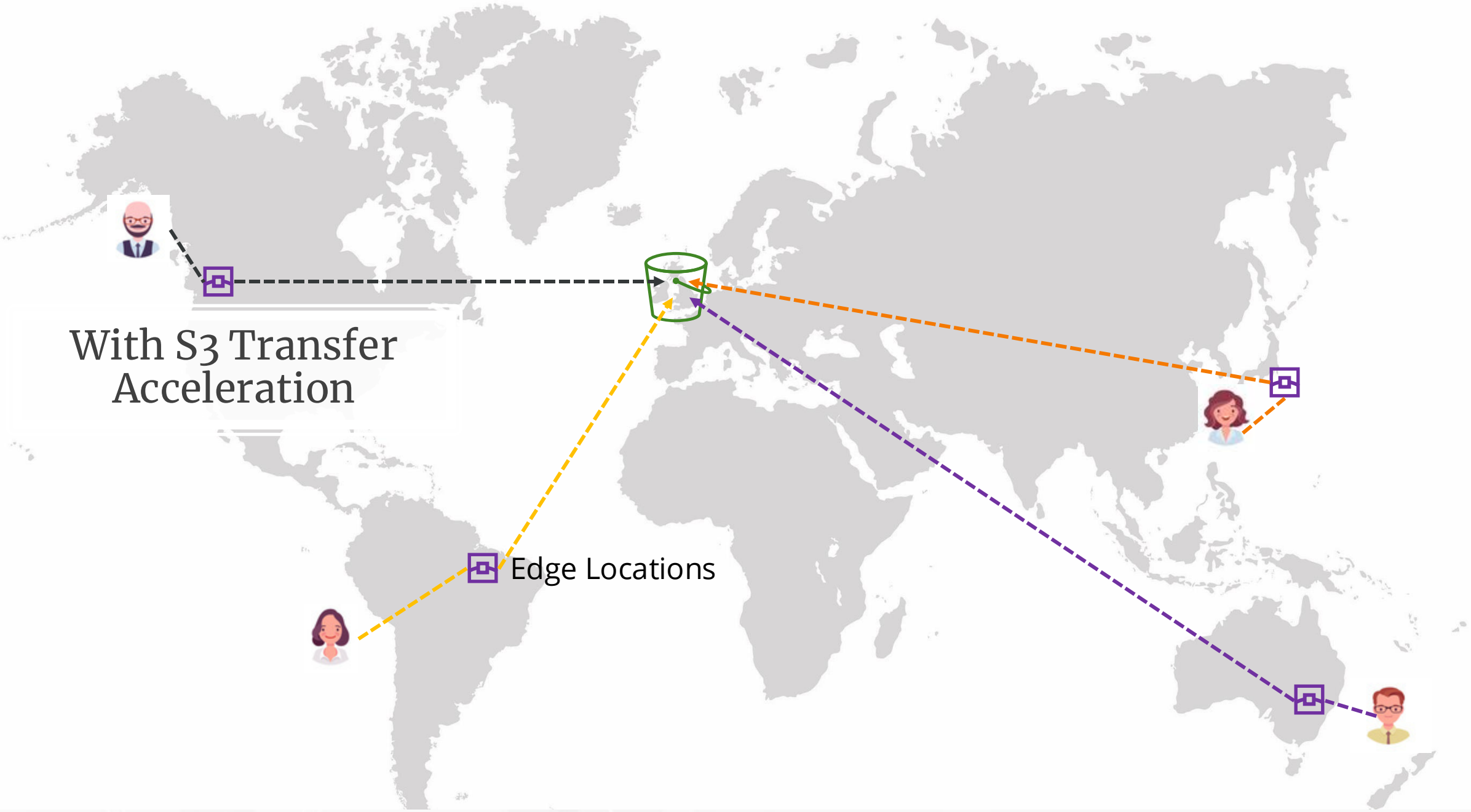
Customer Scenario – Ritual Roast Vlog Contest

Multipart Upload

- Data is broken into parts
- Multipart Upload ID
- Upload up to 10,000 parts
- Each part has an ID and can range from 5MB to 5GB (Last part can be less than 5MB)




Without S3 Transfer Acceleration

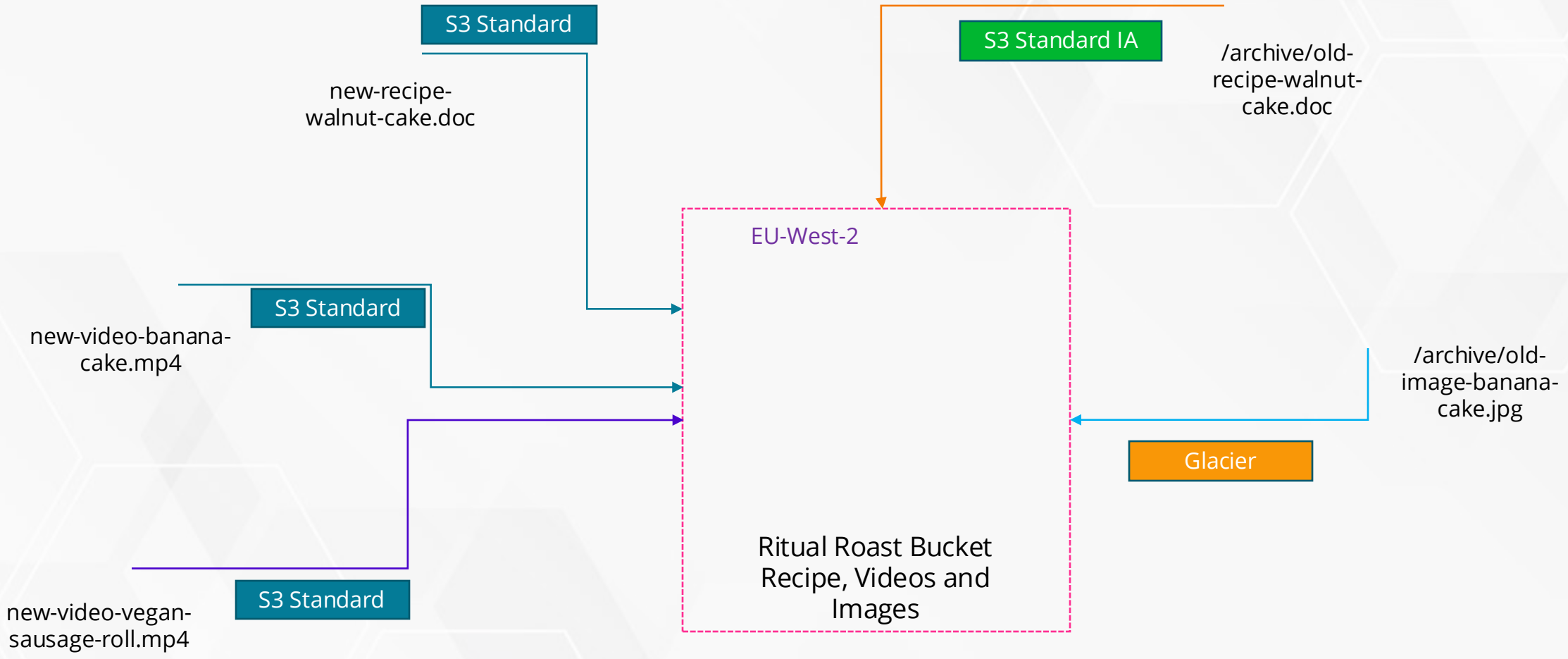

With S3 Transfer Acceleration
Edge Locations

Amazon S3 Lifecycle Management

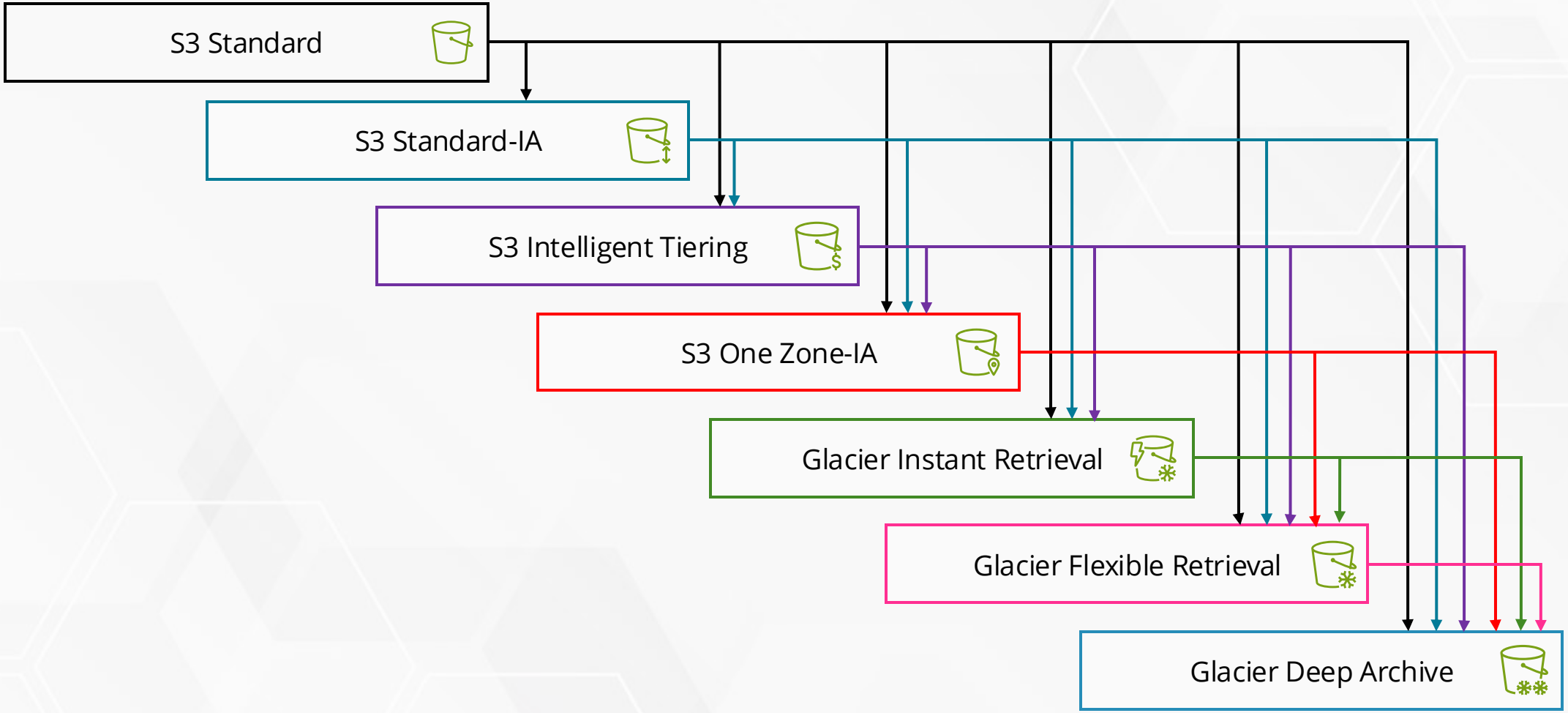
Managing Data Lifecycle in your Organization

IAAS ACADEMY

Amazon S3 Bucket – Objects Storage Classes



Amazon S3 Lifecycle Management

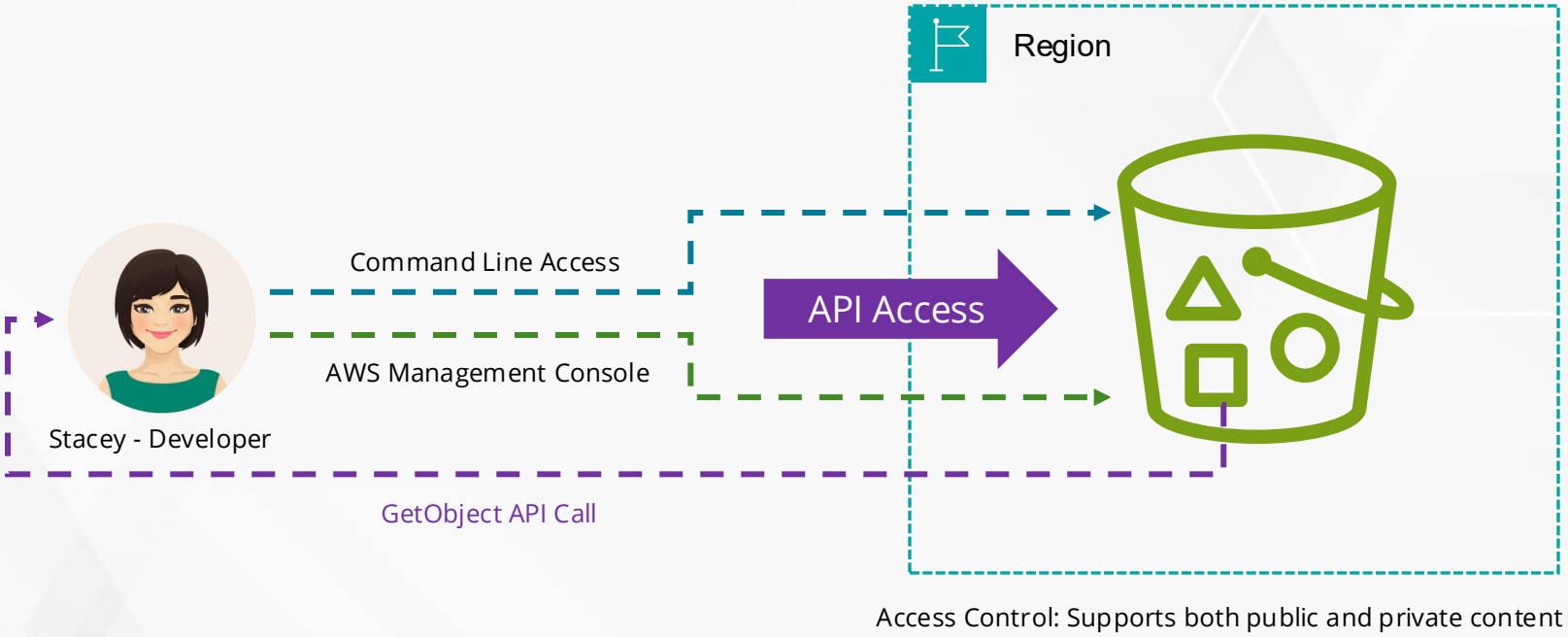


Amazon S3 Static Website Hosting

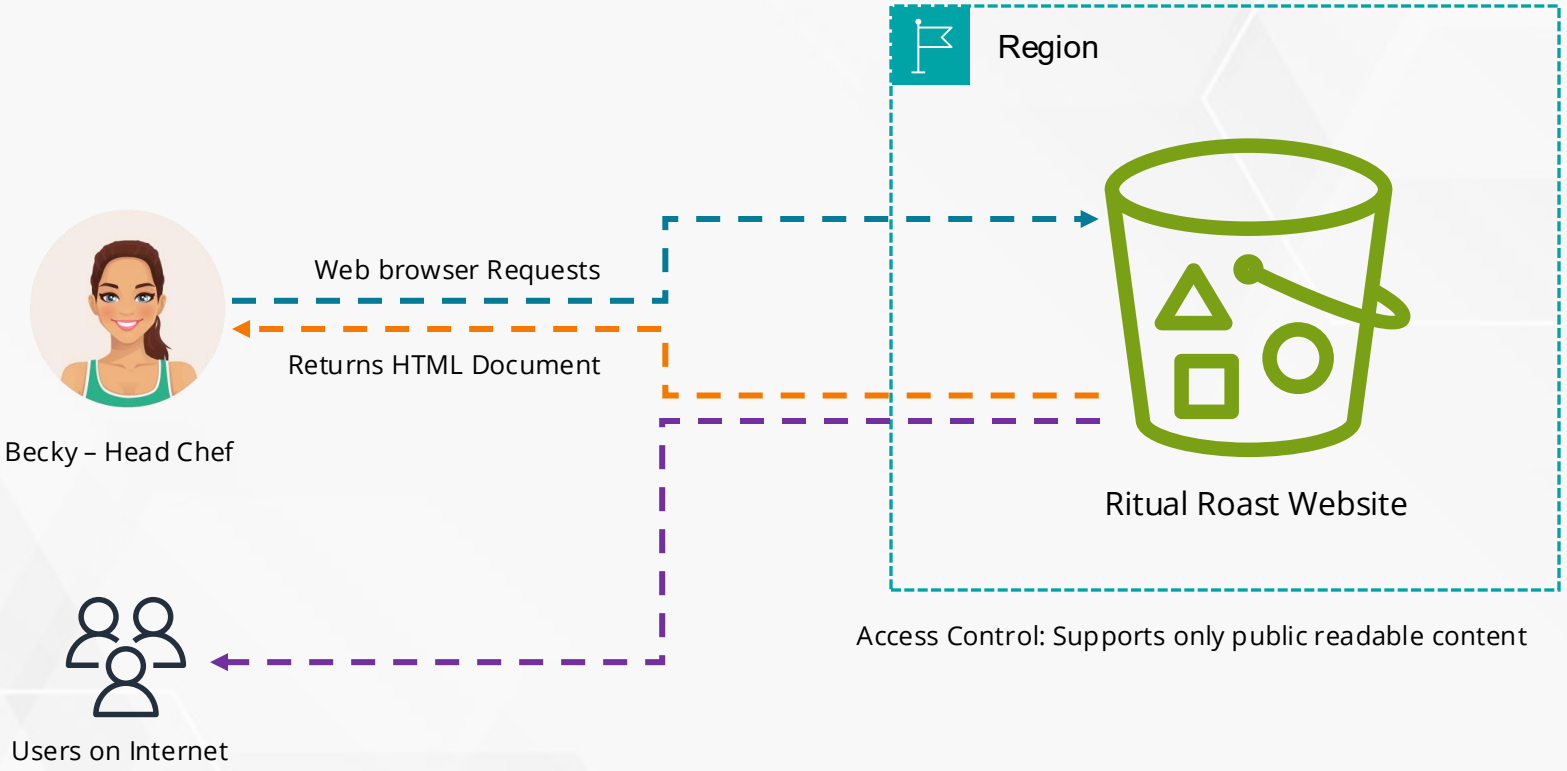
Hosting Front-End Static Websites

IAAS ACADEMY

Accessing Amazon S3 Buckets – API Access



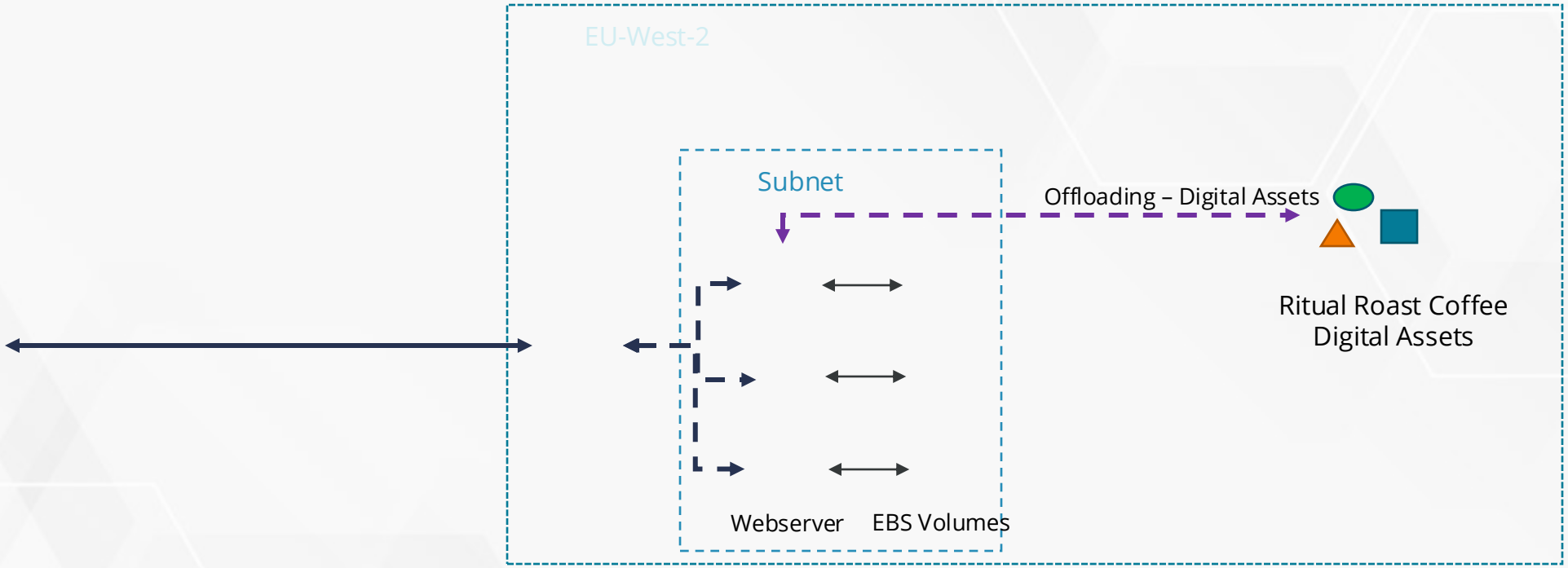
Accessing Amazon S3 Buckets – Website Endpoint



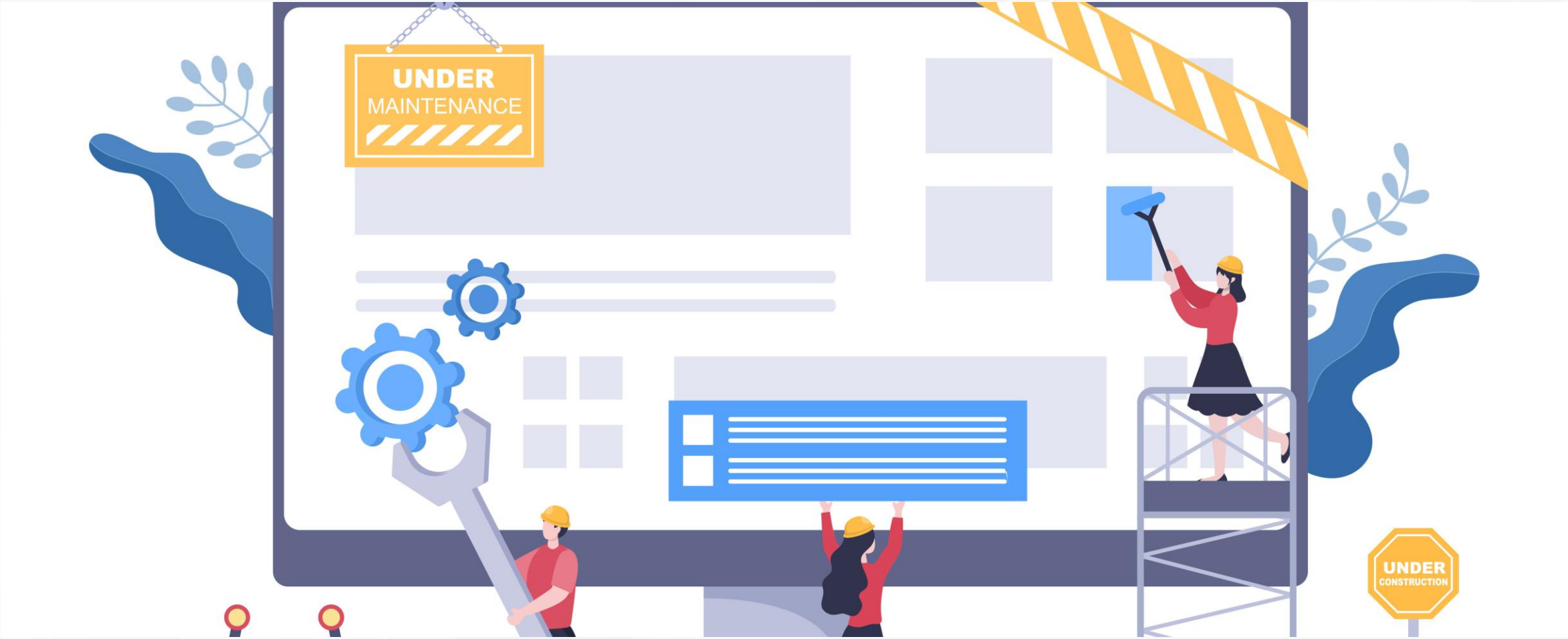
URL Format - Region Dependent
s3-website dash (-) Region - http://bucket-name.s3-website.Region.amazonaws.com
s3-website dot (.) Region - http://bucket-name.s3-website-Region.amazonaws.com

Static Website Hosting – Product Launch

Static Website Hosting – Offloading



Static Website Hosting – Under Maintenance



Static Website Hosting

To configure a bucket to host a static website, follow these steps:

- Create a bucket with the name that will be your website hostname
- Upload static files to the bucket
  - You will need an **index.html** file
  - You will need an **error.html** file
- Make all files public or grant everyone read access
- Enable the bucket for Static Website Hosting
- Optionally configure a DNS CNAME to point to the Amazon URL. However, your **bucket name must match the custom domain name.**

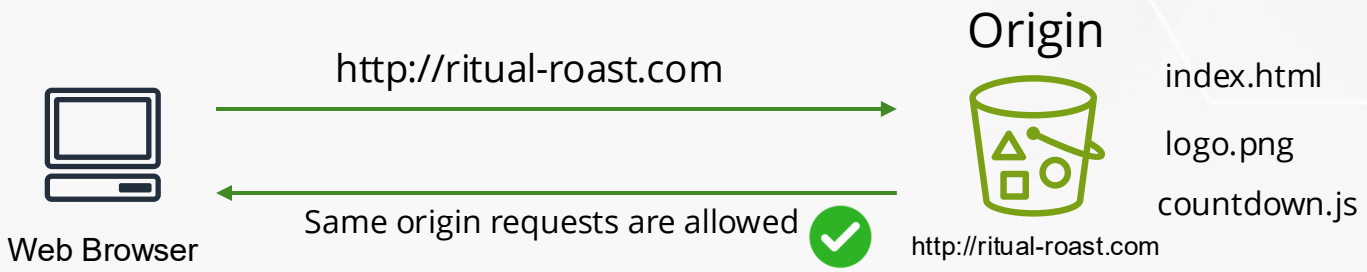
Cross Origin Resource Sharing

Introduction to CORS

IAS ACADEMY

What is CORS

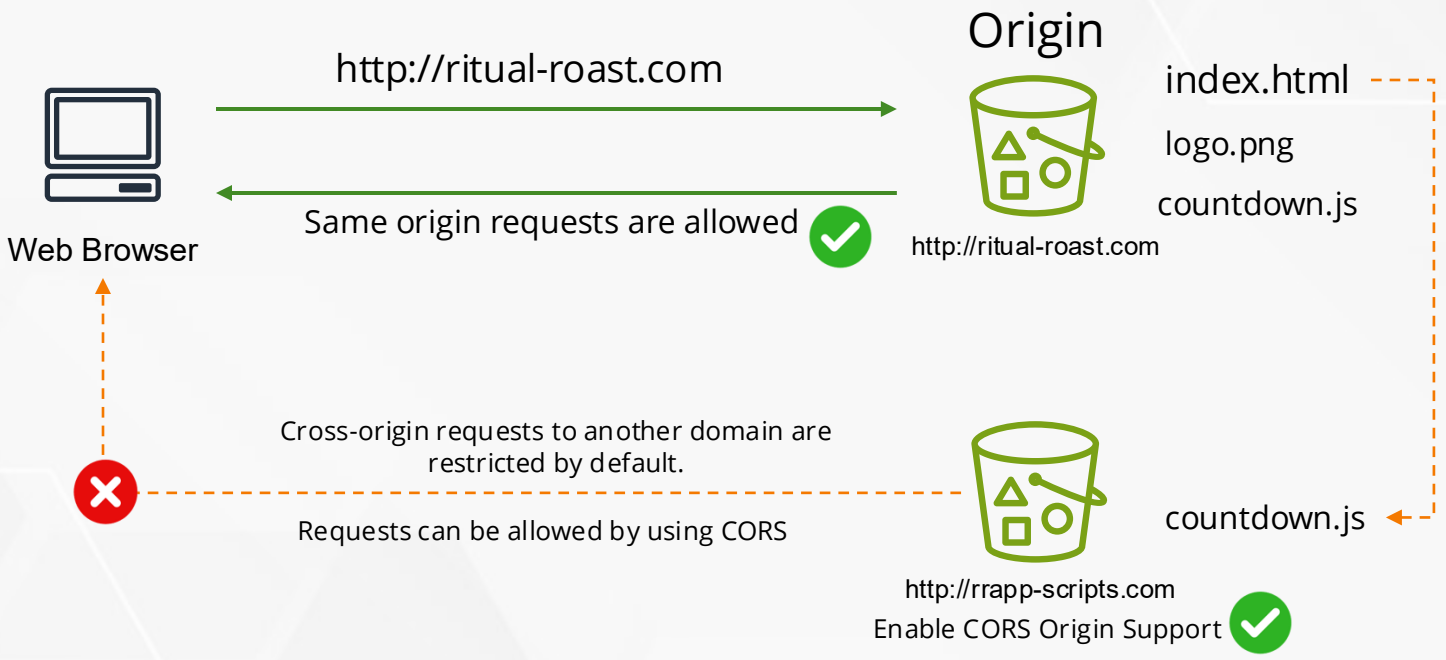
Browser security feature allows client web applications loaded in one domain to interact with resources in a different domain.



What is CORS

Browser security feature allows client web applications loaded in one domain to interact with resources in a different domain.

Origin = scheme(protocol) + host(host) + port
Example: http://ritual-roast.com (port 80 for HTTP or 443 for HTTPS)



Additional Points

- Simple Requests
- Preflight requests & responses
- Access-Control-Allow-Origin (specify the origins that you want to allow cross-domain requests from)
- Access-Control-Max-Age
- Access-Control-Allow-Methods (GET, PUT, POST, DELETE, HEAD)
- Access-Control-Allow-Headers

Example CORS Configuration

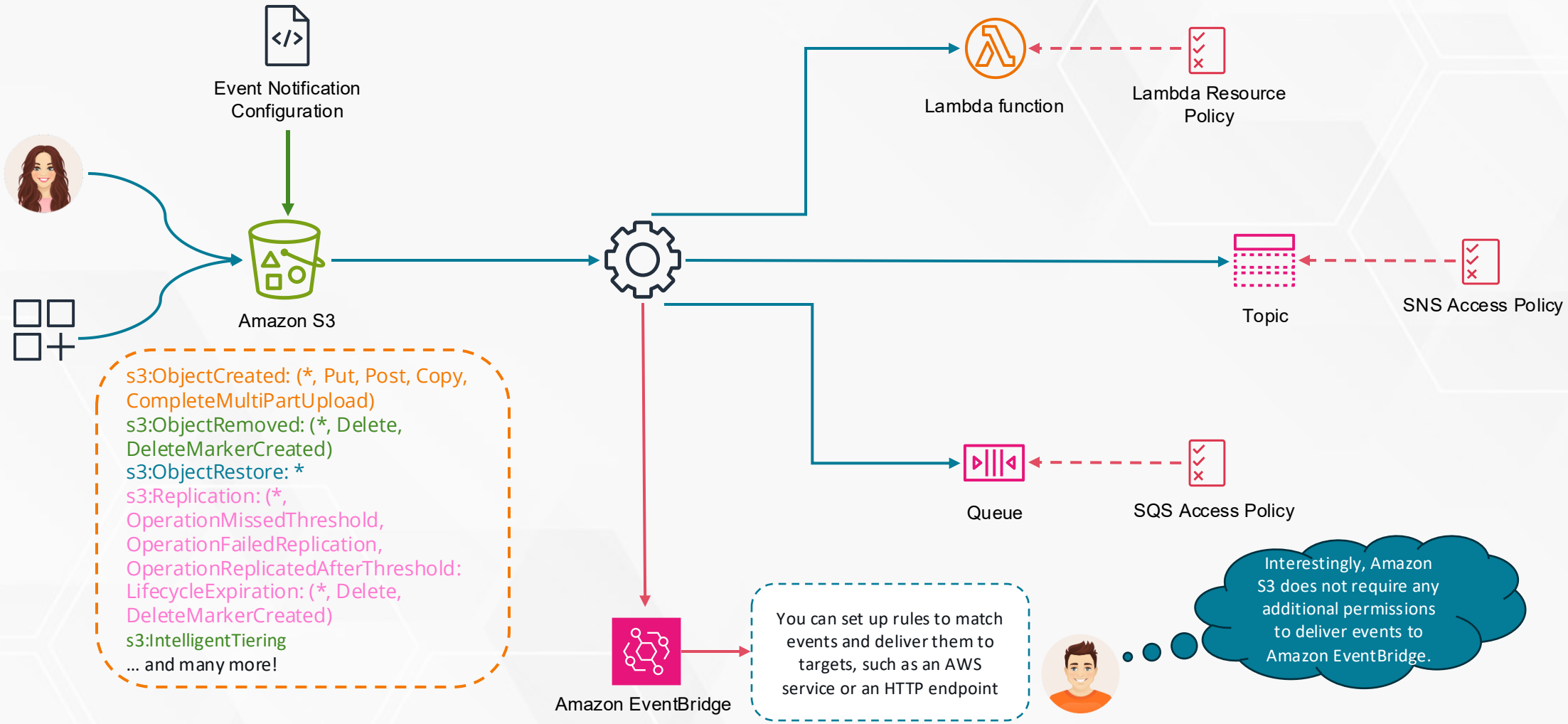
```
[
  {
    "AllowedHeaders": [
      "*"
    ],
    "AllowedMethods": [
      "PUT",
      "POST",
      "DELETE"
    ],
    "AllowedOrigins": [
      "http://www.example1.com"
    ],
    "ExposeHeaders": []
  },
  {
    "AllowedHeaders": [],
    "AllowedMethods": [
      "GET"
    ],
    "AllowedOrigins": [
      "*"
    ],
    "ExposeHeaders": []
  }
]
```

Amazon S3 Event Notification

IAAS ACADEMY

Amazon S3 Event Notification

Use the Amazon S3 Event Notifications feature to receive notifications when certain events happen in your S3 bucket



Sample SNS Access Policy


```
{
    "Version": "2012-10-17",
    "Id": "example-ID",
    "Statement": [
        {
            "Sid": "Example SNS topic policy",
            "Effect": "Allow",
            "Principal": {
                "Service": "s3.amazonaws.com"
            },
            "Action": [
                "SNS:Publish"
            ],
            "Resource": "SNS-topic-ARN",
            "Condition": {
                "ArnLike": {
                    "aws:SourceArn": "arn:aws:s3:*:*:amzn-s3-demo-bucket"
                },
                "StringEquals": {
                    "aws:SourceAccount": "bucket-owner-account-id"
                }
            }
        }
    ]
}
```

Amazon S3 Encryption Options

Encrypting Your Amazon S3 Objects

IAAS ACADEMY

Introduction to Amazon S3 Encryption

Encryption is defined at a per object level, not bucket level. Each object might use a different encryption setting

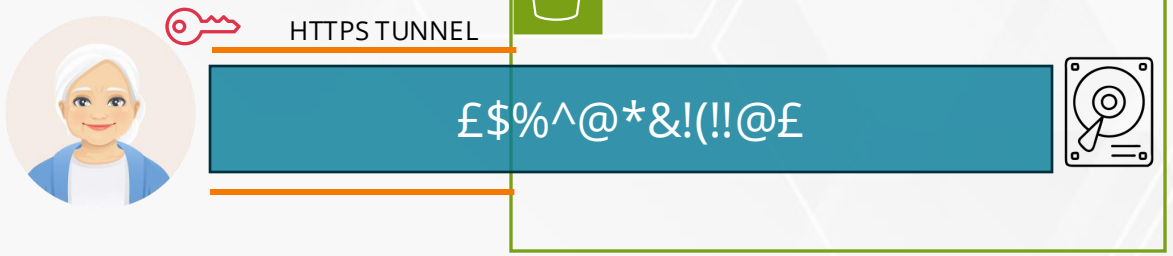
Server-side encryption is now mandatory

Encryption **in transit** vs **at rest**

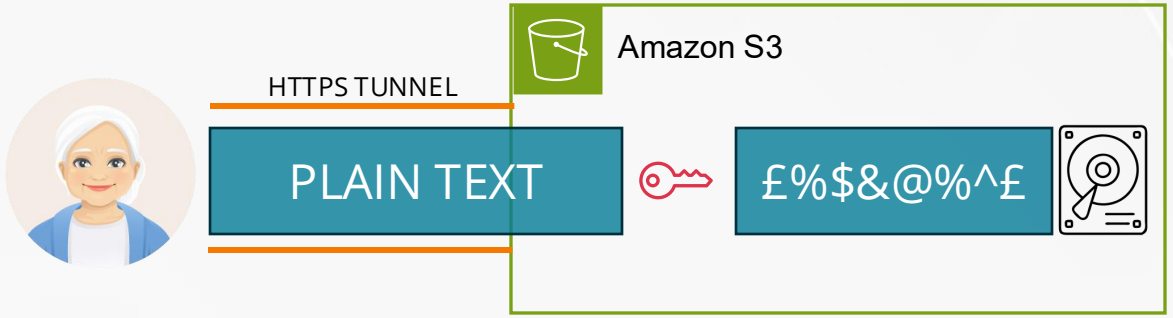
With Client-side encryption:

- Your objects aren't exposed to any third party, including AWS
- Amazon S3 does not play a role in encrypting or decrypting your objects
- You must use the Amazon S3 Encryption Client to automatically encrypted and decrypted as part of your Amazon S3 PutObject and GetObject requests

Client-side Encryption



Server-side Encryption

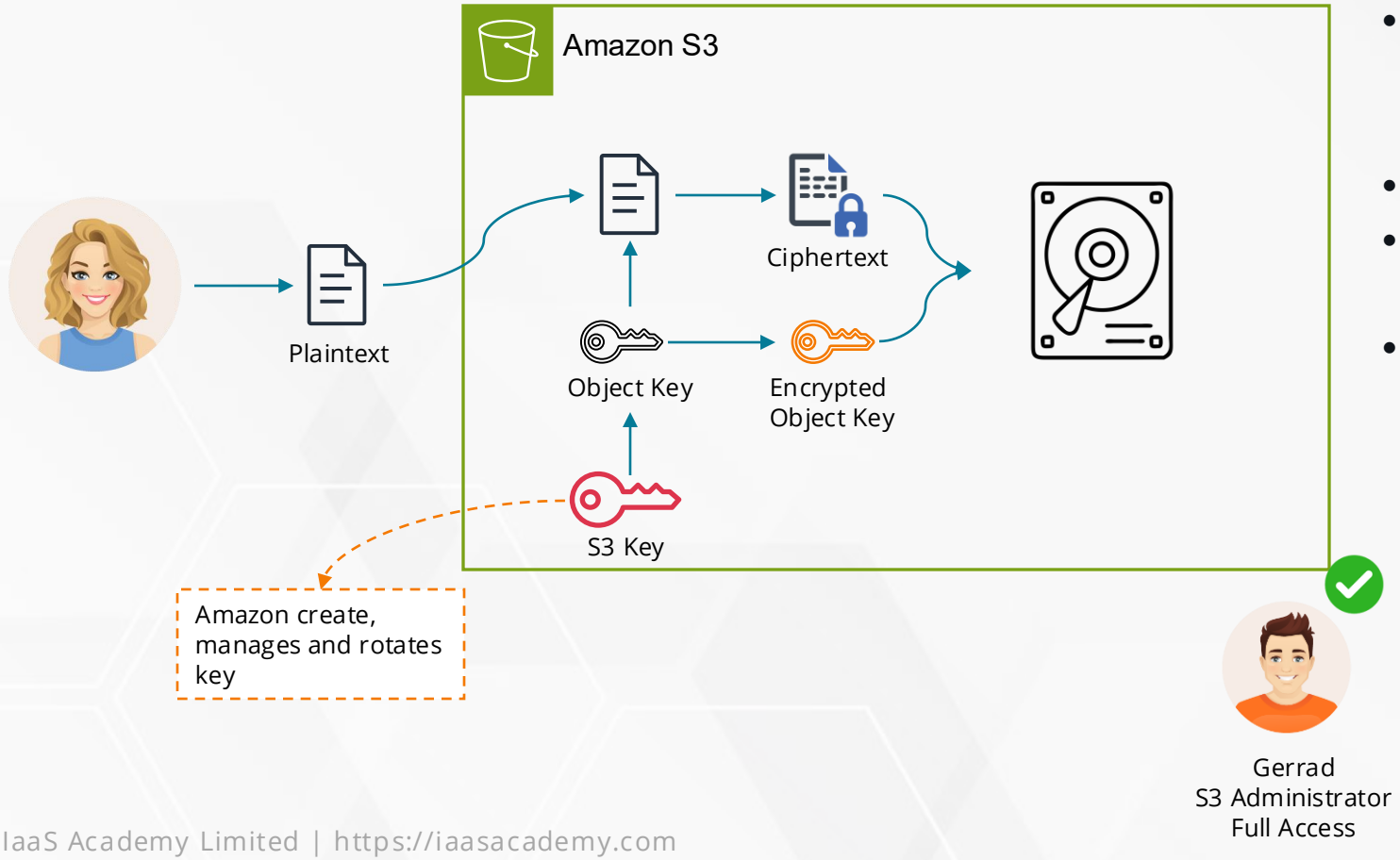


Introduction to Amazon S3 Encryption

AWS manages creation, deletion and usage of keys – Zero Control on Keys!

Other users with admin access can access data = **No Role Separation**

Server-side encryption with Amazon S3 managed keys (SSE-S3)

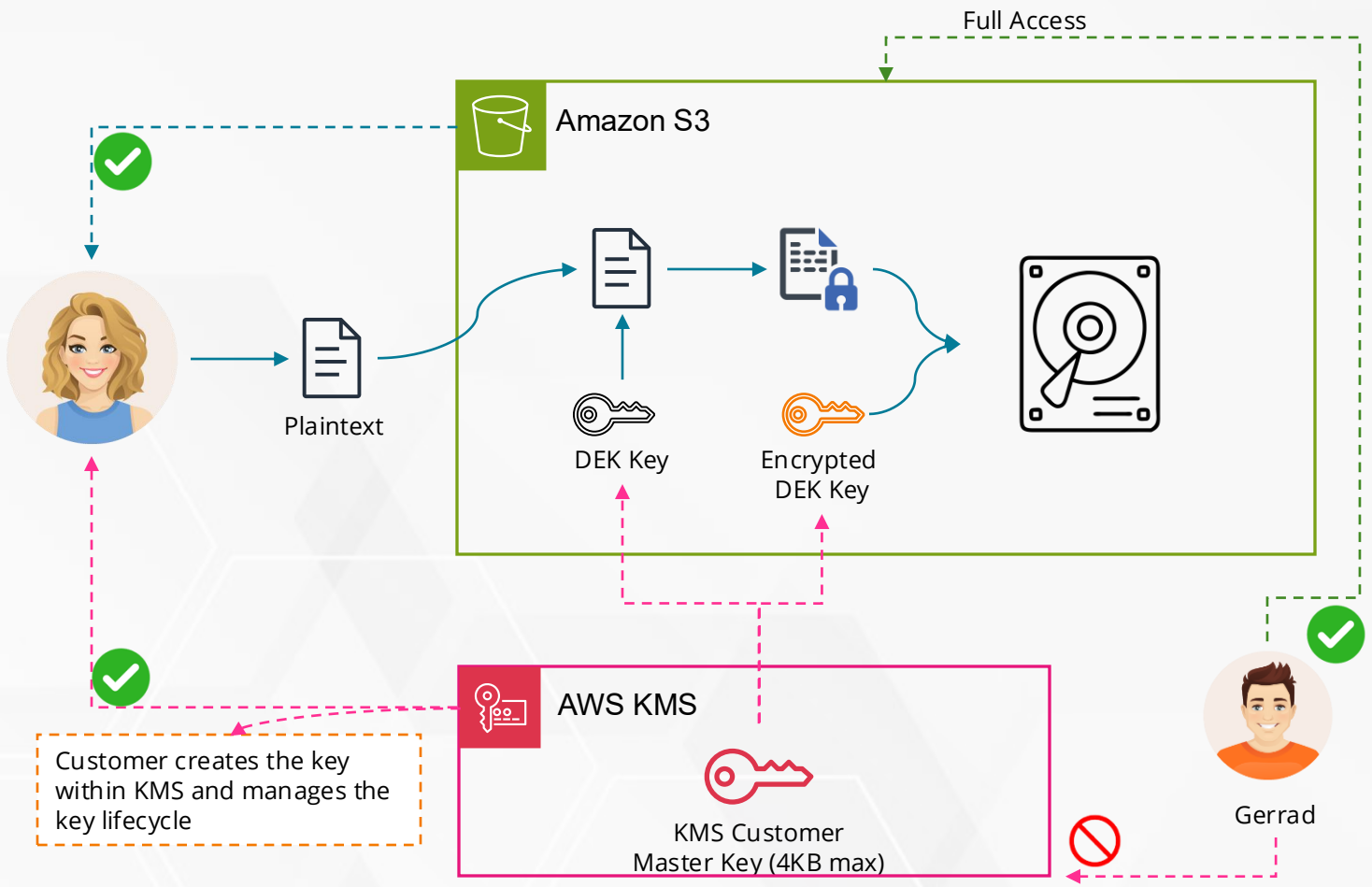


- **Server-side encryption with Amazon S3 managed keys (SSE-S3)** – Default encryption (no cost)
- AWS creates and manages the keys for you
- SS3-S3 is applied to all objects uploaded to the bucket
- Using object creation REST APIs, you must provide the ”**x-amz-server-side-encryption”:”AES256”** request header

Introduction to Amazon S3 Encryption

Note: AWS KMS keys must be in the same Region as the bucket.

Server-side encryption with AWS KMS keys (SSE-KMS)

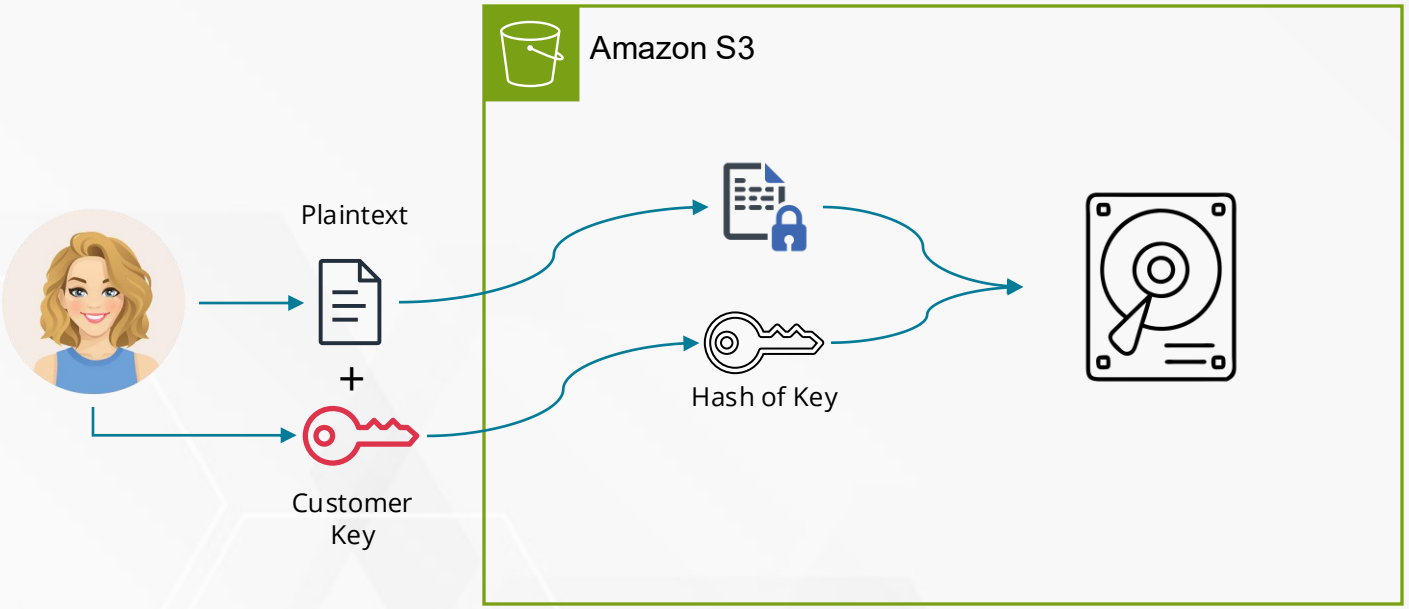


- **Server-side encryption with AWS KMS keys (SSE-KMS)**
- Centrally create, view, edit, monitor, enable or disable, rotate, and schedule deletion of KMS keys.
- Define the policies that control how and by whom KMS keys can be used – **Role Separation**
- Audit their usage to prove that they are being used correctly (CloudTrail)
- Ideal for regulated industries
- Using object creation REST APIs, you must provide the ”**x-amz-server-side-encryption”:”aws:kms”** request header.

Dual-layer server-side encryption with AWS Key Management Service (AWS KMS) keys (DSSE-KMS) applies two layers of encryption to objects when they are uploaded to Amazon S3.

Introduction to Amazon S3 Encryption

Server-side encryption with Customer Provided Keys(SSE-C)



- **Server-side encryption with Customer Provided Keys (SSE-C)**
- You store your data encrypted with your own encryption keys
- Amazon S3 manages data encryption as it writes to disks and data decryption when you access your objects.
- Amazon S3 then removes the encryption key from memory
- To retrieve an object, you must provide the same encryption key as part of your request.

Amazon S3 stores a randomly salted Hash-based Message Authentication Code (HMAC) value of the encryption key to validate future requests. The HAMC cannot be used to derive the value of the encryption key or to decrypt the contents of the encrypted object. That means if you lose the encryption key, you lose the object.

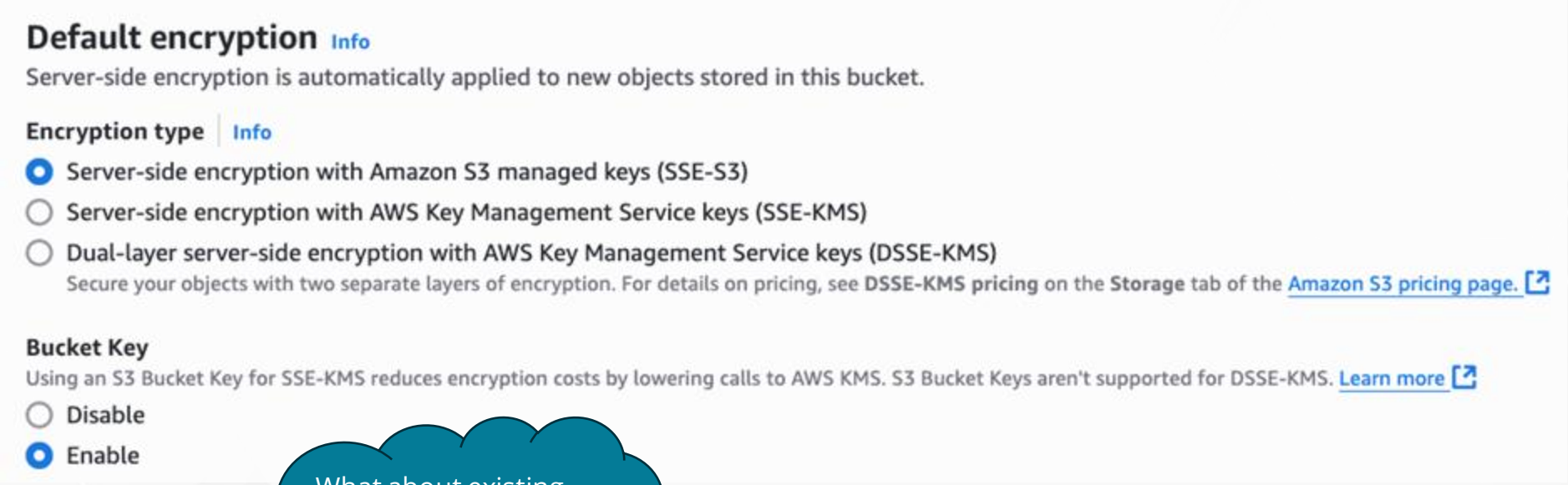
Amazon S3 Encryption – KMS

Additional Concepts on KMS

IAAS ACADEMY

Encrypting Existing Objects

- All new buckets have default SSE-S3 encryption enforced



Menio

That's easy, you can use AWS S3 Batch Operations to encrypt existing objects.
You can encrypt existing objects using **CopyObject** API operation or **copy-object** CLI command

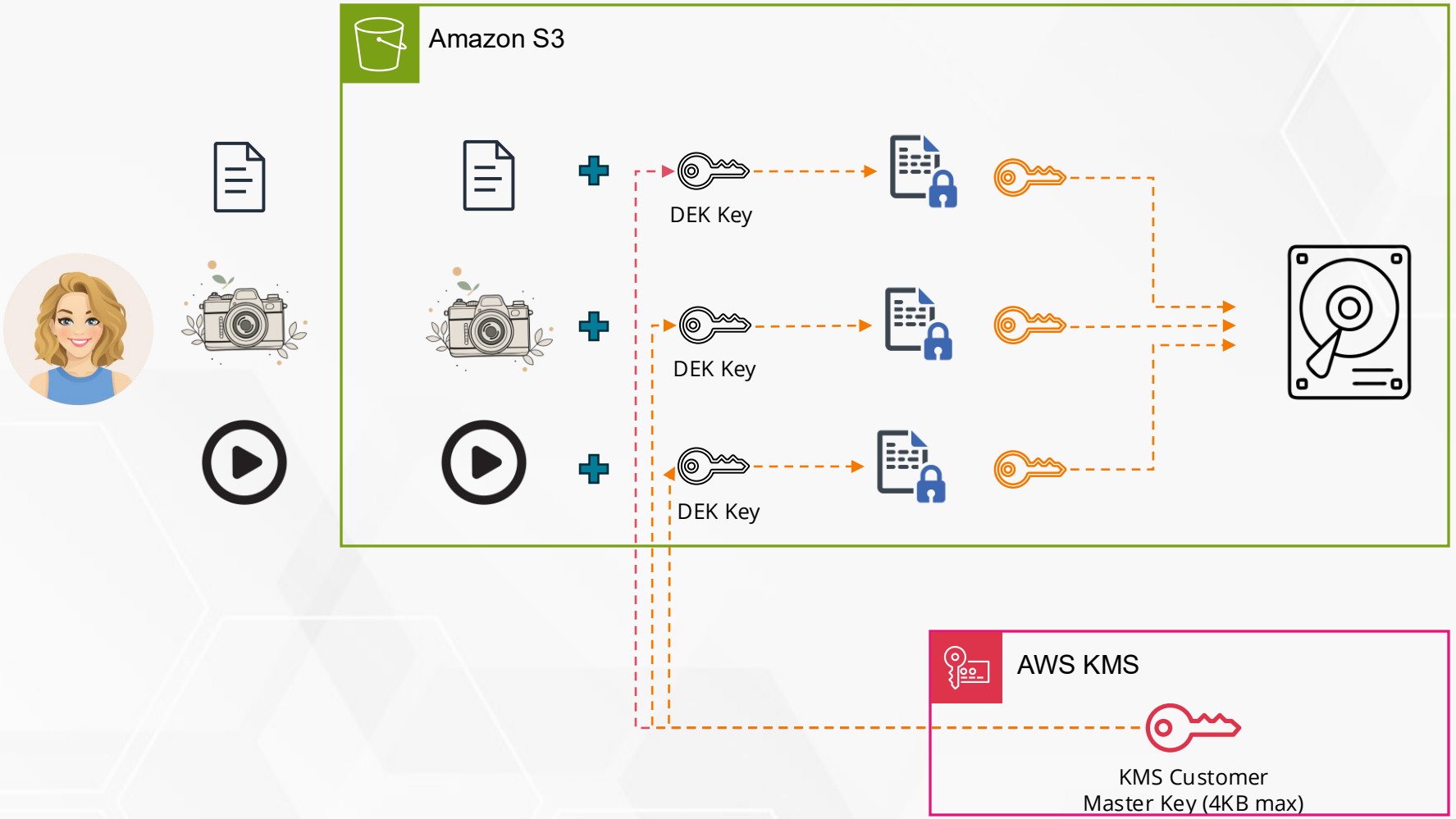
Using Bucket Policies to Enforce Encryption

```
{
    "Version":"2012-10-17",
    "Id":"PutObjectPolicy",
    "Statement":[{
        "Sid":"DenyObjectsThatAreNotSSEKMS",
        "Effect":"Deny",
        "Principal":"*",
        "Action":"s3:PutObject",
        "Resource":"arn:aws:s3:::amzn-s3-demo-bucket1/*",
        "Condition":{
            "Null":{
                "s3:x-amz-server-side-encryption-aws-kms-key-id":"true"
            }
        }
    }
    ]
}
```

You can specify the KMS key using the **x-amz-server-side-encryption-aws-kms-key-id** header or rely on your default bucket encryption configuration. If your **PutObject** request specifies aws:kms in the **x-amz-server-side-encryption** header but does not specify the **x-amz-server-side-encryption-aws-kms-key-id** header, then Amazon S3 assumes that you want to use the AWS managed key.

Uploads denied unless request includes the x-amz-server-side-encryption-aws-kms-key-id

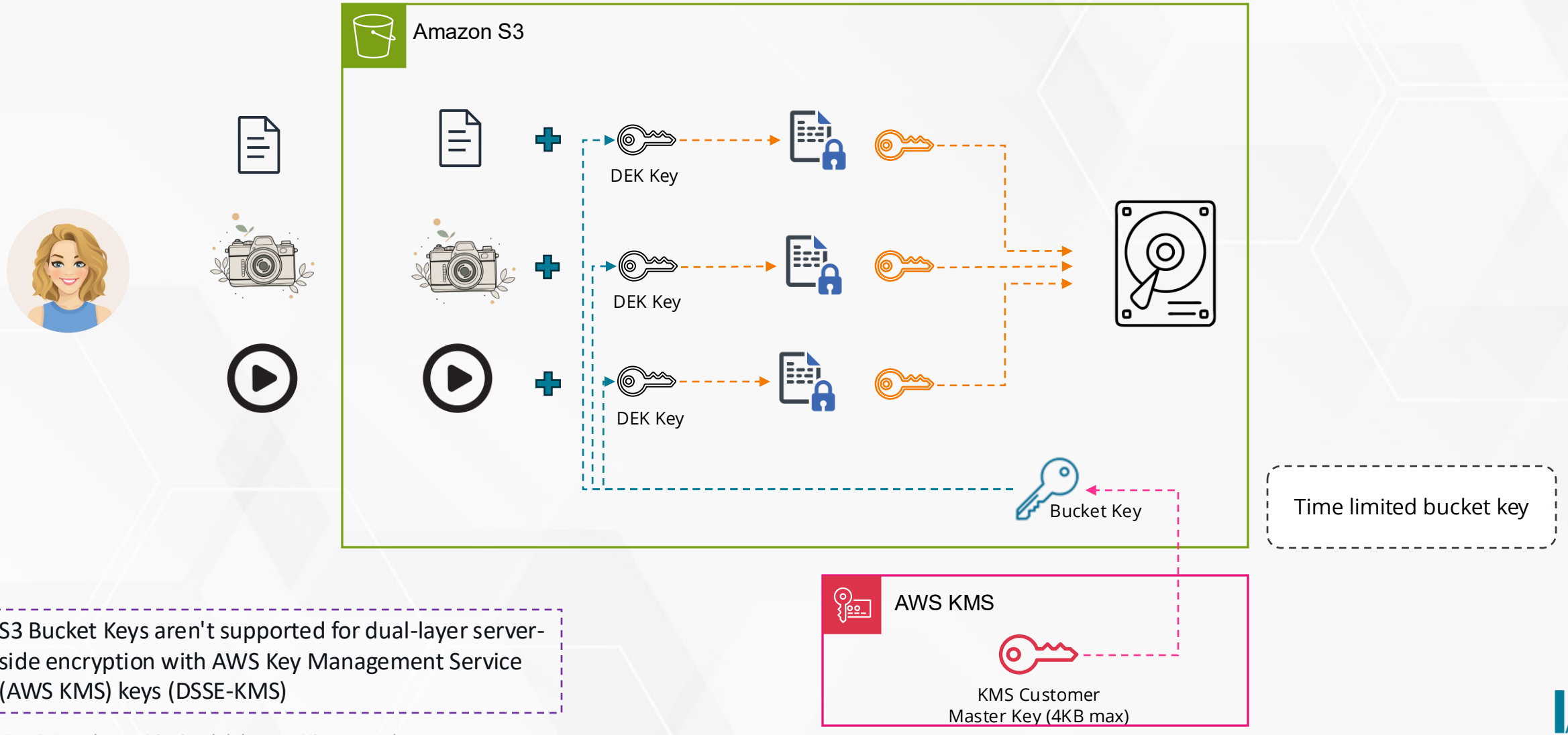
Amazon S3 KMS Requests



KMS Throttling

You have exceeded the rate at which you may call KMS. Reduce the frequency of your calls.
(Service: AWSKMS; Status Code: 400; Error Code: ThrottlingException; Request ID: <ID>

Amazon S3 Bucket Keys



S3 Bucket Keys aren't supported for dual-layer server-side encryption with AWS Key Management Service (AWS KMS) keys (DSSE-KMS)

Amazon S3 Encryption

Testing Encryption Options for Amazon S3

IAAS ACADEMY

Amazon S3 Presigned URLs

Sharing Objects Securely

IAAS ACADEMY

Customer Scenario

Fulfilling the business requirements

"secure, scalable and resilient, ensure that the eBook is only available to genuine sign-ups ."

Gemma Customer

Subscribe for a **FREE** Recipe eBook

Name

Email

Submit

Email

*https://gutfreeweb.s3.amazonaws.com/******.pdf*

```
{
  "Version": "2012-10-17",
  "Id": "Policy1705424029637",
  "Statement": [
    {
      "Sid": "Stmt1705424028548",
      "Effect": "Allow",
      "Principal": "*",
      "Action": "s3:GetObject",
      "Resource": [
        "arn:aws:s3:::[redacted]
        "arn:aws:s3:::[redacted]/*"
      ]
    }
  ]
}
```

Introducing Amazon S3 Pre-Signed URLs

"secure, scalable and resilient, ensure that the eBook is only available to genuine sign-ups ."

Gemma Customer

Subscribe for a **FREE** Recipe eBook

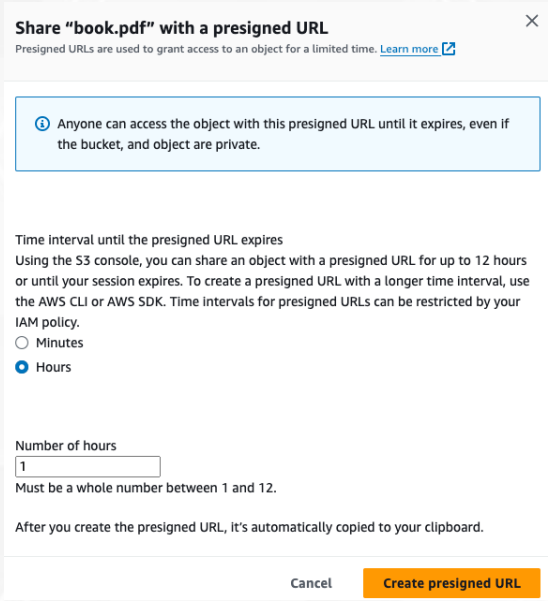
Name

Email

Submit

Generate a Pre-Signed URL Expiry 1 Hour

Email

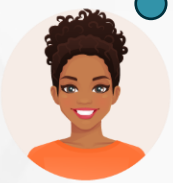


https://gutfree.s3.us-east-1.amazonaws.com/books/book.pdf?response-content-disposition=inline&X-Amz-Security-**Token**=IQoJb3JpZ2luX2VjEGMaCWV1LXdlc3QtMiJGMEQCIG5%2BalpeghxZHd%2BnjbmYQk%2FIdDN%%2FQN69gXQ6HvK8vE%2FRJzWCYOlh%2FxIN9WcAwlxjcilT4jP05BgCpbUAh5RjOSMn%2FNKsOYsYX%2FWNhmidMsglPTkK5AdTC1Gk5fdEl3H&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Date=20240120T191219Z&X-Amz-SignedHeaders=host&**X-Amz-Expires=3600**&**X-Amz-Credential**=ASIA5XZWIW7TAEYOFEEN%2F20240120%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Signature=767af7db23ed4dbf5980d3e259bbecd61bd1b6dd4f7c0081a8ebdd0cff6e02e3

Presigned URLs

- When you create a presigned URL, you must provide your security credentials, and then specify the following:
- An Amazon S3 bucket
- An object key (if downloading this object will be in your Amazon S3 bucket, if uploading this is the file name to be uploaded)
- An HTTP method (GET for downloading objects or PUT for uploading)
- An expiration time interval

If you created a presigned URL using a temporary credential (such when using an IAM Role, the URL expires when the credential expires. In general, a presigned URL expires when the credential you used to create it is revoked, deleted, or deactivated. This is true even if the URL was created with a later expiration time



Amazon S3 Sever Access Logs

Logging Amazon S3 Activity

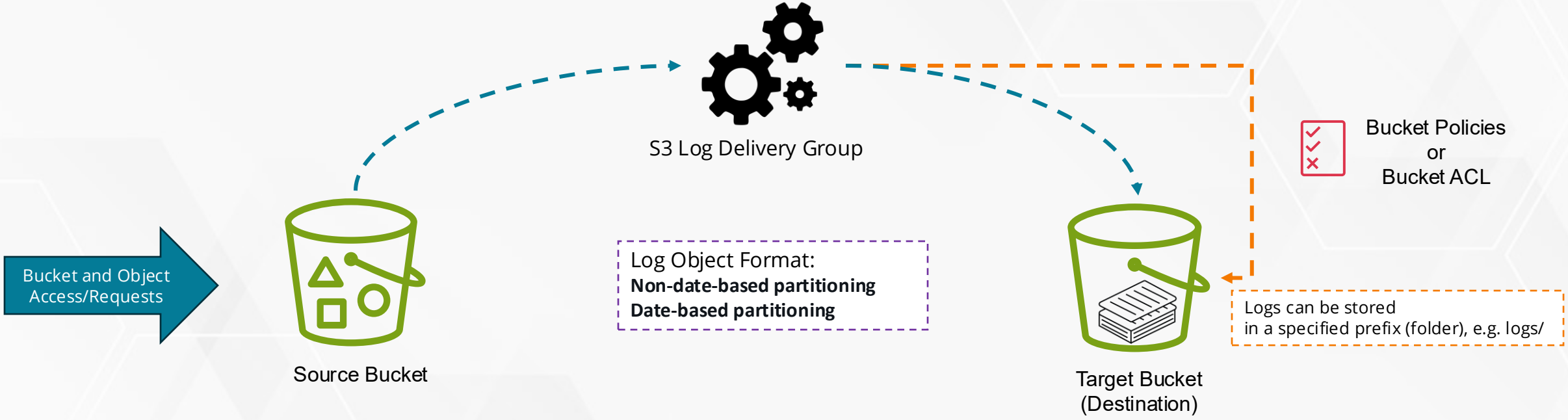
IAS ACADEMY

Amazon S3 Server Access Logging

Both the source and destination buckets must be in the same AWS Region and owned by the same account.

Be careful of infinite loop issue

Best efforts basis can take hours!



Amazon S3 Object Lock

Fulfilling Compliance and Governance Requirements

IAAS ACADEMY

Overview of S3 Object Lock

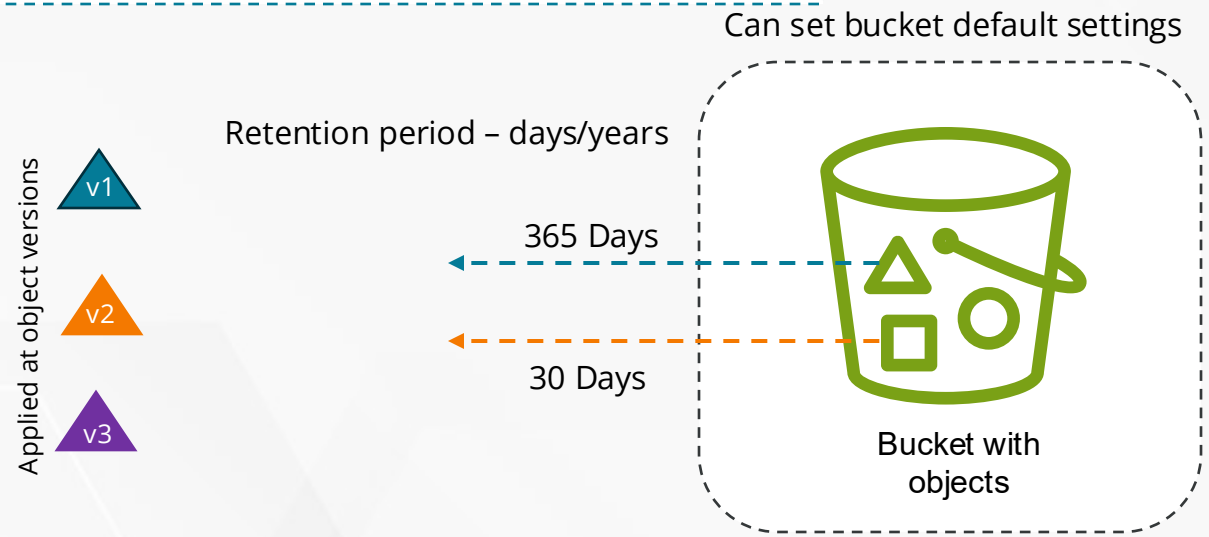
- S3 Object Lock can help prevent Amazon S3 objects from being deleted or overwritten for a fixed amount of time or indefinitely.
- Object Lock uses a write-once-read-many (WORM) mode
- Object Locks requires versioning to be enabled
- Two Options
  - S3 Object Lock **Retention period**
  - S3 Object Lock **Legal hold**

S3 stores lock information in metadata on the object version when locking that specific object version. The lock protects only the version that's specified in the request.

Object Lock Retention period

The only way to delete an object under the compliance mode before its retention date expires is to delete the associated AWS account.

Important Note: When you PUT an object version that has an explicit individual retention mode and period in a bucket, the object version's individual Object Lock settings override any bucket property retention settings



Restrict the minimum and maximum allowable retention periods with the s3:object-lock-remaining-retention-days condition key in the bucket policy

Compliance Mode

Object version can't be overwritten or deleted
Retention mode can't be changed
Retention period can't be shortened

```
aws s3api put-object-lock-configuration --bucket amzn-s3-demo-bucket1 --object-lock-configuration='{ "ObjectLockEnabled": "Enabled", "Rule": { "DefaultRetention": { "Mode": "COMPLIANCE", "Days": 50 }}}'
```

Governance Mode

Cannot overwrite or delete an object version
Cannot alter its lock settings
UNLESS identity has special permissions

s3:BypassGovernanceRetention

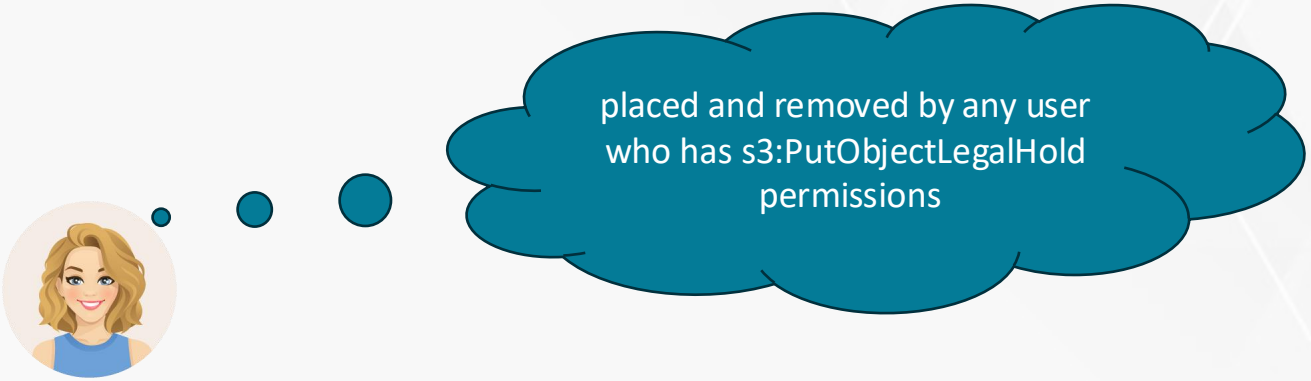
Header: x-amz-bypass-governance-retention:true

Object Lock – Legal holds

Place a *legal hold* on an object version

Legal holds are independent from retention periods

Object version is locked until legal hold is removed



Amazon S3 Select & Glacier Select

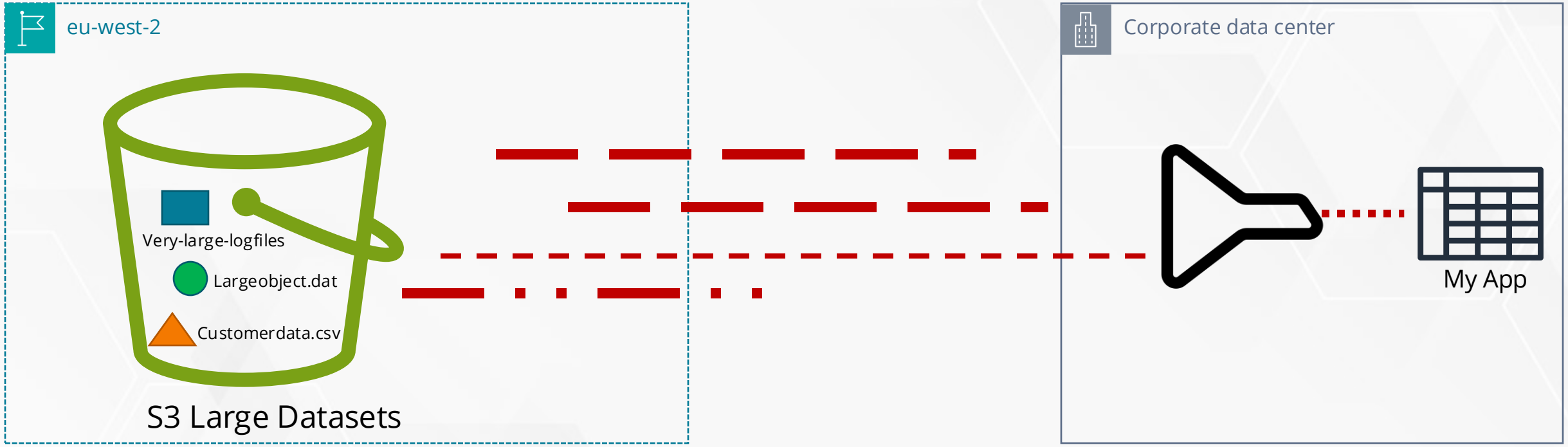
Filtering Data Retrievals

IAAS ACADEMY

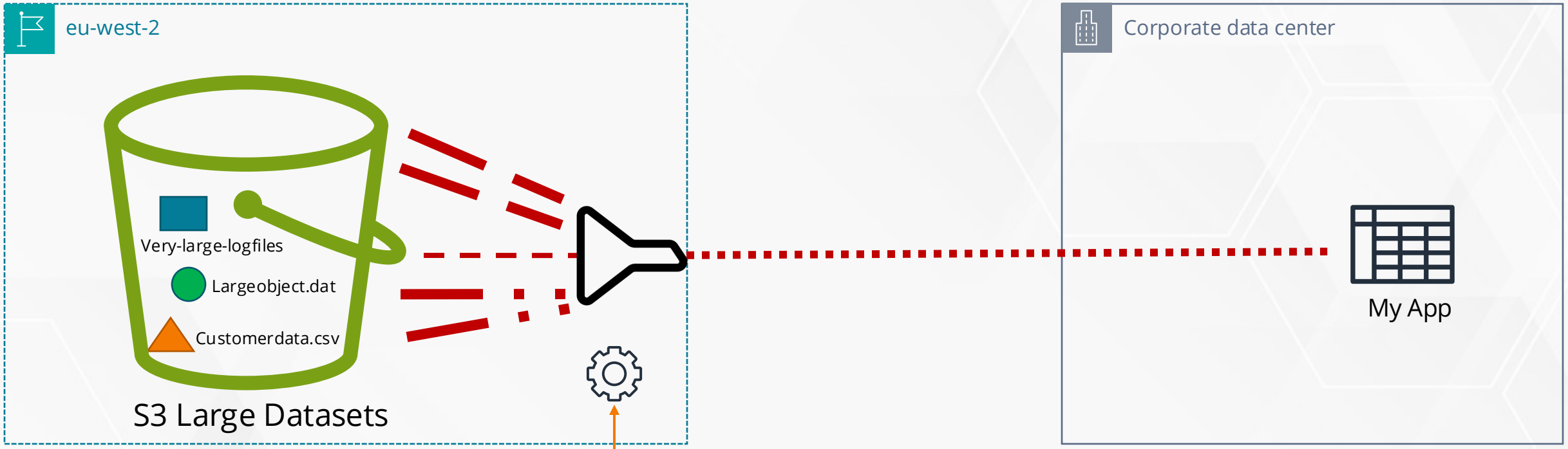
Data Retrieval from Amazon S3

- Maximum Object Size is 5TB
- You typically retrieve objects in their entirety – e.g., 1GB CSV Customer Order across all stores
  - You consume 1GB of Data in data transfer
  - Time Delay in retrieval
  - Your application may filter data at client side, e.g., you only want to know the total breakdown of customers across five stores in Birmingham.
- S3 and Glacier Select allows you to filter data in the S3 environment before download using SQL-Like Statements, and only retrieve a subset of the required data

Without Amazon S3 Select



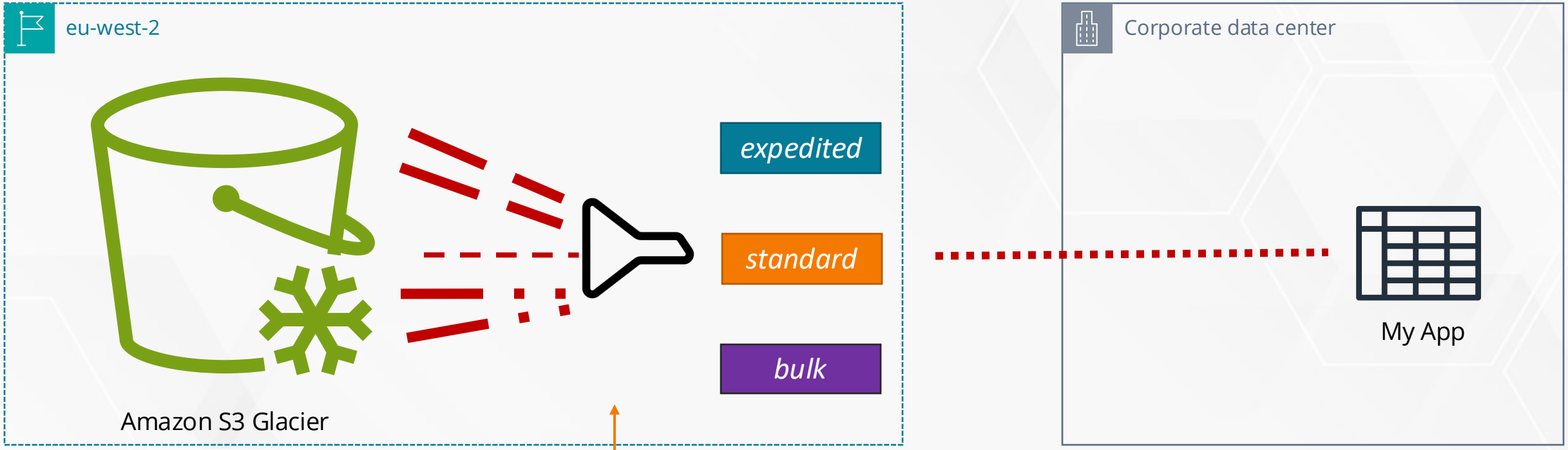
With Amazon S3 Select



Simple SQL expressions to pull out only the bytes you need from those objects

Only required dataset is retrieved by the App
- Cost optimized
- Faster throughput

With Amazon S3 Glacier Select



S3 Glacier Select runs the query in place and writes the output results to Amazon S3.

Only required dataset is retrieved by the App
- Cost optimized
- Faster throughput