
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

STG323

Using S3 Batch Operations to manage S3 buckets at scale

Neha Tyagi (she/her)

Software Development Manager
Amazon S3

Rohan Thomas (he/him)

Senior Product Manager
Amazon S3

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

- Operating at scale
- How S3 Batch Operations can help manage your data
- S3 Batch Operations functionality
- Customer use cases
- Best practices
- Demo
- Recap

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Operating at scale

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Operating at scale in Amazon S3

Industry-leading scalability, availability, and durability

Wide range of cost-optimization capabilities



Easily manage data and access controls

Unmatched security, compliance, and audit capabilities

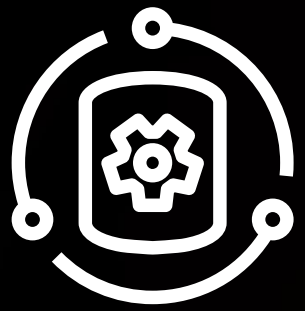
Broadest data movement and hybrid cloud storage options

Industry-leading performance

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

How do you manage your data at scale?



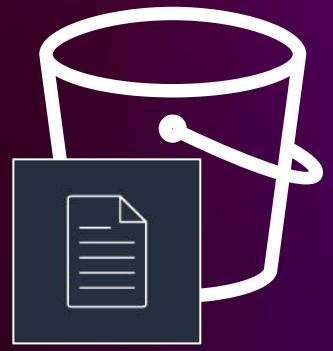
S3 Lifecycle policies



S3 Intelligent-Tiering



S3 Storage Lens



S3 Inventory reports



S3 Batch Operations

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

How Amazon S3 Batch Operations can help manage your data

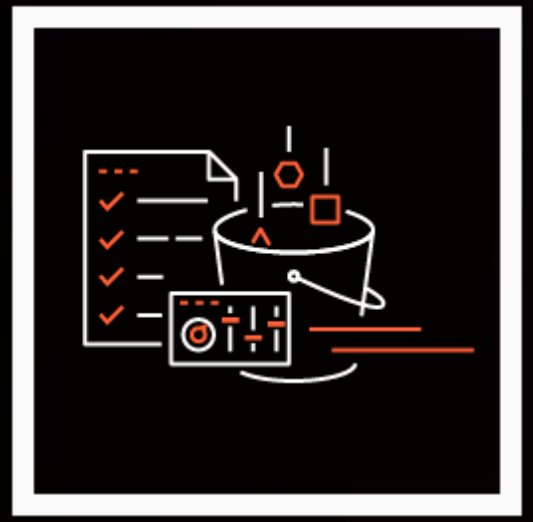
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

S3 Batch Operations use cases

- Copy data to another bucket/regions
- Batch Replication
- Move data between storage classes using the copy API
- Change object access control lists (ACLs)
- Re-encrypt data with the copy API
- Use a Lambda function to create thumbnails for all the images in your bucket
- Bulk add or remove tags to existing objects
- Tag Deletion
- Locking and setting retention period of objects

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

S3 Batch Operations is a managed solution



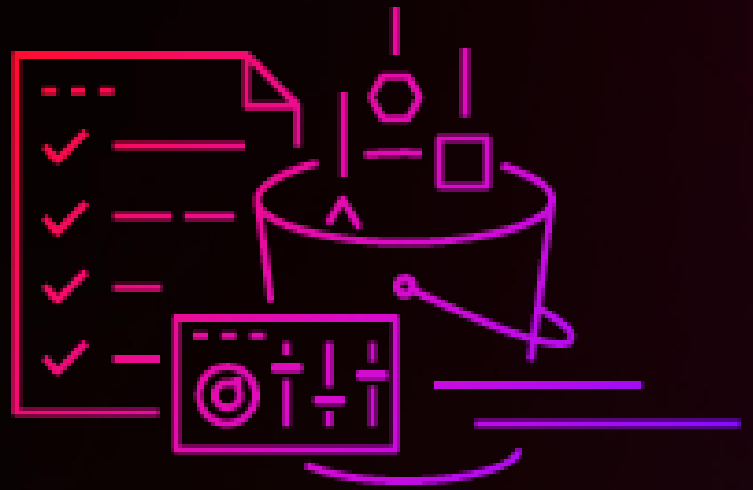
- Guaranteed execution
- Automatic retries
- Progress visibility
- Management controls
- Notifications
- Auditing

No need to build and maintain an application to call APIs in bulk

aws

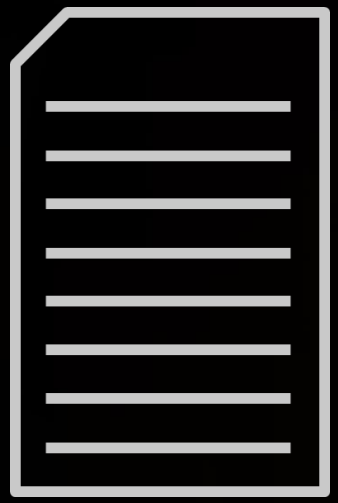
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Save on operations and development time by managing billions of objects with a single request

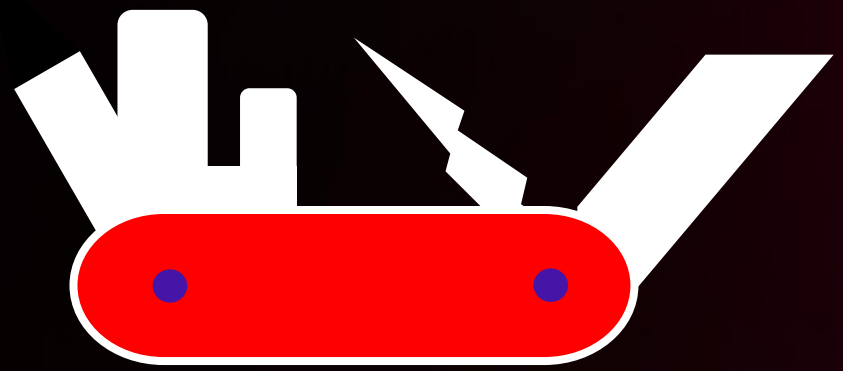


© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

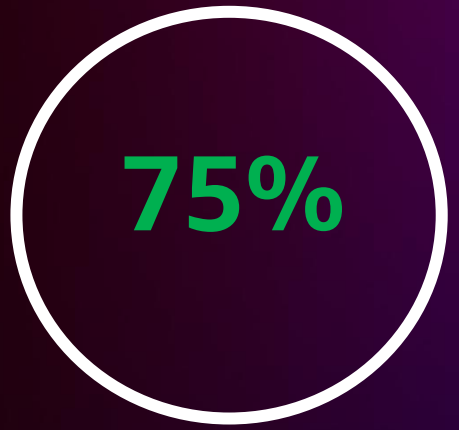
S3 Batch Operations: How it works



Choose objects



Select an operation



View progress

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

S3 Batch Operations

Choose objects

- S3 Inventory Report
- CSV
- Create manifest using S3 Replication configuration.

Select an operation

- Copy
- Restore from S3 Glacier
- Put Access Control List (ACL)
- Replace All Object Tags
- Delete All Object Tags
- Run Lambda Functions
- Object Lock Retention
- Object Lock Legal Hold
- Replicate

View progress

- Object Level Progress
- Time in Active state
- Job Notifications
- Completion Report

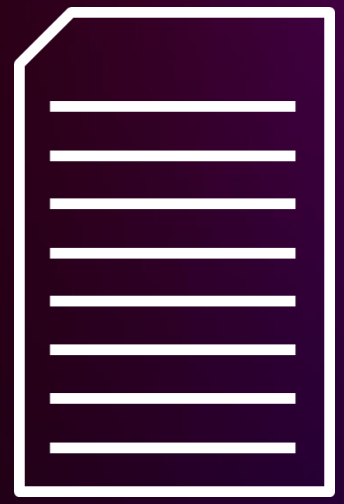
aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon S3 Batch Operations functionality

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Manifest creation

- Your manifest lists the target objects for your job
- Manifest formats
  - S3 Inventory report
  - CSV file
  - Create manifest using S3 Replication configuration
- Input fields
  - Bucket
  - Key name
  - Version ID (optional)



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Setting up permissions

S3 Batch Operations assumes an Identity and Access Management (IAM) role to perform the job

S3 Batch Operations requires permissions to:

- Reading your manifest of objects
- Performing the action in the source and destination (if required) bucket
- Writing the optional completion report (recommended)

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Cross-account considerations

To copy objects to another AWS account:

- Write the S3 Inventory report to the destination account
- Have S3 Batch Operations assume an IAM role in the destination account to copy the objects

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Destination account permissions

```
"Version": "2022-10-10",
"Statement": [{
      "Sid": "AllowBatchOperationsDestinationObjectCOPY",
      "Effect": "Allow",
      "Action": [
         "s3:PutObject",
         "s3:PutObjectVersionAcl",
         "s3:PutObjectAcl",
         "s3:PutObjectVersionTagging",
         "s3:PutObjectTagging",
         "s3:GetObject",
         "s3:GetObjectVersion",
         "s3:GetObjectAcl",
         "s3:GetObjectTagging",
         "s3:GetObjectVersionAcl",
         "s3:GetObjectVersionTagging”],
      "Resource": [
         "arn:aws:s3:::ObjectDestinationBucket/*",
         "arn:aws:s3:::ObjectSourceBucket/*",
         "arn:aws:s3:::ObjectDestinationInventoryReportBucket/*"
      ]
   }
]
```

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Source account permissions

```
{
   "Version": "2022-10-10",
   "Statement": [
      {
         "Sid": "AllowBatchOperationsSourceObjectCOPY",
         "Effect": "Allow",
         "Principal": {
        "AWS": "arn:aws:iam::DestinationAccountNumber:role/BatchOperationsDestinationRoleCOPY"
         },
         "Action": [
            "s3:GetObject",
            "s3:GetObjectVersion",
            "s3:GetObjectAcl",
            "s3:GetObjectTagging",
            "s3:GetObjectVersionAcl",
            "s3:GetObjectVersionTagging"
         ],
         "Resource": "arn:aws:s3:::ObjectSourceBucket/*"
      }
   ]
}
```

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Synchronize buckets with S3 Batch Replication



Migrate data

Create a disaster recovery plan by replicating all data in the bucket, including data that was uploaded before replication was set up

Set up a new region

Hydrate a new S3 bucket with data from the primary or secondary region

Re-run previously failed objects

Re-run replication on a set of objects after fixing user issues

Visibility

- Active job running time
- Completion report
- Optional manifest report
- CloudTrail events

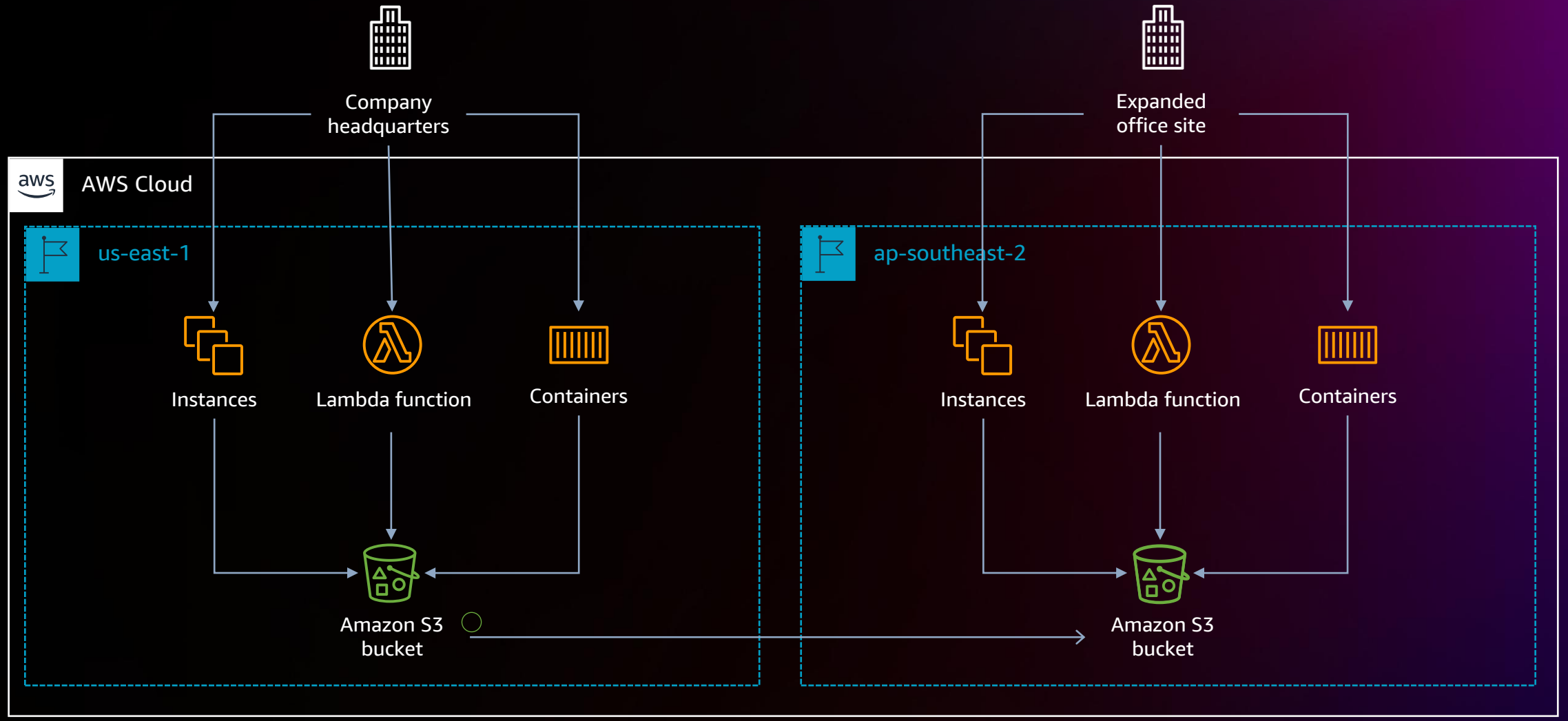
aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Customer use cases

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Customer use case - 1

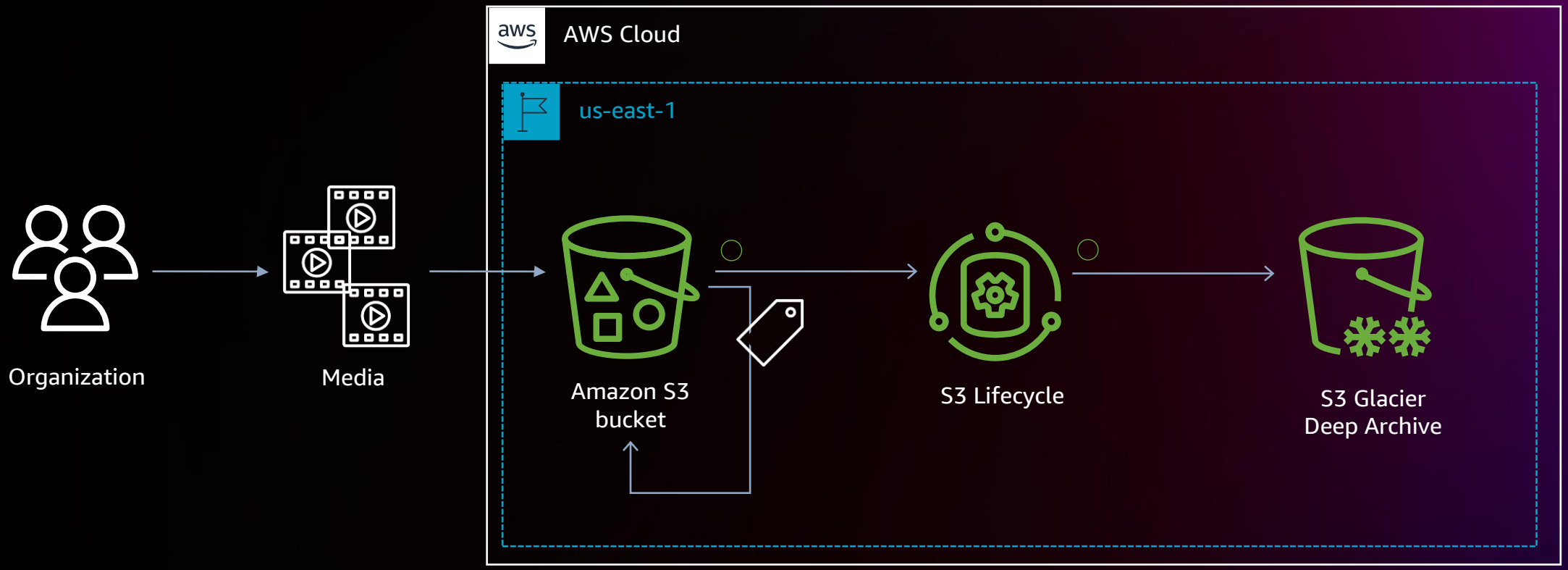


Question: How can you make the existing data available in the new office site?

S3 Batch Replication!

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Customer use case - 2

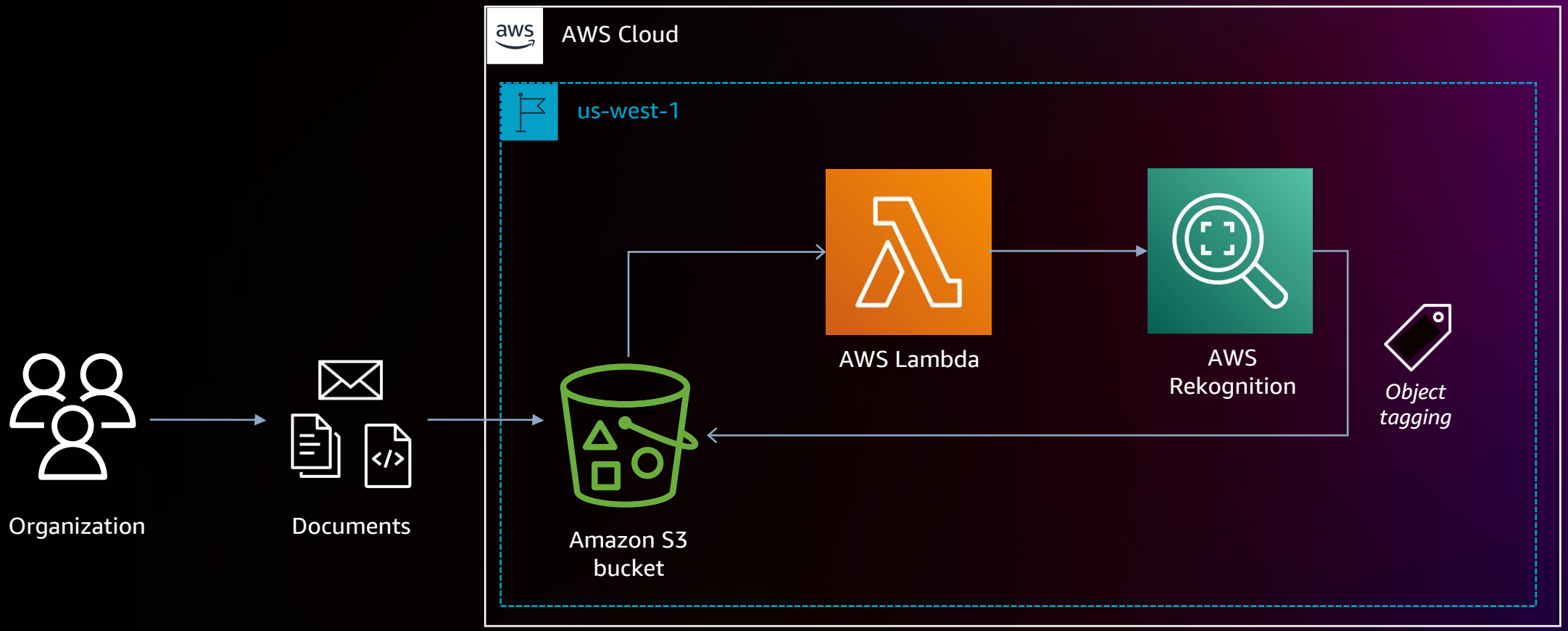


S3 Batch Operations to Tag Objects
Set an S3 Lifecycle Rule based on Object Tag
S3 Lifecycle transitions Objects to S3 Glacier Deep Archive

Question: How can they catalogue the media files and transition them to more cost optimized storage classes?

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Customer use case - 3



Use S3 Batch Operations to call a Lambda function
Use the Lambda function to invoke AWS Rekognition
Use Object Tags to identify objects with PII

Question: How can you daily scan for Personal Identifiable Information (PII)?

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Best practices

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Best practices for setting up and running S3 Batch Operations

- Break up large Batch Operation jobs into a few smaller ones for highest performance
- Always ask for and customize the completion report
- Always download and review the manifest of large scale Batch replication jobs
- Use reserved concurrency or request a quota increase to control Lambda executions

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Recap

- Features to operate at scale
- Amazon S3 Batch Operations dive deep
- S3 Batch Operations functionality
- Use cases and best practices
- Demo – S3 Batch Operations copy using Lambda
- Demo – S3 Batch Replication

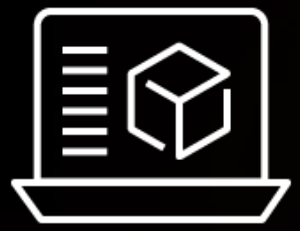
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Q&A

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Continue your AWS Storage learning

Build a learning plan



Set your AWS Storage Learning Plans using **AWS Skill Builder**

Increase your knowledge



Use our **Ramp-Up Guides** to build your storage knowledge

Earn AWS Storage badges



Demonstrate your knowledge by achieving **digital badges**

aws.training/storage

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Question: Optimizing performance

Let’s say we have 4 billion objects in our S3 bucket that we want to add tags to using S3 Batch Operations …

Question: How do we optimize performance in this scenarios?

Our recommendation:

- Take advantage of parallel processing
- Split 4 billion tasks into 4 jobs, each containing 1 billion tasks
- Run the 4 jobs in parallel

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Question: Restoring archived objects

Let’s say we have 400 TB in asynchronous S3 storage classes …

- 100 TB in Glacier Flexible Retrieval
- 100 TB in Glacier Deep Archive
- 100 TB in Intelligent Tiering Archive Access
- 100 TB in Intelligent Tiering Deep Archive Access

Question: How do we restore these archived objects back to S3 Standard?

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Question: Restoring archived objects

Question: How do we restore these archived objects back to S3 Standard?

Our recommendation:

- Make a manifest for *Glacier Flexible Retrieval* and *Glacier Deep Archive* objects
- Make a manifest for *Intelligent Tiering Archive Access* and *Intelligent Tiering Deep Archive Access* objects
- Create separate S3 Batch Operations jobs to restore objects to S3 Standard

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Question: Copying objects

Let’s say we have a bucket with objects ranging from 100 KB to 100 GB that need to be copied to another bucket …

Question: How do we copy all of these objects using S3 Batch Operations?

Our recommendation:

- Write a S3 Lambda function that copies objects between buckets
- Run a S3 Batch Operations job calling the Lambda function

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.