

Amazon S3 Batch Operations

Rob Wilson, Product Manager, Amazon S3

May 23, 2019

Agenda

Operating at scale in Amazon S3

Overview of Amazon S3 Batch Operations

Demo

Deep Dive on Amazon S3 Batch Operations

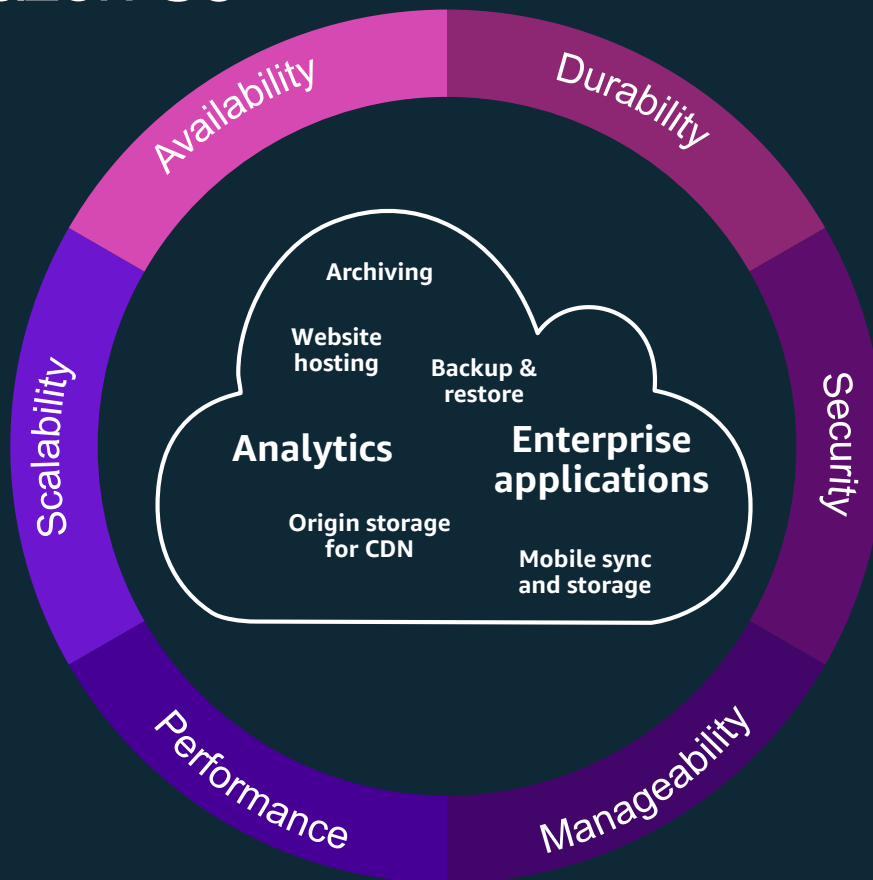
Common Use Cases

Q&A

Operating at scale



Benefits of Amazon S3



Features to help you operate at scale

- Lifecycle management
- S3 Intelligent-Tiering storage class
- Inventory reports
- S3 Batch Operations

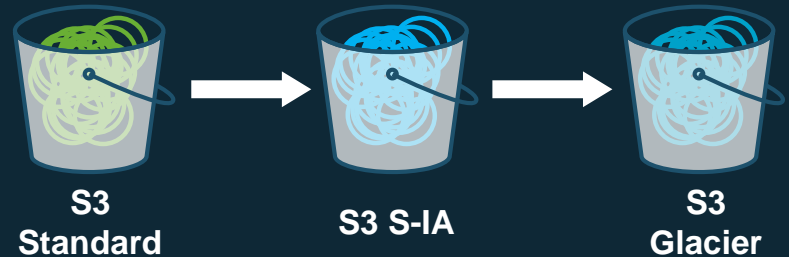


Set S3 Lifecycle Policies to Tier and Expire storage

S3 lifecycle policies tier objects to lower cost storage classes and expire storage

S3 Storage Class Analysis results help you set lifecycle policies

Policies are based on the object's age and can be set by bucket, prefix, or object tag



Your choice of Amazon S3 storage classes



S3 Standard



S3 Intelligent-Tiering



S3 Standard-IA



S3 One Zone-IA



S3 Glacier



S3 Glacier Deep Archive



Frequent

Access frequency

Infrequent

- Active, frequently accessed data
- Milliseconds access
- ≥ 3 AZ
- \$0.0210/GB

- Data with changing access patterns
- Milliseconds access
- ≥ 3 AZ
- \$0.0210 to \$0.0125/GB
- Monitoring fee per Obj.
- Min storage duration

- Infrequently accessed data
- Milliseconds access
- ≥ 3 AZ
- \$0.0125/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

- Re-creatable, less accessed data
- Milliseconds access
- 1 AZ
- \$0.0100/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

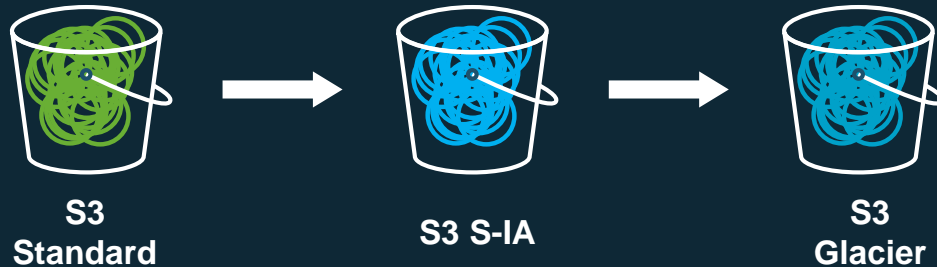
- Archive data
- Select minutes or hours
- ≥ 3 AZ
- \$0.0040/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

- Archive data
- Select hours
- ≥ 3 AZ
- \$0.00099/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

Lifecycle Management Example Policies

Lifecycle rules take action based on object age:

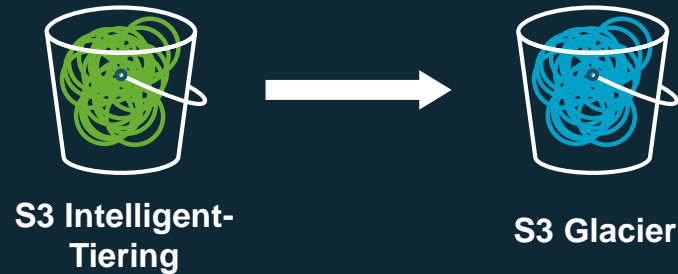
1. Move all objects older than 60 days to S3 S-IA
2. Move all objects older than 180 days to S3 Glacier



Lifecycle Management Example Policies

Lifecycle rules take action based on object age:

1. Move all objects older than 180 days to S3 Glacier



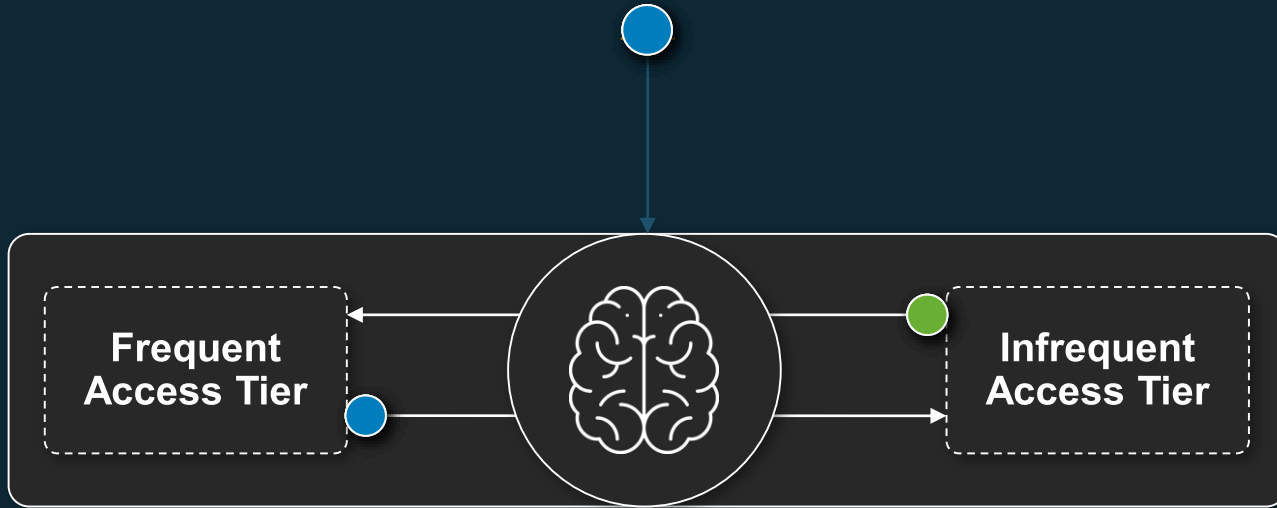
Lifecycle Management Example Policies

Lifecycle rules take action based on object age:

1. Move all objects older than 180 days to S3 Glacier
2. Move all objects older than 365 days to S3 Glacier Deep Archive



S3 Intelligent-Tiering automates cost savings



S3 Intelligent-Tiering storage class



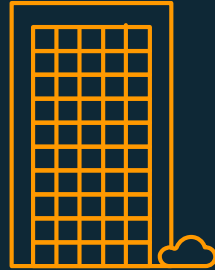
- Automatically optimizes storage costs for data with changing access patterns
- Monitors access patterns and auto-tiers on granular object level
- No performance impact, no operational overhead
- Milliseconds access, ≥ 3 AZs
- Monitoring fee per object, minimum storage duration

Ideal use cases for S3 Intelligent-Tiering



Big Data, Data Lakes

Storage with changing access patterns used by multiple applications



Enterprises

Storage accessed by fragmented applications from various organizations



Startups

Constraint on resources and experience to optimize storage themselves

Dynamic cost optimization with no performance impact and no operational overhead

Amazon S3 Inventory

Regularly generates a list of objects for analytics and auditing.



- ✓ Encryption status
- ✓ Replication status
- ✓ Retention date
- ✓ Size, and more

Overview of Amazon S3 Batch Operations



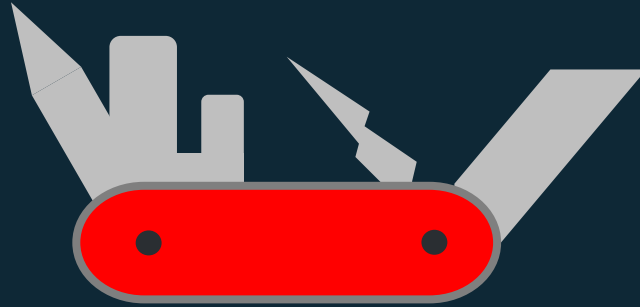
Save on operations and development time by
managing billions of objects with a single request



S3 Batch Operations: How it works



Choose
Objects



Select an
Operation



View
Progress

S3 Batch Operations

Choose
Objects

- S3 Inventory Report
- CSV List

S3 Batch Operations

Choose
Objects

Select an
Operation

- S3 Inventory Report
- CSV List
- Copy
- Restore from S3 Glacier
- Put Access Control List (ACL)
- Replace Object Tag Sets
- Run Lambda Functions

S3 Batch Operations

Choose
Objects

Select an
Operation

View
Progress

- S3 Inventory Report
- CSV List
- Copy
- Restore from S3 Glacier
- Put Access Control List (ACL)
- Replace Object Tag Sets
- Run Lambda Functions
- Object Level Progress
- Job Notifications
- Completion Report

Amazon S3 Management Console list view

S3 batch operations

A job is used to execute batch operations on a list of S3 objects. The list of S3 objects is contained in a manifest object, which can be an S3 inventory report or a list of objects that you generate. After the total number of objects listed in the manifest has been confirmed, the job status will update to *Awaiting confirmation*, and you must **Confirm and run** the job within 14 days. Job events are published to [CloudWatch Events](#). Jobs are deleted 90 days after they finish. [Learn more](#)

Status: All

Region: US West (Oregon)

+ Create job

Clone job

Update priority

Confirm and run

Cancel job



Viewing 1 to 78 of 78 jobs

Job ID	Description	Operation	Date created	Status	Total objects	% Complete	Total failed (rate)	Priority
fcaaea6d-8ead-4225-b7bf-88d4c87c753c	2019-02-20 Set-Tags ProjectDelta	Replace all tags	Feb 20, 2019 3:06:49 PM GMT-0800	Complete	2,717,669	100%	0 (0%)	10
d8f2971a-4b5c-44ec-847e-23af52f840f9	2019-02-19 Set-Tags	Replace all tags	Feb 20, 2019 2:13:55 PM GMT-0800	Complete	2,700,671	100%	0 (0%)	10
5318b940-eafd-4d1b-946a-b611e4e124e4	2019-02-19 Set-Tags	Replace all tags	Feb 20, 2019 2:13:38 PM GMT-0800	Complete	2,700,671	100%	0 (0%)	10
ed4f5617-b7e8-4576-a378-f31d23588c82	2019-02-19 Set-Tags	Replace all tags	Feb 20, 2019 2:13:20 PM GMT-0800	Complete	2,700,671	100%	0 (0%)	10

S3 Batch Operations is a managed solution



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

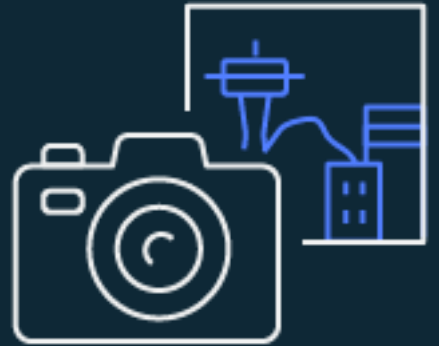
No need to build and maintain an application to call APIs in bulk

Demo



Demo Overview

- Images stored in an Amazon S3 bucket
- S3 Batch Operations invokes an AWS Lambda function that calls Amazon Rekognition
- Results stored in Amazon Elasticsearch
- A Kibana dashboard shows the output



Deep dive on S3 Batch Operations functionality



Manifest creation

Your manifest lists the target objects for your job.

Manifest formats:

- S3 Inventory report
- CSV file

Input fields:

- Bucket
- Key name
- Version ID (optional)



How does S3 Batch Operations execute the job?

S3 Batch Operations calls existing APIs.

Offers similar functionality as calling the API directly

- Includes AWS CloudTrail support
- Allows you to customize the API's parameters
- Requires the same permissions as the existing API
- Makes it simple to use S3 Batch Operations with your existing applications

Setting up permissions

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

S3 Batch Operations requires permissions for:

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

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

Trust Policy needed for S3 Batch Operations

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "batchoperations.s3.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

S3 Batch Operations can be customized

- Specify whether or not a confirmation step is required
- Set different priorities for different types of jobs
- Specify information to include in your completion report

Using S3 Batch Operations priorities

<u>Job Description</u>	<u>Job State</u>	<u>Job Priority</u>
Tag All Objects	Active	10
Copy to Oregon	Active	10
Restore Prefix A	Active	10
Change ACLs	Active	10
Lambda Demo	Ready	10

Using S3 Batch Operations priorities

<u>Job Description</u>	<u>Job State</u>	<u>Job Priority</u>
Tag All Objects	Active	10
Copy to Oregon	Active	10
Restore Prefix A	Active	10
Change ACLs	Active	10
Lambda Demo	Ready	10

Using S3 Batch Operations priorities

<u>Job Description</u>	<u>Job State</u>	<u>Job Priority</u>
Tag All Objects	Active	10
Copy to Oregon	Active	10
Restore Prefix A	Active	10
Change ACLs	Active	10
Lambda Demo	Ready	15

Using S3 Batch Operations priorities

<u>Job Description</u>	<u>Job State</u>	<u>Job Priority</u>
Tag All Objects	Active	10
Copy to Oregon	Active	10
Restore Prefix A	Active	10
Change ACLs	Pausing	10
Lambda Demo	Active	15

Using S3 Batch Operations priorities

<u>Job Description</u>	<u>Job State</u>	<u>Job Priority</u>
Tag All Objects	Active	10
Copy to Oregon	Active	10
Restore Prefix A	Active	10
Change ACLs	Paused	10
Lambda Demo	Active	15

Regional list view

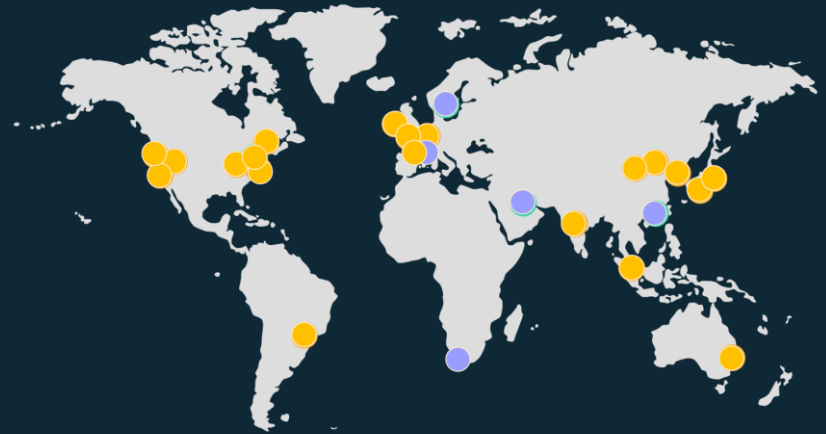
S3 Batch Operations shows job activity at the account level.

ListJobs API shows:

- All jobs in the region
- Job ID and description
- Operation type
- Creation date
- Job state and job priority
- Job size and progress

Choose region for jobs based on:

- Object location for most operations
- Destination region for all copy jobs



Job progress and notifications

Amazon CloudWatch Events inform you when your job changes state.

Detailed job progress at the object level is visible through:

- DescribeJob API
- ListJobs API (view up to 1,000 jobs at a time filtered by job states)
- Amazon S3 Management Console

Completion reports

A record of all work performed by S3 Batch Operations

Lists:

- Bucket, Key name, Version ID
- Success/Failure
- Status code
- Code description
- Additional detail or result set

S3 Batch Operations Pricing

Jobs	\$0.25 per job
Objects	\$1.00 per million object operations performed

You are charged for S3 Batch Operations jobs, objects, and requests in addition to any charges associated with the operation that S3 Batch Operations performs on your behalf, including data transfer, requests, and other charges.


Common use cases



Encrypt your existing objects

Copy objects to the same bucket and specify the desired type of encryption

Server-side encryption

Choose the encryption type for all objects copied to the destination. [Learn more](#) 

Server-side encryption configuration

None
If this bucket is set up for default encryption, then S3 will encrypt objects as per the bucket settings.

Amazon S3-managed encryption keys (SSE-S3)

Amazon KMS master-key (SSE-KMS)

Copy objects to a new location


Works within regions, across regions, and across accounts

PUT copy

Your PUT copy destination must be in the US West (Oregon) Region, where this job is being created.

PUT copy destination bucket

[Browse](#)[View](#)[Create bucket](#)

Format: s3://mybucket. [Learn more](#) 

Copy objects to a new storage class

Can be used as an alternative to lifecycle policies

Storage class


Choose a storage class based on your use case and access requirements. [Learn more](#) or see [Amazon S3 pricing](#)

Storage class	Designed for	Availability Zones	Min storage duration	Min billable object size	Monitoring and automation fees	Retrieval fees
<input checked="" type="radio"/> Standard	Frequently accessed data	≥ 3	-	-	-	-
<input type="radio"/> Intelligent-Tiering	Long-lived data with changing or unknown access patterns	≥ 3	30 days	-	Per-object fees apply	-
<input type="radio"/> Standard-IA	Long-lived, infrequently accessed data	≥ 3	30 days	128KB	-	Per-GB fees apply
<input type="radio"/> One Zone-IA	Long-lived, infrequently accessed, non-critical data	≥ 1	30 days	128KB	-	Per-GB fees apply
<input type="radio"/> Glacier	Archive data with retrieval times ranging from minutes to hours	≥ 3	90 days	-	-	Per-GB fees apply

Restore objects from S3 Glacier

Can restore and copy using the same manifest

Restore

Retrieval fees apply. See [S3 pricing](#) 

Number of days that the restored copy is available

The restored copy in Reduced Redundancy Storage (RRS) class is automatically deleted after a specified number of days.

days

Restore tier

- Bulk retrieval
Typically within 5-12 hours for Glacier and within 48 hours for Glacier Deep Archive.
- Standard retrieval
Typically within 3-5 hours for Glacier and within 12 hours for Glacier Deep Archive.

Pricing for an S3 Glacier restore job

<u>Usage</u>	<u>Cost</u>
1 Job	\$0.25 (\$0.25 / job)
1,000,000 objects	\$1.00 (\$1.00 / 1M objects)
1,000,000 restore requests	\$25.00 (\$0.025 / 1k requests)
977 GB restored (bulk tier)	\$2.44 (\$0.0025 / GB)
	Total: \$28.69

Simplify recurring and one-time workloads

Recurring

- Apply new object tags to drive lifecycle policies
- Perform bulk S3 Glacier restores to run periodic reports
- Standardize data formats using AWS Lambda functions
- Copy data as part of a workflow or for backups

One-time

- Change object access control lists (ACLs)
- Move data between storage classes using the copy API
- Encrypt data with the copy API
- Use a Lambda function to create thumbnails for all the images in your bucket
- Copy data to another bucket
- Bulk tag existing objects

Q&A

Thank you!

Rob Wilson, Product Manager, Amazon S3