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

- Enterprise applications
- Analytics
- Archiving
- Website hosting
- Backup & restore
- Origin storage for CDN
- Mobile sync and storage
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**
  - Active, frequently accessed data
  - Milliseconds access
  - > 3 AZ
  - $0.0210/GB

- **S3 Intelligent-Tiering**
  - Data with changing access patterns
  - Milliseconds access
  - > 3 AZ
  - $0.0210 to $0.0125/GB
  - Monitoring fee per Obj.
  - Min storage duration

- **S3 Standard-IA**
  - Infrequently accessed data
  - Milliseconds access
  - > 3 AZ
  - $0.0125/GB
  - Retrieval fee per GB
  - Min storage duration
  - Min object size

- **S3 One Zone-IA**
  - Re-creatable, less accessed data
  - Milliseconds access
  - 1 AZ
  - $0.0100/GB
  - Retrieval fee per GB
  - Min storage duration
  - Min object size

- **S3 Glacier**
  - Archive data
  - Select minutes or hours
  - > 3 AZ
  - $0.0040/GB
  - Retrieval fee per GB
  - Min storage duration
  - Min object size

- **S3 Glacier Deep Archive**
  - 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

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.
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

75%
S3 Batch Operations

Choose Objects

- S3 Inventory Report
- CSV List
S3 Batch Operations

Choose Objects
- S3 Inventory Report
- CSV List

Select an Operation
- Copy
- Restore from S3 Glacier
- Put Access Control List (ACL)
- Replace Object Tag Sets
- Run Lambda Functions
S3 Batch Operations

Choose Objects
- S3 Inventory Report
- CSV List

Select an Operation
- Copy
- Restore from S3 Glacier
- Put Access Control List (ACL)
- Replace Object Tag Sets
- Run Lambda Functions

View Progress
- Object Level Progress
- Job Notifications
- Completion Report
## Amazon S3 Management Console list view

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.

### S3 batch operations

<table>
<thead>
<tr>
<th>Job ID</th>
<th>Description</th>
<th>Operation</th>
<th>Date created</th>
<th>Status</th>
<th>Total objects</th>
<th>% Complete</th>
<th>Total failed (rate)</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>fc4b4b6b-8ead-4225-b7bf-88d4e87c75d5c</td>
<td>2019-02-20 Set-Tags ProjectDelta</td>
<td>Replace all tags</td>
<td>Feb 20, 2019 3:06:49 PM GMT-0800</td>
<td>Complete</td>
<td>2,717,669</td>
<td>100%</td>
<td>0 (0%)</td>
<td>10</td>
</tr>
<tr>
<td>db2971a-4b5c-4ac-847e-23af28340f39</td>
<td>2019-02-19 Set-Tags</td>
<td>Replace all tags</td>
<td>Feb 20, 2019 2:13:55 PM GMT-0800</td>
<td>Complete</td>
<td>2,700,671</td>
<td>100%</td>
<td>0 (0%)</td>
<td>10</td>
</tr>
<tr>
<td>5318b394-c6fd-4d1b-9456-b5112e12a444</td>
<td>2019-02-19 Set-Tags</td>
<td>Replace all tags</td>
<td>Feb 20, 2019 2:13:38 PM GMT-0800</td>
<td>Complete</td>
<td>2,700,671</td>
<td>100%</td>
<td>0 (0%)</td>
<td>10</td>
</tr>
<tr>
<td>ed4f5617-b7e3-4576-a376-f312d388c82</td>
<td>2019-02-19 Set-Tags</td>
<td>Replace all tags</td>
<td>Feb 20, 2019 2:13:20 PM GMT-0800</td>
<td>Complete</td>
<td>2,700,671</td>
<td>100%</td>
<td>0 (0%)</td>
<td>10</td>
</tr>
</tbody>
</table>
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 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

<table>
<thead>
<tr>
<th>Job Description</th>
<th>Job State</th>
<th>Job Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tag All Objects</td>
<td>Active</td>
<td>10</td>
</tr>
<tr>
<td>Copy to Oregon</td>
<td>Active</td>
<td>10</td>
</tr>
<tr>
<td>Restore Prefix A</td>
<td>Active</td>
<td>10</td>
</tr>
<tr>
<td>Change ACLs</td>
<td>Active</td>
<td>10</td>
</tr>
<tr>
<td>Lambda Demo</td>
<td>Ready</td>
<td>10</td>
</tr>
</tbody>
</table>
# Using S3 Batch Operations priorities

<table>
<thead>
<tr>
<th>Job Description</th>
<th>Job State</th>
<th>Job Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tag All Objects</td>
<td>Active</td>
<td>10</td>
</tr>
<tr>
<td>Copy to Oregon</td>
<td>Active</td>
<td>10</td>
</tr>
<tr>
<td>Restore Prefix A</td>
<td>Active</td>
<td>10</td>
</tr>
<tr>
<td>Change ACLs</td>
<td>Active</td>
<td>10</td>
</tr>
<tr>
<td>Lambda Demo</td>
<td>Ready</td>
<td>10</td>
</tr>
</tbody>
</table>
## Using S3 Batch Operations priorities

<table>
<thead>
<tr>
<th>Job Description</th>
<th>Job State</th>
<th>Job Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tag All Objects</td>
<td>Active</td>
<td>10</td>
</tr>
<tr>
<td>Copy to Oregon</td>
<td>Active</td>
<td>10</td>
</tr>
<tr>
<td>Restore Prefix A</td>
<td>Active</td>
<td>10</td>
</tr>
<tr>
<td>Change ACLs</td>
<td>Active</td>
<td>10</td>
</tr>
<tr>
<td>Lambda Demo</td>
<td>Ready</td>
<td>15</td>
</tr>
</tbody>
</table>
## Using S3 Batch Operations priorities

<table>
<thead>
<tr>
<th>Job Description</th>
<th>Job State</th>
<th>Job Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tag All Objects</td>
<td>Active</td>
<td>10</td>
</tr>
<tr>
<td>Copy to Oregon</td>
<td>Active</td>
<td>10</td>
</tr>
<tr>
<td>Restore Prefix A</td>
<td>Active</td>
<td>10</td>
</tr>
<tr>
<td>Change ACLs</td>
<td>Pausing</td>
<td>10</td>
</tr>
<tr>
<td>Lambda Demo</td>
<td>Active</td>
<td>15</td>
</tr>
</tbody>
</table>
Using S3 Batch Operations priorities

<table>
<thead>
<tr>
<th>Job Description</th>
<th>Job State</th>
<th>Job Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tag All Objects</td>
<td>Active</td>
<td>10</td>
</tr>
<tr>
<td>Copy to Oregon</td>
<td>Active</td>
<td>10</td>
</tr>
<tr>
<td>Restore Prefix A</td>
<td>Active</td>
<td>10</td>
</tr>
<tr>
<td>Change ACLs</td>
<td>Paused</td>
<td>10</td>
</tr>
<tr>
<td>Lambda Demo</td>
<td>Active</td>
<td>15</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Jobs</th>
<th>$0.25 per job</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objects</td>
<td>$1.00 per million object operations performed</td>
</tr>
</tbody>
</table>

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

s3://nwxtryingthingsout

Browse  View  Create bucket

Format: s3://mybucket. Learn more

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.
### Copy objects to a new storage class

Can be used as an alternative to lifecycle policies

<table>
<thead>
<tr>
<th>Storage class</th>
<th>Designed for</th>
<th>Availability Zones</th>
<th>Min storage duration</th>
<th>Min billable object size</th>
<th>Monitoring and automation fees</th>
<th>Retrieval fees</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>Frequently accessed data</td>
<td>≥ 3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Intelligent-Tiering</td>
<td>Long-lived data with changing or unknown access patterns</td>
<td>≥ 3</td>
<td>30 days</td>
<td>-</td>
<td>Per-object fees apply</td>
<td>-</td>
</tr>
<tr>
<td>Standard-IA</td>
<td>Long-lived, infrequently accessed data</td>
<td>≥ 3</td>
<td>30 days</td>
<td>128KB</td>
<td>-</td>
<td>Per-GB fees apply</td>
</tr>
<tr>
<td>One Zone-IA</td>
<td>Long-lived, infrequently accessed, non-critical data</td>
<td>≥ 1</td>
<td>30 days</td>
<td>128KB</td>
<td>-</td>
<td>Per-GB fees apply</td>
</tr>
<tr>
<td>Glacier</td>
<td>Archive data with retrieval times ranging from minutes to hours</td>
<td>≥ 3</td>
<td>90 days</td>
<td>-</td>
<td>-</td>
<td>Per-GB fees apply</td>
</tr>
</tbody>
</table>
Restore objects from S3 Glacier

Can restore and copy using the same manifest
## Pricing for an S3 Glacier restore job

<table>
<thead>
<tr>
<th>Usage</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Job</td>
<td>$0.25 ($0.25 / job)</td>
</tr>
<tr>
<td>1,000,000 objects</td>
<td>$1.00 ($1.00 / 1M objects)</td>
</tr>
<tr>
<td>1,000,000 restore requests</td>
<td>$25.00 ($0.025 / 1k requests)</td>
</tr>
<tr>
<td>977 GB restored (bulk tier)</td>
<td>$2.44 ($0.0025 / GB)</td>
</tr>
<tr>
<td><strong>Total:</strong></td>
<td><strong>$28.69</strong></td>
</tr>
</tbody>
</table>
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