Implementation Guide

Data Transfer from Amazon S3 Glacier Vaults to Amazon S3



Copyright © 2025 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Data Transfer from Amazon S3 Glacier Vaults to Amazon S3: Implementation Guide

Copyright © 2025 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon's trademarks and trade dress may not be used in connection with any product or service that is not Amazon's, in any manner that is likely to cause confusion among customers, or in any manner that disparages or discredits Amazon. All other trademarks not owned by Amazon are the property of their respective owners, who may or may not be affiliated with, connected to, or sponsored by Amazon.

Table of Contents

| Guidance overview | . 1 |
|---|----------|
| Features and benefits | . 3 |
| Use cases | . 5 |
| Concepts and definitions | . 5 |
| Architecture overview | 7 |
| Architecture diagram | . 7 |
| Translation of S3 Glacier vault archive descriptions to S3 object names | . 9 |
| Creating custom file names for S3 objects | 10 |
| AWS Well-Architected design considerations | 11 |
| Operational excellence | 11 |
| Security | 11 |
| Reliability | 11 |
| Performance efficiency | 12 |
| Cost optimization | 12 |
| Sustainability | 12 |
| Architecture details | 13 |
| AWS services in this Guidance | 13 |
| Third-party software support | 14 |
| Plan your deployment | 15 |
| Supported AWS Regions | 15 |
| Cost | 16 |
| Cost table calculation | 16 |
| Sample cost tables | 18 |
| AWS CloudTrail cost | 20 |
| Potential cost savings | 20 |
| Security | 20 |
| Amazon DynamoDB | 20 |
| CloudWatch Logs | 21 |
| IAM roles | 21 |
| Quotas | 21 |
| Quotas for AWS services in this Guidance | 21 |
| AWS CloudFormation quotas | 21 |
| Lambda concurrent execution quota | 22 |
| Amazon S3 Glacier Initiate Job quota | 22 |

| Amazon S3 file size limit | 22 |
|--|--|
| Amazon S3 storage class considerations | 22 |
| Amazon S3 Glacier resource considerations | 22 |
| Amazon S3 Glacier Vault Lock policy considerations | 23 |
| Deploy the Guidance | 24 |
| Deployment process overview | 24 |
| Prerequisites | 25 |
| AWS CloudFormation template | 26 |
| Step 1: Launch the stack | 27 |
| Step 2: Launch the transfer workflow | 30 |
| (Optional) Download the vault inventory file | 30 |
| (Optional) Provide the vault inventory file | 33 |
| Step 3: Resume the transfer workflow | 35 |
| Monitor the Guidance with AppRegistry | 37 |
| Activate CloudWatch Application Insights | 37 |
| Confirm cost allocation tags | 39 |
| Activate cost allocation tags | 39 |
| AWS Cost Explorer | 40 |
| | 44 |
| Update the Guidance | 41 |
| Update the Guidance Troubleshooting | 41 42 |
| Update the Guidance Troubleshooting Problem: Transfer workflow has not progressed after 14 hours | 41 42 42 |
| Update the Guidance Troubleshooting Problem: Transfer workflow has not progressed after 14 hours Resolution | 41 42 42 42 |
| Update the Guidance Troubleshooting Problem: Transfer workflow has not progressed after 14 hours Resolution Problem: Transfer workflow must be stopped | 41 42 42 42 42 |
| Update the Guidance Troubleshooting Problem: Transfer workflow has not progressed after 14 hours Resolution Problem: Transfer workflow must be stopped Resolution | 41 42 42 42 42 42 |
| Update the Guidance Troubleshooting Problem: Transfer workflow has not progressed after 14 hours Resolution Problem: Transfer workflow must be stopped Resolution Uninstall the Guidance | 41 42 42 42 42 42 43 |
| Update the Guidance Troubleshooting Problem: Transfer workflow has not progressed after 14 hours Resolution Problem: Transfer workflow must be stopped Resolution Uninstall the Guidance Using the AWS Management Console | 41 42 42 42 42 42 43 |
| Update the Guidance Troubleshooting Problem: Transfer workflow has not progressed after 14 hours Resolution Problem: Transfer workflow must be stopped Resolution Uninstall the Guidance Using the AWS Management Console Using AWS Command Line Interface | 41 42 42 42 42 42 43 43 43 |
| Update the Guidance Troubleshooting Problem: Transfer workflow has not progressed after 14 hours Resolution Problem: Transfer workflow must be stopped Resolution Uninstall the Guidance Using the AWS Management Console Using AWS Command Line Interface Deleting the S3 buckets | 41 42 42 42 42 42 43 43 43 |
| Update the Guidance Troubleshooting Problem: Transfer workflow has not progressed after 14 hours Resolution Problem: Transfer workflow must be stopped Resolution Uninstall the Guidance Using the AWS Management Console Using AWS Command Line Interface Deleting the S3 buckets Deleting the DynamoDB tables | 41 42 42 42 42 42 43 43 43 43 |
| Update the Guidance Troubleshooting Problem: Transfer workflow has not progressed after 14 hours Resolution Problem: Transfer workflow must be stopped Resolution Uninstall the Guidance Using the AWS Management Console Using AWS Command Line Interface Deleting the S3 buckets Deleting the DynamoDB tables Deleting the CloudWatch Logs | 41 42 42 42 42 42 42 43 43 43 43 44 44 |
| Update the Guidance Troubleshooting Problem: Transfer workflow has not progressed after 14 hours Resolution Problem: Transfer workflow must be stopped Resolution Uninstall the Guidance Using the AWS Management Console Using AWS Command Line Interface Deleting the S3 buckets Deleting the DynamoDB tables Deleting the CloudWatch Logs Use the Guidance | 41 42 42 42 42 42 42 43 43 43 43 43 44 44 |
| Update the Guidance Troubleshooting Problem: Transfer workflow has not progressed after 14 hours Resolution Problem: Transfer workflow must be stopped Resolution Uninstall the Guidance Using the AWS Management Console Using AWS Command Line Interface Deleting the S3 buckets Deleting the DynamoDB tables Deleting the CloudWatch Logs Validate your inventory | 41 42 42 42 42 42 43 43 43 43 43 44 44 46 |
| Update the Guidance Troubleshooting Problem: Transfer workflow has not progressed after 14 hours Resolution Problem: Transfer workflow must be stopped Resolution Uninstall the Guidance Using the AWS Management Console Using AWS Command Line Interface Deleting the S3 buckets Deleting the S3 buckets Deleting the CloudWatch Logs Use the Guidance Validate your inventory Access the CloudWatch dashboard | 41 42 42 42 42 42 43 43 43 43 43 43 43 43 43 44 44 46 46 |
| Update the Guidance Troubleshooting Problem: Transfer workflow has not progressed after 14 hours Resolution Problem: Transfer workflow must be stopped Resolution Uninstall the Guidance Using the AWS Management Console Using AWS Command Line Interface Deleting the S3 buckets Deleting the DynamoDB tables Deleting the CloudWatch Logs Use the Guidance Validate your inventory Access the CloudWatch dashboard Manage your Amazon S3 storage | 41 42 42 42 42 42 42 43 43 43 43 43 43 43 43 44 44 46 46 46 |
| Update the Guidance Troubleshooting Problem: Transfer workflow has not progressed after 14 hours Resolution Problem: Transfer workflow must be stopped Resolution Uninstall the Guidance Using the AWS Management Console Using AWS Command Line Interface Deleting the S3 buckets Deleting the DynamoDB tables Deleting the CloudWatch Logs Use the Guidance Validate your inventory Access the CloudWatch dashboard Manage your Amazon S3 storage | 41 42 42 42 42 42 42 43 43 43 43 43 43 43 43 44 46 46 46 46 46 |

| Reference | 49 |
|----------------------------|------|
| Anonymized data collection | 49 |
| Contributors | 50 |
| Revisions | 51 |
| Notices | . 52 |

Automatically copy your Amazon S3 Glacier vault archives to an S3 bucket and storage classes

Data Transfer from Amazon S3 Glacier Vaults to Amazon S3 is a serverless Guidance that automates and optimizes the restore, copy, and transfer process of <u>Amazon Simple Storage</u> <u>Service Glacier</u> (Amazon S3 Glacier) vault archives. The Guidance copies all of the vault's archives to a defined <u>Amazon Simple Storage Service</u> (Amazon S3) bucket destination and <u>storage class</u>. Then you can attach <u>tags</u> to help you categorize your data, such as with data classification or cost allocation. A prebuilt <u>Amazon CloudWatch</u> dashboard provides a visualization of the copy operation progress.

<u> Important</u>

Amazon S3 and Amazon S3 Glacier are different AWS services.

Amazon S3 Glacier is an object storage service for low-cost data archiving and long-term backup. It stores *archives* in *vaults*. It doesn't offer storage classes. The Amazon S3 Glacier service provides a console. However, any archive operation, such as upload, download, or deletion, requires you to use the AWS CLI or write code. There is no console support for archive operations.

Amazon S3 is an object storage service for any type of data. It stores *objects* in *buckets*. It offers different storage classes for frequent access, infrequent access, archives, and optimized tiering. You can interact with the Amazon S3 service by using the Amazon S3 console or <u>AWS Command Line Interface</u> (AWS CLI).

The S3 Glacier Instant Retrieval, S3 Glacier Flexible Retrieval, and S3 Glacier Deep Archive storage classes are features of the Amazon S3 service. The S3 Glacier Flexible Retrieval storage class offers the same features as the Amazon S3 Glacier service. The Amazon S3 Glacier service doesn't offer storage classes.

For example, Saanvi works at AnyCompany Archives. Five years ago, she used the Amazon S3 Glacier service to store scanned copies of historical documents in a vault. AnyCompany just announced that they will have a different online exhibit each month, featuring documents that are stored in the S3 Glacier vault. To address this change of business:

• Saanvi wants to take advantage of the storage classes offered with the Amazon S3 service, including more flexibility in how files are stored and accessed.

- Using Data Transfer from Amazon S3 Glacier Vaults to Amazon S3, Saanvi can copy all of her document archives from her S3 Glacier vault to an S3 bucket. She can assign them to the S3 storage classes that best fit her use cases. For example, she can use the S3 Standard storage class for documents that will be featured in the first exhibit and accessed daily, and the S3 Glacier Deep Archive storage class for documents that won't be featured in any of the exhibits.
- Now that the documents are stored in the Amazon S3 service, Saanvi can also apply <u>S3 Lifecycle</u> configurations, tag her data, and use the Amazon S3 console.

(i) Note

This Guidance doesn't delete the original archives or the source S3 Glacier vault. You must manually delete the archives and vault. For more information, refer to <u>Deleting an Archive</u> <u>in Amazon S3 Glacier</u> in the *Amazon S3 Glacier Developer Guide*. If your source S3 Glacier vault has a <u>Vault Lock policy</u> that prevents deletion, you must delete this policy before deleting the original archives. However, if your Vault Lock policy is in the Locked state, you can't delete it. See <u>S3 Glacier Vault Lock</u> and <u>Abort Vault Lock</u> (DELETE lock-policy) in the Amazon S3 Glacier Developer Guide for more information.

This implementation guide provides an overview of the Data Transfer from Amazon S3 Glacier Vaults to Amazon S3 Guidance, its reference architecture and components, considerations for planning the deployment, and configuration steps for deploying the Guidance to the Amazon Web Services (AWS) Cloud.

The intended audience for using this Guidance's features and capabilities in their environment includes solution architects, business decision makers, DevOps engineers, data scientists, and cloud professionals. Practical experience with the AWS Cloud, Amazon S3 Glacier vaults, Amazon S3 buckets, and Amazon S3 storage classes is preferred.

Use this navigation table to quickly find answers to these questions:

| If you want to | Read |
|---|------|
| Know the cost for running this Guidance. | Cost |
| The estimated cost for running this Guidance in the US East (Ohio) Region is USD \$153.57 | |

| If you want to | Read |
|--|-----------------------------|
| to copy 100,000 S3 Glacier vault archives, totaling 100 TB of data, from an S3 Glacier vault to an S3 bucket. | |
| Understand the security considerations for this Guidance. | <u>Security</u> |
| Know how to plan for quotas for this Guidance. | <u>Quotas</u> |
| This Guidance uses <u>AWS Lambda</u> functions to transfer data. This affects your account-wide Lambda concurrency limit. | |
| Know which AWS Regions support this Guidance. | Supported AWS Regions |
| View or download the AWS CloudForm ation template included in this Guidance to automatically deploy the infrastructure resources (the "stack") for this Guidance. | AWS CloudFormation template |
| Access the source code and optionally use the AWS Cloud Development Kit (AWS CDK) to deploy the Guidance. | <u>GitHub repository</u> |

Features and benefits

The Guidance provides the following features:

Automation

Automate the process of restoring, copying, and transferring archives from a vault in the Amazon S3 Glacier service to a bucket in the Amazon S3 service. After you move your data, you can use the Amazon S3 console. A prebuilt Amazon CloudWatch dashboard helps you monitor metrics and visualize the copy operation progress.

Ability to assign storage classes

When you use this Guidance to move your data from the Amazon S3 Glacier service into the Amazon S3 service, you choose a storage class to assign to all of your objects. After your data is stored in the Amazon S3 service, you can change the storage classes to fit the use case for each file. We recommend carefully reviewing each storage class and its pricing details before deploying this Guidance. See Amazon S3 storage class considerations for more information.

Visibility and access to data

After the Guidance stores S3 Glacier archives as objects in the destination S3 bucket, you can add tags to data. Tagging offers benefits such data classification, permissions controls, object lifecycle management, and cost allocation. For more information, see <u>Categorizing your storage using tags</u> in the *Amazon Simple Storage Service User Guide*.

Flexibility to cancel your transfer and resume later

The Guidance tracks the progress of the transfer. You can stop and restart transfers without needing to retransfer existing archives. See <u>Problem: Transfer workflow must be stopped</u> for more information.

Cost optimization

Copy S3 Glacier vault archives to an S3 bucket and assign more <u>cost-effective storage classes</u>, such as:

- The low-cost S3 Glacier Deep Archive storage class for data that you rarely access
- The S3 Glacier Instant Retrieval storage class if you'll need your data quarterly but within milliseconds
- The S3 Standard storage class for data you'll need daily

You can also configure and apply <u>S3 Lifecycles</u> to transition your objects automatically into different storage classes, based on rules you set.

Integration with Service Catalog AppRegistry and Application Manager, a capability of AWS Systems Manager

This Guidance includes a <u>Service Catalog AppRegistry</u> resource to register the Guidance's CloudFormation template and its underlying resources as an application in both Service Catalog

AppRegistry and <u>Application Manager</u>. With this integration, centrally manage the Guidance's resources and enable application search, reporting, and management actions.

Use cases

Performance and cost optimization

Balancing cloud storage performance and storage cost is crucial for organizations. You can use this Guidance to help you optimize your storage infrastructure with the features that the Amazon S3 service offers. With tagging, lifecycle configurations, and a variety of storage classes, you can improve your performance while minimizing costs.

Cloud archiving

Many organizations store their most fundamental asset—their data—in locations that are slow to retrieve and lack flexibility. You can use this Guidance to help you automate, monitor, and seamlessly move your data when and where you need it.

Concepts and definitions

This section describes key concepts and defines terminology specific to this Guidance:

archive

Any data stored in an Amazon S3 Glacier vault, such as a photo, video, or document. An archive is similar to an Amazon S3 object: it's the base unit of storage in the Amazon S3 Glacier service. For more information, see <u>Archive</u> in the *Amazon S3 Glacier Developer Guide*.

chunk

Term used to describe a *part* in a multipart upload or download for the Amazon S3 Glacier service. This Guidance uses multipart upload to transfer the archives. For more information, see <u>Uploading Large Archives in Parts (Multipart Upload)</u> and <u>Retrieving S3 Glacier Archives Using AWS</u> Management Console in the *Amazon S3 Glacier Developer Guide*.

inventory

A point in time snapshot or listing of the archives stored within an S3 Glacier vault. For more information, see <u>Downloading a Vault Inventory in Amazon S3 Glacier</u> in the *Amazon S3 Glacier Developer Guide*.

tag

A key-value pair used to categorize storage in the Amazon S3 service. For more information, see Categorizing your storage using tags in the Amazon S3 User Guide.

vault

A container in the Amazon S3 Glacier service for storing archives. An S3 Glacier vault is similar to an S3 bucket. For more information, see <u>Vault</u> in the *Amazon S3 Glacier Developer Guide*.

workflow_run

An identifier used to represent the transfer of an S3 Glacier vault to an S3 bucket. The Guidance randomly generates the workflow_run value on the first run (or you can choose the value). The Guidance uses this value when resuming a transfer.

Note

For a general reference of AWS terms, see the AWS Glossary.

Architecture overview

This section provides a reference implementation architecture diagram for the components deployed with this Guidance.

Architecture diagram

Deploying this Guidance with the default parameters deploys the following components in your AWS account.



Data Transfer from Amazon S3 Glacier Vaults to Amazon S3 architecture on AWS

i Note

AWS CloudFormation resources are created from <u>AWS Cloud Development Kit (AWS CDK)</u> (AWS CDK) constructs.

The high-level process flow for the Guidance components deployed with the <u>AWS CloudFormation</u> template is as follows:

- 1. Invoke a transfer workflow by using a <u>Systems Manager document</u> (SSM document).
- 2. The Systems Manager document starts an <u>AWS Step Functions</u> Orchestrator execution.
- 3. The Step Functions Orchestrator execution initiates a nested Step Functions Get Inventory workflow to retrieve the inventory file.
- 4. Upon completion of the inventory retrieval, the Guidance invokes the Initiate Retrieval nested Step Functions workflow.
- 5. When a job is ready, the <u>Amazon S3 Glacier</u> service sends a notification to an <u>Amazon Simple</u> <u>Notification Service</u> (Amazon SNS) topic indicating job completion.
- 6. The Guidance stores all job completion notifications in the <u>Amazon Simple Queue Service</u> (Amazon SQS) Notifications queue.
- 7. When an archive job is ready, the Amazon SQS Notifications queue invokes the <u>AWS Lambda</u> Notifications Processor function. This Lambda function prepares the initial steps for archive retrieval.
- 8. The Lambda Notifications Processor function places chunks retrieval messages in Amazon SQS Chunks Retrieval queue for chunk processing.
- 9. The Amazon SQS Chunks Retrieval queue invokes the Lambda Chunk Retrieval function to process each chunk.
- 10. The Lambda Chunk Retrieval function downloads the chunk from the Amazon S3 Glacier service.
- 11. The Lambda Chunk Retrieval function uploads a <u>multipart upload</u> part to the <u>Amazon S3</u> service.
- 12After a new chunk is downloaded, the Guidance stores chunk metadata in <u>Amazon DynamoDB</u> (etag, checksum_sha_256, tree_checksum).
- 13. The Lambda Chunk Retrieval function verifies whether all chunks for that archive have been processed. If yes, it inserts an event into the Amazon SQS Validation queue to invoke the Lambda Validate function.
- 14. The Lambda Validate function does the following:
 - a. Performs an integrity check against the tree hash in the inventory.
 - b. Calculates a checksum and passes it to the into the close multipart upload call. If that hash is wrong, Amazon S3 rejects the request.

- 15A DynamoDB <u>stream</u> invokes the Lambda Metrics Processor function to update the transfer process metrics in DynamoDB.
- 16. The Step Functions Orchestrator execution enters an async wait, pausing until the archive retrieval workflow concludes before initiating the Step Functions Cleanup workflow.
- 17. The DynamoDB stream invokes the Lambda Async Facilitator function, which unlocks asynchronous waits in Step Functions.
- 18. The <u>Amazon EventBridge</u> rules periodically initiate Step Functions Extend Download Window and Update CloudWatch Dashboard workflows.
- 19Customers monitor the transfer progress by using the Amazon CloudWatch dashboard.

Translation of S3 Glacier vault archive descriptions to S3 object names

To create the key name for each of the new objects in the Amazon S3 service, this Guidance uses the **ArchiveDescription** value for each **ArchiveId** listed in the Amazon S3 Glacier inventory file. The following are examples.

1. If the **ArchiveDescription** is a single string value, such as data01, the Guidance translates that value to an S3 object key name in the destination S3 bucket.



- 2. If the ArchiveDescription value is blank, then the Guidance does the following:
 - a. Copies the archive.
 - b. Uses the Archiveld as the S3 object key name.
 - c. Adds the prefix 00undefined to the S3 object key names and stores the objects in the destination S3 bucket.



3. If multiple **Archiveld** entries have the same value for the **ArchiveDescription** field (for example, duplicatefile02.txt), then the Guidance appends a timestamp suffix to the name of the original file. This resolves the potential issue of having duplicate S3 object key names copied over one another. The timestamp used is the CreationDate of the archive.

| | duplicatefile02.txt |
|----|--|
| Ľ | duplicatefile02.txt-2019-10-17T00:51:33Z |
| Ľ٩ | duplicatefile02.txt-2019-10-17T00:51:34Z |

Creating custom file names for S3 objects

You can provide custom S3 object key names for each **Archiveld** that's copied to your S3 bucket. To do this, provide a **NamingOverrideFile** to the Guidance when you <u>launch the transfer workflow</u>, using the **NamingOverrideFile** input parameter. Use the following process.

 Create a data file in CSV format. The file must contain only two columns: GlacierArchiveID and FileName (separated by a comma). The following table is an example.

| GlacierArchiveID | FileName |
|------------------------------|----------------------|
| WVfrXME2KC6JIedfadJF937412-e | Mydata.txt |
| yLam5H76JXYSKKIY34404D-Kwcrk | Myfolder/mydata2.txt |

- Obtain a copy of your vault inventory file for the Amazon S3 Glacier service. For more information, see <u>Downloading a Vault Inventory in Amazon S3 Glacier</u> in the Amazon S3 Glacier Developer Guide.
- 3. Copy all the **Archiveld** values from your S3 Glacier vault inventory file. Paste them into the **GlacierArchiveID** column of your **NamingOverride** CSV file.
- 4. In the FileName column, for each ArchiveID, enter your desired S3 object key name.

Note

If you provide an empty value for the **FileName**, the Guidance uses the original value for **ArchiveDescription** from the S3 Glacier archive.

- 5. Upload the CSV file to any S3 bucket and create a presigned URL for the file.
- 6. Use this presigned URL as the value of the **NamingOverride File** input parameter used when launching the transfer workflow.

AWS Well-Architected design considerations

This Guidance uses the best practices from the <u>AWS Well-Architected Framework</u>, which helps customers design and operate reliable, secure, efficient, and cost-effective workloads in the cloud.

This section describes how the design principles and best practices of the Well-Architected Framework benefit this Guidance.

Operational excellence

This section describes how we architected this Guidance using the principles and best practices of the <u>operational excellence pillar</u>.

• This Guidance pushes metrics to CloudWatch at various stages to provide visibility into archive transfer progress.

Security

This section describes how we architected this Guidance using the principles and best practices of the <u>security pillar</u>.

- All interservice communications use applicable <u>AWS Identity and Access Management</u> (IAM) roles.
- All roles used by the Guidance follow least privilege access. They only contain the minimum permissions required to accomplish the transfer.
- All data storage, including the S3 buckets, encrypts the data at rest.

Reliability

This section describes how we architected this Guidance using the principles and best practices of the <u>reliability pillar</u>.

- The Guidance uses a serverless architecture to achieve high availability and recovery from failure.
- The Guidance protects against state machine definition errors through a suite of automated tests.
- Data processing uses Lambda functions. Data is stored in DynamoDB and Amazon S3, which persist in multiple Availability Zones by default.

Performance efficiency

This section describes how we architected this Guidance using the principles and best practices of the performance efficiency pillar.

- The Guidance uses a serverless architecture with the ability to scale horizontally as needed.
- The Guidance is tested and deployed daily to achieve consistency as AWS services change.

Cost optimization

This section describes how we architected this Guidance using the principles and best practices of the cost optimization pillar.

- The Guidance uses a serverless architecture that only charges customers for what they use.
- DynamoDB global secondary indexes are selected to reduce the pricing for queries.

Sustainability

This section describes how we architected this Guidance using the principles and best practices of the sustainability pillar.

• The Guidance uses managed serverless services to minimize the environmental impact of the backend services compared to continually operating on-premises services.

Architecture details

This section describes the components and AWS services that make up this Guidance and the architecture details on how these components work together.

AWS services in this Guidance

| AWS service | Description |
|------------------------|---|
| Amazon S3 | Core. Stores inventory and archives transferr ed from the Amazon S3 Glacier service. |
| Amazon S3 Glacier | Core. Maintains the vault being transferred to the Amazon S3 service. |
| <u>Amazon DynamoDB</u> | Core. Stores workflow-related data, including hash validation values and progress metrics for S3 Glacier archives. |
| <u>AWS Glue</u> | Core. Reorders and splits the S3 Glacier vault inventory file into partitions for processing by multiple Lambda invocations. Parses file names from Inventory/OverrideNaming file. |
| AWS Lambda | Core. Performs workflow compute, including archive transfer and validation. |
| <u>Amazon SNS</u> | Supporting. Decouples Lambda workflows by providing communication between the Guidance, the Amazon S3 Glacier service, and Amazon SQS queues. |
| Amazon SQS | Supporting. Decouples Lambda workflows. |
| Amazon CloudWatch | Supporting. Stores the Guidance logs and metrics. Presents a custom dashboard to |

| AWS service | Description |
|---------------------|---|
| | provide visibility of the archive copy operation progress. |
| Amazon EventBridge | Supporting. Invokes supporting workflows that promote the smooth operation of the Guidance. |
| IAM | Supporting. Provides permissions for this Guidance's resources to perform actions. |
| AWS Systems Manager | Supporting. Provides a Systems Manager <u>document</u> (SSM document) for Guidance's workflow invocations. |
| AWS Step Functions | Supporting. Orchestrates partitioning the inventory with AWS Glue. Also orchestrates Lambda invocations to request S3 Glacier vault archive retrieval. |

Third-party software support

This Guidance uses the value stored in the **ArchiveDescription** for each S3 Glacier vault archive (as listed in the S3 Glacier inventory file) as the key name for the new S3 object that it creates. The Guidance supports copying S3 Glacier vaults using either FastGlacier or CloudBerry software as follows.

- FastGlacier (v1-v4) The Guidance extracts the value for /ArchiveMetadata/Path of /m/p from the XML metadata stored in the ArchiveDescription field in the S3 Glacier inventory file. It then converts that value to a string that forms the S3 object key name.
- **CloudBerry (v5.9)** The Guidance extracts the value for **Path** from the JSON metadata stored in the **ArchiveDescription** field in the S3 Glacier inventory file. It then converts that value to a string that forms the S3 object key name.

Plan your deployment

This section describes the Regions, <u>cost</u>, <u>security</u>, <u>quota</u>, and other considerations before deploying the Guidance.

Supported AWS Regions

This Guidance uses AWS services that are not currently available in all AWS Regions. For the most current availability of AWS services by Region, see the AWS Regional Services List.

🚯 Note

This Guidance launches in the US East (Ohio) Region by default. To launch the Guidance in a different AWS Region, use the Region selector in the console navigation bar. See <u>Step 1:</u> <u>Launch the stack</u> for more information.

| Region name | | | |
|--------------------------|-----------------------|--|--|
| US East (N. Virginia) | Canada (Central) | | |
| US East (Ohio) | China (Beijing) | | |
| US West (N. California) | China (Ningxia) | | |
| US West (Oregon) | Europe (Frankfurt) | | |
| Africa (Cape Town) | Europe (Ireland) | | |
| Asia Pacific (Hong Kong) | Europe (London) | | |
| Asia Pacific (Mumbai) | Europe (Milan) | | |
| Asia Pacific (Osaka) | Europe (Paris) | | |
| Asia Pacific (Seoul) | Europe (Stockholm) | | |
| Asia Pacific (Singapore) | Middle East (Bahrain) | | |

| Region name | |
|-----------------------|---------------------------|
| Asia Pacific (Sydney) | South America (São Paulo) |
| Asia Pacific (Tokyo) | AWS GovCloud (US-West) |

Cost

You are responsible for the cost of the AWS services used while running this Guidance. As of this revision, the cost for running this Guidance with the default settings in the US East (Ohio) Region is approximately **\$153.57** for 100,000 S3 Glacier vault archives (1GB each) and **\$1,229.21** for 10,000,000 S3 Glacier vault archives (10MB each). These costs assume that the destination bucket is also in US East (Ohio) Region. Refer to Sample cost tables for more details.

🚯 Note

If the destination bucket is not in the same region as the Glacier vault, a "Data Transfer OUT From Amazon S3 Glacier" fee will be added. See <u>Data transfer pricing</u> for more information. This cost should be considered when planning your data storage and transfer strategies to avoid unexpected charges.

See the pricing webpage for each AWS service used in this Guidance. Estimated costs vary based on the number of archives processed and the total volume of data to copy from an S3 Glacier vault.

We recommend creating a <u>budget</u> through AWS Cost Explorer to help manage costs. Prices are subject to change. For full details, see the pricing webpage for each <u>AWS service used in this</u> <u>Guidance</u>.

Cost table calculation

The following table shows how the sample cost tables were calculated.

| Guidance component | Туре | [A] - Unit cost | [B] - Value | [C] - Estimated cost |
|-----------------------|-----------------------|-----------------|---------------------|-------------------------|
| Additional services | Per 1,000 requests | \$0.01 | <# of S3 Glacier | [A] × [B] / 1,000 |

| Guidance component | Туре | [A] - Unit cost | [B] - Value | [C] - Estimated cost |
|---|-----------------------|-----------------|---|-------------------------|
| | | | vault archives> | |
| Amazon S3 multipart upload create requests | Per 1,000 requests | \$0.005 | <# of S3 Glacier vault archives> | [A] × [B] / 1,000 |
| Amazon S3 multipart upload complete requests | Per 1,000 requests | \$0.03 | <# of S3 Glacier vault archives> | [A] × [B] / 1,000 |
| Guidance runtime (Lambda and Step Functions) | Per GB | \$0.00143 | <size of<br="">S3 Glacier vault in GBs></size> | [A] × [B] |
| Guidance runtime (Lambda and Step Functions) | Per 1,000 requests | \$0.0447 | <# of S3 Glacier vault archives> | [A] × [B] / 1,000 |
| DynamoDB writes/reads for transfer metadata | Per 1,000 requests | \$0.02 | <# of S3 Glacier vault archives> | [A] × [B] / 1,000 |

| Guidance component | Туре | [A] - Unit cost | [B] - Value | [C] - Estimated cost |
|---|--------|---|---|-------------------------|
| Data Transfer OUT from S3 Glacier | Per GB | \$0.02 <i>if destination</i> <i>bucket is in a</i> <i>different region</i> <i>than S3 Glacier</i> <i>vault.</i> \$0.00 <i>if destination</i> <i>bucket is in the</i> <i>same region as</i> <i>S3 Glacier vault.</i> | <size of<br="">S3 Glacier vault in GBs></size> | [A] × [B] |

Sample cost tables

The following tables provide two sample cost breakdowns for deploying this Guidance with the default parameters in the US East (Ohio) Region, with an S3 Glacier vault size of 100 TB. These cost breakdowns are based on the destination bucket is also being in the US East (Ohio) Region, the same region as the S3 Glacier vault.

(i) Note

Costs associated with storing data in the Amazon S3 service are nearly continuous and aren't included in these estimates.

Scenario 1: 100,000 S3 Glacier vault archives

| AWS service | Dimensions | Cost [USD] |
|----------------|------------|------------|
| Step Functions | | \$0.07 |

| AWS service | Dimensions | Cost [USD] |
|----------------------|---------------|----------------|
| Lambda | | \$140.00 |
| DynamoDB | | \$2.00 |
| Amazon S3 | Transfer cost | \$5.00 |
| Additional services: | | \$6.50 |
| Amazon SQS | | |
| Amazon SNS | | |
| • AWS Glue | | |
| CloudWatch | | |
| | Total: | \$153.57 [USD] |

If the destination bucket is in a different region than US East (Ohio), an additional price of \$2048 (\$0.02 x size of S3 Glacier vault in GBs) should be added to the total.

Scenario 2: 10,000,000 S3 Glacier vault archives

| AWS service | Dimensions | Cost [USD] |
|---|---------------|------------|
| Step Functions | | \$3.21 |
| Lambda | | \$411.00 |
| DynamoDB | | \$221.00 |
| Amazon S3 | Transfer cost | \$465.00 |
| Additional services: Amazon SQS | | \$129.00 |
| Amazon SNSAWS Glue | | |
| CloudWatch | | |

| AWS service | Dimensions | Cost [USD] |
|-------------|------------|------------------|
| | Total: | \$1,229.21 [USD] |

If the destination bucket is in a different region than US East (Ohio), an additional price of \$2048 (\$0.02 x size of S3 Glacier vault in GBs) should be added to the total.

AWS CloudTrail cost

You can use <u>AWS CloudTrail</u> to log, continuously monitor, and retain account activity related to actions across your AWS infrastructure. When you create additional trails—for example, to capture data or insight on generated events—AWS CloudTrail charges apply. See <u>AWS CloudTrail pricing</u> for more information.

Potential cost savings

As of this revision, you can save \$0.00261 per GB per month by storing your archives in the Amazon S3 service with the S3 Glacier Deep Archive storage class applied. For example, if you have 100 TB stored in your S3 Glacier vault, you can save \$261.00 per month by storing that data in the Amazon S3 service with the S3 Glacier Deep Archive storage class applied.

The cost to run the Guidance scales with the size of the S3 Glacier vault and the number of archives. Cost savings only scale with the S3 Glacier vault size.

Security

When you build systems on AWS infrastructure, security responsibilities are shared between you and AWS. This <u>shared responsibility model</u> reduces your operational burden because AWS operates, manages, and controls the components including the host operating system, the virtualization layer, and the physical security of the facilities in which the services operate. For more information about AWS security, visit <u>AWS Cloud Security</u>.

Amazon DynamoDB

All user data stored in DynamoDB is encrypted at rest using encryption keys stored in <u>AWS Key</u> <u>Management Service</u> (AWS KMS). We recommend enforcing <u>AWS managed keys</u> because you have permission to <u>audit their use</u> in AWS CloudTrail logs. Refer to <u>Managing encrypted tables in</u> <u>DynamoDB</u> for more information. Consider enabling DynamoDB Data Plane Events for CloudTrail logging to gain insights into the data operations in DynamoDB tables, according to your use cases and your regulatory and compliance requirements. Refer to <u>Logging DynamoDB operations by using AWS CloudTrail</u> for more information. Additionally, consider implementing <u>AWS Config</u> to actively monitor DynamoDB configuration changes

CloudWatch Logs

We recommend <u>changing the retention period</u> of your <u>CloudWatch Logs</u> according to your use cases and your regulatory and compliance requirements.

IAM roles

IAM roles allow you to assign granular access policies and permissions to services and users on the AWS Cloud. This Guidance creates IAM roles that grant the Guidance's resources permission to access the S3 Glacier vault, write logs, and create EventBridge targets.

Quotas

Service quotas, also referred to as limits, are the maximum number of service resources or operations for your AWS account.

Quotas for AWS services in this Guidance

Make sure you have a sufficient quota for each of the <u>services implemented in this Guidance</u>. For more information, see <u>AWS service quotas</u>.

Use the following links to go to the page for that service. To view the service quotas for all AWS services in the documentation without switching pages, view the information in the <u>Service</u> <u>endpoints and quotas</u> page in the PDF instead.

AWS CloudFormation quotas

Your AWS account has AWS CloudFormation quotas that you should be aware of when <u>launching</u> <u>the stack</u> in this Guidance. By understanding these quotas, you can avoid limitation errors that would prevent you from deploying this Guidance successfully. For more information, see <u>AWS</u> <u>CloudFormation quotas</u> in the AWS CloudFormation User Guide.

Lambda concurrent execution quota

Your AWS account has a quota on the number of concurrent Lambda executions that can be running. For more information, see <u>Lambda quotas</u> in the *AWS Lambda Developer Guide*. This Guidance uses 230–250 concurrent Lambda executions when running at maximum capacity.

Amazon S3 Glacier Initiate Job quota

This Guidance optimizes your transfer by requesting archives in order. Other random restore requests can impact throughput.

The Amazon S3 Glacier service maintains a service quota of <u>35 random restore requests</u> per PiB stored per day. If you continue to initiate your archive retrievals as the Guidance runs, Amazon S3 Glacier responses might slow down. You might also see Amazon S3 Glacier <u>ThrottlingExceptions</u> if you initiate archive retrievals external to the Guidance.

Amazon S3 file size limit

The Amazon S3 service restricts file sizes to 5 TB. The Guidance won't transfer archives larger than 5 TB. The Guidance's CloudWatch dashboard indicates the number of archives that meet this condition. The Guidance stores inventory data for these archives in the Inventory S3 bucket under \$WORKFLOW_RUN/not_migrated/.

Amazon S3 storage class considerations

When you deploy this Guidance, you must choose a storage class to apply to all of your transferred data. Before you choose this storage class, consider the availability, durability, minimum storage duration, and cost of each storage class. After your data is stored in the Amazon S3 service, you can change the storage class for each object. Some storage classes have minimum durations, so it's important to plan accordingly. For more information, see <u>Comparing the Amazon S3 storage</u> <u>classes</u> in the *Amazon Simple Storage Service User Guide* and <u>Amazon S3 pricing</u>.

Amazon S3 Glacier resource considerations

This Guidance uses the entirety of your Amazon S3 Glacier service resources. You won't be able to use your Amazon S3 Glacier archive while the Guidance is running.

Amazon S3 Glacier Vault Lock policy considerations

This Guidance doesn't delete the original archives or the source S3 Glacier vault. You must manually delete the archives and vault. For more information, refer to <u>Deleting an Archive in</u> Amazon S3 Glacier in the Amazon S3 Glacier Developer Guide.

If your source S3 Glacier vault has a <u>Vault Lock policy</u> that prevents deletion, you must delete this policy before deleting the original archives. However, if your Vault Lock policy is in the Locked state, you can't delete it. See <u>S3 Glacier Vault Lock</u> and <u>Abort Vault Lock (DELETE lock-policy)</u> in the Amazon S3 Glacier Developer Guide for more information.

Deploy the Guidance

This Guidance uses <u>AWS CloudFormation templates and stacks</u> to automate its deployment. The CloudFormation template specifies the AWS resources included in this Guidance and their properties. The CloudFormation stack provisions the resources that are described in the template.

🚯 Note

This Guidance doesn't delete the original archives or the source S3 Glacier vault. You must manually delete the archives and vault. For more information, refer to <u>Deleting an Archive</u> <u>in Amazon S3 Glacier</u> in the *Amazon S3 Glacier Developer Guide*. If your source S3 Glacier vault has a <u>Vault Lock policy</u> that prevents deletion, you must delete this policy before deleting the original archives. However, if your Vault Lock policy is in the Locked state, you can't delete it. See <u>S3 Glacier Vault Lock</u> and <u>Abort Vault Lock</u> (DELETE lock-policy) in the Amazon S3 Glacier Developer Guide for more information.

Deployment process overview

Follow the step-by-step instructions in this section to configure and deploy the Guidance into your account.

Before you launch the Guidance, review the <u>cost</u>, <u>architecture</u>, <u>security</u>, and other considerations discussed earlier in this guide.

Time to deploy: Approximately 5–10 minutes

🚯 Note

The archive transfer can take up to one day to complete. If you need to move your data faster, contact <u>AWS Support</u> (AWS Developer Support plan or above).

Step 1: Launch the stack

Step 2: Launch the transfer workflow

Step 3: Resume the transfer workflow

<u> Important</u>

This Guidance includes an option to send anonymized operational metrics to AWS. We use this data to better understand how customers use this Guidance and related services and products. AWS owns the data gathered though this survey. Data collection is subject to the AWS Privacy Notice.

To opt out of this feature, download the template, modify the AWS CloudFormation mapping section, and then use the AWS CloudFormation console to upload your updated template and deploy the Guidance. For more information, see the <u>Anonymized data</u> <u>collection</u> section of this guide.

Prerequisites

Before deployment, ensure that there are no new uploads or deletes of archives occurring on your source Glacier vault. Your Glacier vault content must be static.

- Ensure you have reviewed the cost section before deployment.
- Create a new destination Amazon S3 bucket. This bucket will be the destination storage location for your Glacier vault archives. For more information, refer to <u>Creating a bucket</u> in the Amazon Simple Storage Service User Guide.
 - The new destination Amazon S3 bucket should be in the same Region as the S3 Glacier vault, otherwise an excessive additional cost of "Data Transfer OUT From Amazon S3 Glacier" will be added, see Data transfer pricing.

🚯 Note

The destination Amazon S3 bucket must be created in the same account as the Glacier vault, which is also the account where the Guidance is deployed. Currently, the Guidance does not support cross-account transfers.

- It is advisable to review and modify any Service Control Policies (SCP) on the destination bucket that may block or prevent PUT operations.
- If you are using CloudTrail on your destination Amazon S3 bucket, please review and modify the CloudTrail export configurations to prevent excessive API charges.

• Ensure that your account has permissions to deploy the CloudFormation template and create the necessary AWS IAM roles. Your account must have permissions to grant access to the source Glacier vault and destination Amazon S3 bucket.

AWS CloudFormation template

You can download the CloudFormation template for this Guidance before deploying it.

View template

data-transfer-from-amazon-s3-glacier-vaults-to-amazon-s3.template – Use this template to launch the Guidance and all associated components. The default configuration deploys the core and supporting services found in the <u>AWS services in this Guidance</u> section, but you can customize the template to meet your specific needs.

🚺 Note

AWS CloudFormation resources are created from AWS Cloud Development Kit (AWS CDK) (AWS CDK) constructs.

This AWS CloudFormation template deploys Data Transfer from Amazon S3 Glacier Vaults to Amazon S3 in the AWS Cloud.

🔥 Important

- Notifications The Amazon S3 Glacier service sends one notification per archive to the vault Amazon SNS topic (if it exists). If you don't want subscribers to the Amazon SNS topic to receive these notifications, confirm that the S3 Glacier vault being transferred doesn't have notifications enabled. For more information, see <u>Configuring Vault</u> Notifications in Amazon S3 Glacier in the Amazon S3 Glacier Developer Guide.
- Inventory This Guidance copies your S3 Glacier vault archives once when you launch the stack. If you make changes to your vault after launching this Guidance, the Guidance doesn't replicate those changes.

 Simultaneous workflows – Running multiple transfer workflows simultaneously can exceed S3 Glacier quotas and induce throttling. We recommend that you only run one transfer workflow at a time.

Step 1: Launch the stack

Follow the step-by-step instructions in this section to configure and deploy the Guidance into your account.

Time to deploy: Approximately 5–10 minutes

Note

The archive transfer can take up to one day to complete. If you need to move your data faster, contact AWS Support (AWS Developer Support plan or above).

 Sign in to the <u>AWS Management Console</u> and select the button to launch the dataretrieval-from-amazon-s3-glacier-vaults-to-amazon-s3.template AWS CloudFormation template.

Launch solution

2. The template launches in the US East (Ohio) Region by default. To launch the Guidance in a different AWS Region, use the Region selector in the console navigation bar.

1 Note

This Guidance uses AWS services that are not currently available in all AWS Regions. For the most current availability by Region, see the <u>AWS Regional Services List</u>.

- 3. On the **Create stack** page, verify that the correct template URL is in the **Amazon S3 URL** text box and choose **Next**.
- 4. On the **Specify stack details** page, assign a name to your Guidance stack. For information about naming character limitations, see <u>IAM and AWS STS quotas</u>, <u>name requirements</u>, <u>and character</u> limits in the AWS Identity and Access Management User Guide.

5. Under **Parameters**, review the parameters for this Guidance template and modify them as necessary. This Guidance uses the following default values.

| Parameter | Default | Description |
|-------------------|---------|--|
| DestinationBucket | empty | The name of the destination Amazon S3 bucket. |
| | | ✔ Important Referring a bucket in a different region than S3 Glacier vault will incur additiona l costs. Refer to the cost section for more details. ✔ Note ↑ Note |

| Parameter | Default | Description |
|-----------------------|---------|---|
| DynamoDB Backup | false | Enter true to enable DynamoDB table backups for tables created by the Guidance. |
| Lambda Tracing | false | Enter true to enable <u>AWS X-Ray</u> tracing for Lambda functions created by the Guidance. |
| Step Function Logging | false | Enter true to enable logging for Step Functions created by the Guidance. |
| Step Function Tracing | false | Enter true to enable X-Ray tracing for Step Functions created by the Guidance. |

(i) Note

It is advisable to review and modify any <u>Service Control Policies</u> (SCP) on the destination bucket that may block or prevent PUT operations. If you are using CloudTrail on your destination Amazon S3 bucket, please review and modify the CloudTrail export configurations to prevent excessive API charges.

6. Select Next.

- 7. On the Configure stack options page, choose Next.
- 8. On the **Review** page, review and confirm the settings. Select the box acknowledging that the template will create IAM resources.
- 9. Choose **Create stack** to deploy the stack.

You can view the status of the stack in the AWS CloudFormation console in the **Status** column. You should receive a CREATE_COMPLETE status in approximately 5–10 minutes.

Step 2: Launch the transfer workflow

🔥 Important

Avoid starting multiple transfers simultaneously. Multiple transfers can exceed S3 Glacier quotas and induce throttling.

When you transfer multiple vaults, initiate transfers sequentially by starting a new transfer only after the previous one completes.

- 1. Sign in to the Systems Manager console.
- 2. Under Shared Resources, select Documents.
- On the Documents page, select Owned by me, and choose the document called Launch-Data-Retrieval-for-Glacier-S3-\$<<u>NAME_OF_STACK</u>>.
- 4. Choose Execute Automation.

(Optional) Download the vault inventory file

If you want to download the vault inventory file as part of this migration, follow these steps.

- 1. Sign in to the Systems Manager console.
- 2. Under Shared Resources, select Documents.
- On the Documents page, select Owned by me, and choose the document called Launch-Data-Retrieval-for-Glacier-S3-\$<<u>NAME_OF_STACK</u>>.
- 4. Choose **Execute Automation**.
- 5. Enter the following under Input Parameters on the Execute Automation Runbook page.

Note

To enable the data transfer to a bucket in a different region from the S3 Glacier vault, you must manually adjust the automation document's content by setting the InputPayload parameter allow_cross_region_data_transfer to true. By default, the Guidance restricts cross-region transfers to avoid potential costs.

| Parameter | Value | Notes |
|---|--------------------------------|---|
| AcknowledgeAdditio nalCostForCrossReg ionTransfer | NO | Select YES only if you are aware of the excessive additional cost when selecting a destination bucket in a different region than the S3 Glacier vault. See <u>Amazon S3 Glacier data</u> <u>transfer pricing</u> . |
| ProvidedInventory | <requires input=""></requires> | Input with two options [YES, NO] indicate if the inventory is provided. |
| VaultName | <requires input=""></requires> | Enter the name of your S3 Glacier vault. |

| Parameter | Value | Notes |
|--------------------|--------------------------------|---|
| WorkflowRun | <requires input=""></requires> | Input to specify the workflow identifier of the workflow that needs to be resumed. Note Input to specify the workflow identifier of the workflow that needs to be resumed. |
| Description | <optional input=""></optional> | Provide an extended description for this migration . |
| NamingOverrideFile | <optional input=""></optional> | Provide a presigned URL of the NamingOverride file and the bucket that is storing the file if you want to <u>customize</u> <u>S3 object key names</u> . |
| S3StorageClass | <requires input=""></requires> | Select the S3 storage class for the migrated archives. See <u>Amazon S3 pricing</u> . |

- 6. Choose **Execute**.
- 7. It takes approximately less than 1 minute to execute the document and launch the transfer. To confirm that the transfer process started successfully, refer to the *runScript* step status under the *Executed Steps* section.

(i) Note

To monitor the progress of the transfer after launching the workflow, please refer to the Guidance's <u>Cloudwatch dashboard</u>. Please note that it might take 5-12 hours before the

dashboard **Workflow Run ID** dropdown menu entries get updated with the new entry after launching the transfer.

(Optional) Provide the vault inventory file

If you want to provide the vault inventory file, follow these steps.

- 1. Sign in to the CloudFormation console.
 - a. On the **Stacks** page, select this Guidance's installation stack.
 - b. Choose the **Output** tab and find the inventory bucket name.
- 2. Sign in to the Amazon S3 console and choose the inventory bucket from step 1b.
 - a. Choose **Create folder**. Enter the workflow_run value from the **WorkflowRun** parameter input. For more information about creating and using folders, see <u>Organizing objects in the</u> Amazon S3 console by using folders in the Amazon Simple Storage Service User Guide.
 - b. Create a subfolder called original_inventory in this new folder.
 - c. Copy the vault inventory file into this folder. For more information, see <u>Downloading a Vault</u> <u>Inventory in Amazon S3 Glacier</u> in the *Amazon S3 Glacier Developer Guide*.
 - d. Rename the vault inventory file to inventory.csv.
- 3. Sign in to the Systems Manager console.
- 4. Under Shared Resources, select Documents.
- On the Documents page, select Owned by me, and choose the document called Launch-Data-Retrieval-for-Glacier-S3-\$<<u>NAME_OF_STACK</u>>.
- 6. Choose **Execute Automation**.
- 7. Enter the following under Input Parameters on the Execute Automation Runbook page.

Note

To enable the data transfer to a bucket in a different region from the S3 Glacier vault, you must manually adjust the automation document's content by setting the InputPayload parameter allow_cross_region_data_transfer to true. By default, the Guidance restricts cross-region transfers to avoid potential costs.

| Parameter | Value | Notes |
|---|--------------------------------|---|
| AcknowledgeAdditio nalCostForCrossReg ionTransfer | NO | Select YES only if you are aware of the excessive additional cost when selecting a destination bucket in a different region than the S3 Glacier vault. See <u>Amazon S3 Glacier data</u> <u>transfer pricing</u> . |
| ProvidedInventory | Yes | |
| WorkflowRun | <requires input=""></requires> | Provide the name of your workflow run, which becomes the workflow_ run value. |
| Description | <optional input=""></optional> | Provide an extended description for this migration . |
| NamingOverrideFile | <optional input=""></optional> | Provide a presigned URL of the NamingOverride file and the bucket that is storing the file if you want to <u>customize</u> <u>S3 object key names</u> . |
| S3StorageClass | <requires input=""></requires> | Select the S3 storage class for the migrated archives. See Amazon S3 pricing. |

8. Choose **Execute** It takes approximately less than 1 minute to execute the document and launch the transfer. To confirm that the transfer process started successfully, refer to the *runScript* step status under the *Executed Steps* section.

(i) Note

To monitor the progress of the transfer after launching the workflow, please refer to the Guidance's <u>Cloudwatch dashboard</u>. Please note that it might take 5-12 hours before the dashboard **Workflow Run ID** dropdown menu entries get updated with the new entry after launching the transfer.

Step 3: Resume the transfer workflow

- 1. Sign in to the Systems Manager console.
- 2. Under Shared Resources, select Documents.
- On the Documents page, select Owned by me, and choose the document called Resume-Data-Retrieval-for-Glacier-S3-\$<<u>NAME_OF_STACK</u>>.
- 4. Choose **Execute Automation**.
- 5. Enter the following under **Input Parameters** on the **Execute Automation Runbook** page.

| Parameter | Value | Notes |
|---|--------------------------------|---|
| AcknowledgeAdditio nalCostForCrossReg ionTransfer | NO | Select YES only if you are aware of the excessive additional cost when selecting a destination bucket in a different region than the S3 Glacier vault. See <u>Amazon S3 Glacier data</u> transfer pricing. |
| ProvidedInventory | <requires input=""></requires> | Input with two options [YES, NO] indicate if the inventory is provided. |
| WorkflowRun | <requires input=""></requires> | Provide the same workflow_run value used in the initial execution of the transfer workflow. |

| Parameter | Value | Notes |
|--------------------|--------------------------------|--|
| Description | <optional input=""></optional> | Provide an extended description for this migration . |
| NamingOverrideFile | <optional input=""></optional> | Provide a presigned URL of the NamingOverride file and the bucket that is storing the file if you want to <u>customize</u> <u>S3 object key names</u> . |
| S3 Storage class | <requires input=""></requires> | Select the S3 storage class for the migrated archives. See <u>Amazon S3 pricing</u> . |

(i) Note

To monitor the progress of the transfer after launching the workflow, please refer to the Guidance's <u>Cloudwatch dashboard</u>. Please note that it might take 5-12 hours before the dashboard **Workflow Run ID** dropdown menu entries get updated with the new entry after launching the transfer.

Monitor the Guidance with Service Catalog AppRegistry

This Guidance includes a Service Catalog AppRegistry resource to register the CloudFormation template and underlying resources as an application in both <u>Service Catalog AppRegistry</u> and <u>AWS</u> <u>Systems Manager Application Manager</u>.

AWS Systems Manager Application Manager gives you an application-level view into this Guidance and its resources so that you can:

- Monitor its resources, costs for the deployed resources across stacks and AWS accounts, and logs associated with this Guidance from a central location.
- View operations data for the resources of this Guidance (such as deployment status, CloudWatch alarms, resource configurations, and operational issues) in the context of an application.

The following figure depicts an example of the application view for the Guidance stack in Application Manager.

| Components (2) | AWS-Systems-Manager-Application-Manager C Start runbook | | | |
|---|---|--|--|--|
| Name Alarms | Application information View in AppRegistry 🖄 | | | |
| AWS-Systems-Manager-Application-Manager AWS-Systems-Manager-A | Application type Name Application monitoring AWS-AppRegistry AWS-Systems-Manager-Application-Manager Application monitoring | | | |
| | Description Service Catalog application to track and manage all your resources for the solution | | | |
| | Overview Resources Instances Compliance Monitoring Opsitems Logs Runbooks Cost | | | |
| | Insights and Alarms Info View all Monitor your application health with Amazon CloudWatch. Cost View resource costs per application using AWS Cost Explorer. | | | |
| | Cost (USD) | | | |

Guidance stack in Application Manager

Activate CloudWatch Application Insights

1. Sign in to the <u>Systems Manager console</u>.

- 2. In the navigation pane, choose **Application Manager**.
- 3. In **Applications**, search for the application name for this Guidance and select it.

The application name will have App Registry in the **Application Source** column, and will have a combination of the Guidance name, Region, account ID, or stack name.

- 4. In the **Components** tree, choose the application stack you want to activate.
- 5. In the Monitoring tab, in Application Insights, select Auto-configure Application Insights.

| Overview Resources Provisioning Com | oliance Monitoring OpsItems Logs Runbooks Cost |
|--|--|
| Application Insights (0) Info Problems detected by severity | O View Ignored Problems Actions ▼ Add an application |
| Q Find problems | Last 7 days C [±] 1 > ∅ |
| Problem su ▼ Status ▼ Se | everity abla Source abla Start time abla Insights abla Start time abl |
| Ad | vanced monitoring is not enabled |
| When you onboard your first application, a service-linker Insights and includes the permiss | d role (SLR) is created in your account. The SLR is predefined by CloudWatch Application ions the service requires to monitor AWS services on your behalf. |
| Aut | o-configure Application Insights |

Monitoring for your applications is now activated and the following status box appears:

| Overview Resources Provisioning Compliance | Monitoring Opsitems Logs F | Runbooks Cost |
|---|--|-------------------------|
| Application Insights (0) Info Problems detected by severity | ○ View Ignored Problems Actions ▼ | Add an application |
| Q. Find problems | Last 7 days 🔻 | C < 1 > © |
| Problem su V Status V Severity V | Source V Start time | |
| Application monitoring has been successfully enabled. It will view results. | ll take some time to display any results. Please use | e the refresh button to |

Confirm cost tags associated with the Guidance

After you activate cost allocation tags associated with the Guidance, you must confirm the cost allocation tags to see the costs for this Guidance. To confirm cost allocation tags:

- 1. Sign in to the Systems Manager console.
- 2. In the navigation pane, choose **Application Manager**.
- 3. In Applications, choose the application name for this Guidance and select it.
- 4. In the **Overview** tab, in **Cost**, select **Add user tag**.

| Cost View resource costs | per application using AWS Cost Explorer. | View all |
|-----------------------------|---|-----------------------------------|
| _ | | |
| To enable cost | tracking, add the "AppManagerCFNStackKey stack. | " user tag to your CloudFormation |
| | Adding the user tag will require redeployn | nent of the stack. |
| | Add user tag | |

5. On the Add user tag page, enter confirm, then select Add user tag.

The activation process can take up to 24 hours to complete and the tag data to appear.

Activate cost allocation tags associated with the Guidance

After you confirm the cost tags associated with this Guidance, you must activate the cost allocation tags to see the costs for this Guidance. The cost allocation tags can only be activated from the management account for the organization.

To activate cost allocation tags:

- 1. Sign in to the AWS Billing and Cost Management and Cost Management console.
- 2. In the navigation pane, select **Cost Allocation Tags**.
- 3. On the **Cost allocation tags** page, filter for the AppManagerCFNStackKey tag, then select the tag from the results shown.
- 4. Choose Activate.

AWS Cost Explorer

You can see the overview of the costs associated with the application and application components within the Application Manager console through integration with AWS Cost Explorer. Cost Explorer helps you manage costs by providing a view of your AWS resource costs and usage over time.

- 1. Sign in to the AWS Cost Management console.
- 2. In the navigation menu, select **Cost Explorer** to view the Guidance's costs and usage over time.

Update the Guidance

If you have previously deployed the Guidance, follow this procedure to update the CloudFormation stack to get the latest version of the Guidance's framework.

- 1. Sign in to the <u>AWS CloudFormation console</u>, select your existing Data Transfer from Amazon S3 Glacier Vaults to Amazon S3 CloudFormation stack, and select **Update**.
- 2. Select Replace current template.
- 3. Under Specify template:
 - a. Select Amazon S3 URL.
 - b. Copy the link of the latest template.
 - c. Paste the link in the Amazon S3 URL box.
 - d. Verify that the correct template URL shows in the **Amazon S3 URL** text box, and choose **Next**. Choose **Next** again.
- 4. Under **Parameters**, review the parameters for the template and modify them as necessary. For details about the parameters, see <u>Step 1: Launch the stack</u>.
- 5. Choose Next.
- 6. On the **Configure stack options** page, choose **Next**.
- 7. On the **Review** page, review and confirm the settings. Select the box acknowledging that the template creates IAM resources.
- 8. Choose **View change set** and verify the changes.
- 9. Choose **Update stack** to deploy the stack.

You can view the status of the stack in the AWS CloudFormation console in the **Status** column. You should receive an UPDATE_COMPLETE status in approximately 5-10 minutes.

Troubleshooting

This section provides troubleshooting instructions for deploying and using the Guidance.

Problem: Transfer workflow has not progressed after 14 hours

Your transfer workflow has been launched for more than 14 hours, but the **Downloaded** count on the CloudWatch dashboard has not increased.

Resolution

- 1. Sign in to the Step Functions console.
- 2. Under **State machines**, select the state machine called OrchestratorStateMachine\$CFN_ID and choose **View Details**.
- 3. Select the most recent execution and choose View details.
- 4. Note failures. If the workflow is still running the InventoryRetrieval workflow, there might be an issue with the Amazon S3 Glacier service generating the inventory file. Contact <u>Support</u> if you have an AWS Developer Support plan or above.

Problem: Transfer workflow must be stopped

Your transfer workflow is ongoing, but you want to stop it.

Resolution

- 1. Follow steps 1–3 in Problem: Transfer workflow has not progressed after 14 hours.
- 2. Under Actions, choose Stop execution.
- 3. If you plan to <u>resume the transfer workflow later</u>, find the workflow_run value from the **Execution Input and Output** tab on this page. You need this value to resume the workflow.

i Note

After you stop the execution, no new archives are requested from the Amazon S3 Glacier service. The archives that were already requested will download. These downloads take 4–8 hours to complete.

Uninstall the Guidance

You can uninstall the Data Transfer from Amazon S3 Glacier Vaults to Amazon S3 Guidance from the AWS Management Console or by using the AWS Command Line Interface. Manually delete the following resources:

- S3 buckets (other than the output bucket if you intend to keep the transferred archives)
- DynamoDB tables
- CloudWatch Logs

This Guidance does not automatically delete these resources in case you have stored data to retain.

Using the AWS Management Console

- 1. Sign in to the <u>CloudFormation console</u>.
- 2. On the Stacks page, select this Guidance's installation stack.
- 3. Choose Delete.

Using AWS Command Line Interface

Determine whether the AWS Command Line Interface (AWS CLI) is available in your environment. For installation instructions, see <u>What Is the AWS Command Line Interface</u> in the AWS CLI User Guide. After confirming that the AWS CLI is available, run the following command.

\$ aws cloudformation delete-stack --stack-name <installation-stack-name>

Deleting the S3 buckets

This Guidance is configured to retain the Guidance-created S3 buckets if you decide to delete the CloudFormation stack, to prevent accidental data loss. After uninstalling the Guidance, you can manually delete the S3 buckets if you don't need to retain the data. Follow these steps to delete the S3 buckets.

1. Sign in to the <u>Amazon S3 console</u>.

- 2. Choose Buckets from the left navigation pane.
- 3. Locate the <*stack-name*> S3 buckets.
- 4. Select each S3 bucket and choose Empty.
- 5. Select each S3 bucket and choose Delete.

To delete the S3 bucket using the AWS CLI, run the following command:

```
$ aws s3 rb s3://<bucket-name> --force
```

Deleting the DynamoDB tables

This Guidance is configured to retain the Guidance-created DynamoDB tables if you decide to delete the CloudFormation stack, to prevent accidental data loss.

- 1. Sign in to the DynamoDB console.
- 2. Choose **Tables** from the left navigation pane.
- 3. Locate the <<u>stack-name</u>> DynamoDB tables.
- 4. Select each DynamoDB table and choose Delete.
- 5. Confirm the deletion.

To delete the DynamoDB table using the AWS CLI, run the following command:

\$ aws dynamodb delete-table --table-name <table.name>

Deleting the CloudWatch Logs

- 1. Sign in to the CloudWatch console.
- 2. Choose Log groups from the left navigation pane.
- 3. Locate the <<u>stack-name</u>> CloudWatch Logs.
- 4. Select each CloudWatch Log and choose Actions, then Delete log groups.
- 5. Confirm the deletion.

To delete the CloudWatch Logs using the AWS CLI, run the following command:

\$ aws logs delete-log-group --log-group-name <log-group-name>

Use the Guidance

This section provides a user guide for using the AWS Guidance.

Validate your inventory

After you deploy this Guidance and transfer your archives, you can validate your Amazon S3 inventory to confirm whether all of your archives transferred. We recommend confirming that all of your archives transferred before deleting your original archive. See <u>Amazon S3 Inventory</u> for more information.

Access the CloudWatch dashboard

- 1. Sign in to the <u>CloudWatch console</u>.
- 2. Choose **Dashboards** from the left navigation pane.
- 3. Select the dashboard that starts with Data-Transfer-from-Amazon-S3-Glacier-to-Amazon-S3-Dashboard-.
- 4. In the dashboard, you can find:
 - The total number of archives in the inventory file, and their collective size.
 - The total number of archives that you requested for download, and their collective size.
 - The total number of archives that are staged by the Amazon S3 Glacier service for download, and their collective size.
 - The total number of archives that won't be downloaded because they're larger than 5 GB in size.
 - The total number of downloaded archives.
- 5. When the number of downloaded archives matches the total number of requested archives, the transfer is complete.

Manage your Amazon S3 storage

After you transfer your Amazon S3 Glacier data to the Amazon S3 service, you can change your storage classes to fit your use cases. For information about how to do this, see the following resources:

- Managing your Amazon S3 storage in the Amazon Simple Storage Service User Guide
- Using Amazon S3 storage classes in the Amazon Simple Storage Service User Guide

Developer guide

This section provides the source code for the Guidance.

Source code

Visit our <u>GitHub repository</u> to download the source files for this Guidance and to share your customizations with others.

The AWS CDK generates the Data Transfer from Amazon S3 Glacier Vaults to Amazon S3 templates. See the README.md file for additional information.

Reference

This section includes information about an optional feature for collecting unique metrics for this Guidance and a list of builders who contributed to this Guidance.

Anonymized data collection

This Guidance includes an option to send anonymized operational metrics to AWS. We use this data to better understand how customers use this Guidance and related services and products. When invoked, the following information is collected and sent to AWS:

- Guidance ID The Guidance identifier
- Unique ID (UUID) Randomly generated, unique identifier for each Data Transfer from Amazon
 S3 Glacier Vaults to Amazon S3 deployment
- Timestamp Data-collection timestamp
- Data The Guidance collects and sends the following statistics:
 - Region Selected AWS Region
 - Version Version of the Guidance deployment
 - StorageClass Selected destination S3 storage class
 - RetrievalTier Type of Amazon S3 Glacier retrieval (for example, Bulk)
 - VaultSize Source S3 Glacier vault size
 - ArchiveCount Number of S3 Glacier archives

AWS owns the data gathered though this survey. Data collection is subject to the <u>Privacy Notice</u>. To opt out of this feature, complete the following steps before launching the AWS CloudFormation template.

- 1. Download the AWS CloudFormation template to your local hard drive.
- 2. Open the AWS CloudFormation template with a text editor.
- 3. Modify the Environment variable section for the SendAnonymizedStats Lambda function from:

"SEND_ANONYMIZED_STATS": "Yes"

to:

"SEND_ANONYMIZED_STATS": "No"

- 4. Sign in to the AWS CloudFormation console.
- 5. Select Create stack.
- 6. On the Create stack page, Specify template section, select Upload a template file.
- 7. Under **Upload a template file**, choose **Choose file** and select the edited template from your local drive.
- 8. Choose **Next** and follow the steps in <u>Launch the stack</u> in the Deploy the Guidance section of this guide.

Contributors

- Bassem Wanis
- Kamyar Ziabari
- Nathaniel Schaaf
- Garvit Singh
- Simon Krol
- Evgeny Minkevich

Revisions

Publication date: December 2023

Check the <u>CHANGELOG.md</u> file in the GitHub repository to see all notable changes and updates to the software. The changelog provides a clear record of improvements and fixes for each version.

Notices

Customers are responsible for making their own independent assessment of the information in this document. This document: (a) is for informational purposes only, (b) represents AWS current product offerings and practices, which are subject to change without notice, and (c) does not create any commitments or assurances from AWS and its affiliates, suppliers, or licensors. AWS products or services are provided "as is" without warranties, representations, or conditions of any kind, whether express or implied. AWS responsibilities and liabilities to its customers are controlled by AWS agreements, and this document is not part of, nor does it modify, any agreement between AWS and its customers.

Data Transfer from Amazon S3 Glacier Vaults to Amazon S3 is licensed under the terms of theApache License, Version 2.0.