
AWS Data Exchange Subscriber Coordinator Implementation Guide



AWS Data Exchange Subscriber Coordinator: Implementation Guide

Copyright © 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

Welcome	1
Overview	2
Cost	2
Architecture overview	2
Components	4
Considerations	5
Amazon S3 bucket	5
AWS Lambda function	5
AWS Data Exchange product subscription	5
Permissions	5
Security	6
AWS CloudFormation template	7
Automated deployment	8
Prerequisites	8
Launch the stack	8
New data set revision	10
Additional resources	11
Uninstall the solution	12
Using the AWS Management Console	12
Using AWS Command Line Interface	12
Operational metrics	13
Source code	14
Contributors	15
Revisions	16
Notices	17

Automatically receive new AWS Data Exchange dataset revisions and initiate downstream processing when new data becomes available from a data provider

Publication date: *December 2020*

This implementation guide describes architectural considerations and provides configuration steps for deploying AWS Data Exchange Subscriber Coordinator in the Amazon Web Services (AWS) Cloud. It includes links to an [AWS CloudFormation](#) template that launches and configures the AWS services required to deploy this solution using AWS best practices for security and availability.

The guide is intended for IT infrastructure architects, administrators, and DevOps professionals who have practical experience architecting in the AWS Cloud.

Overview

This solution is for [AWS Data Exchange](#) customers who [subscribe to data products](#) and want to automate their processes. It automates both receiving new dataset revisions and initiating downstream processing when new data becomes available from a data provider. Through seamless integration with [Amazon CloudWatch](#) and [Amazon Simple Storage Service](#) (Amazon S3), you can automatically download the latest revisions of the datasets when they are released and begin processing that data without delay.

This guide provides infrastructure and configuration information for planning and deploying the AWS Data Exchange Subscriber Coordinator in the AWS Cloud.

Cost

You are responsible for the cost of the AWS services used while running this solution. As of the date of publication, the estimated cost for running this solution with the four Lambda functions and the one Step Functions associated with this template using the default settings in the US East (N. Virginia) Region is **\$0.01 per month** for 10 subscriptions per day for 30 days. This calculation assumes that there is no free-tier consumption available in the account.

AWS service	Quantity	Cost
AWS Lambda	1500 runs in a month at 3000 ms	\$0.01
AWS Step Functions	300 runs in a month	\$0.00002
Total cost		\$0.01

This cost estimate does not include costs associated with Amazon S3 storage or AWS Data Exchange usage. Prices are subject to change. For full details, refer to the pricing webpage for each AWS service used in this solution.

Architecture overview

Deploying this solution with the default parameters builds the following environment in the AWS Cloud.

AWS Data Exchange Subscriber
Coordinator Implementation Guide
Architecture overview

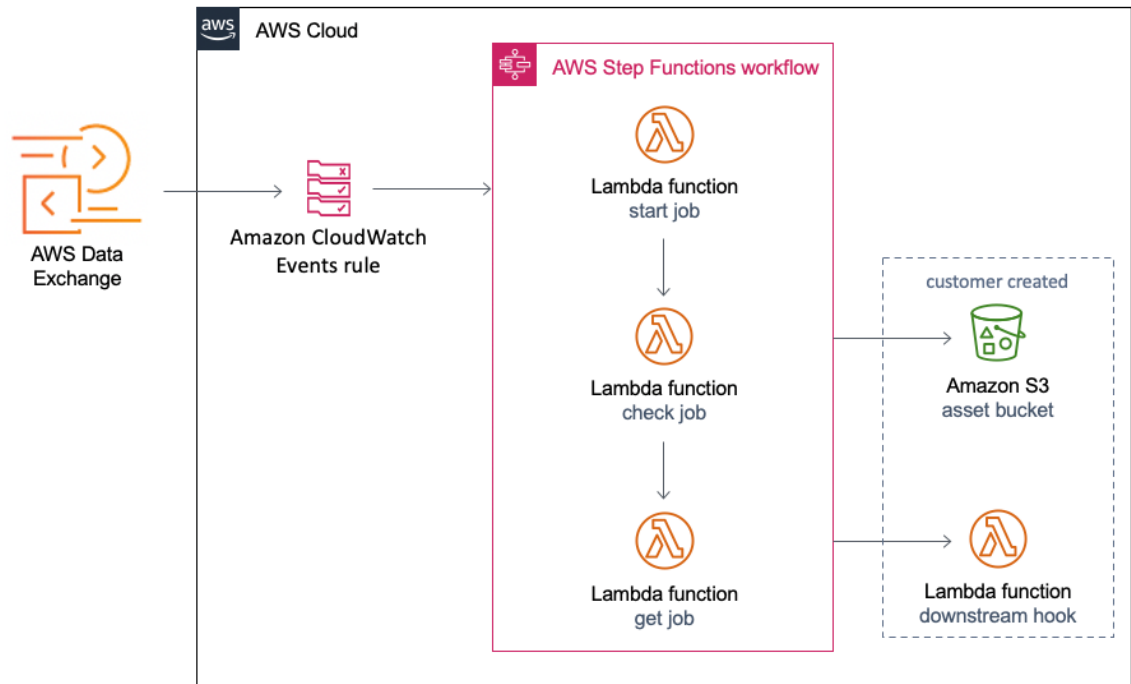


Figure 1: AWS Data Exchange Subscriber Coordinator architecture diagram

The AWS CloudFormation template deploys all of the serverless infrastructure necessary to coordinate the download steps required for AWS Data Exchange. This includes the following:

- Three [AWS Lambda](#) functions to start an AWS Data Exchange job, periodically check its progress, and retrieve the job for the downstream hook function.
- An [AWS Step Functions](#) to orchestrate the AWS functions that download new revisions to the AWS Data Exchange dataset.
- An [Amazon CloudWatch Events](#) rule to identify the publishing of any new revisions by publishers.
- [AWS Identity and Access Management](#) (IAM) roles that provide an AWS Step Functions and three associated Lambda functions with permissions to initiate their API calls.

This solution does not deploy any AWS Data Exchange resources. Also, the Amazon S3 asset storage bucket and the downstream Lambda hook function are not deployed by the template. You must create these separately.

Components

When a publisher creates a new dataset revision, the AWS Data Exchange Subscriber Coordinator initializes the following:

- It first creates a job to receive the new data in the specified Amazon S3 bucket.
- After all the assets are delivered, it invokes the specified AWS Lambda function. You can configure this Lambda function to initiate downstream processing on the new data or send a notification to interested parties.

Deployment considerations

Amazon S3 bucket

Before launching the AWS CloudFormation template, you must have an Amazon S3 bucket to store the assets downloaded from AWS Data Exchange. This bucket must be in the same AWS Region in which the CloudFormation template is deployed. Use this bucket to complete the `SubscriptionBucket` parameter in the CloudFormation template.

AWS Lambda function

When you deploy the AWS Data Exchange Subscriber Coordinator, you must provide a Lambda function ARN as a parameter. That Lambda function is initiated when the download of assets is completed. Example use cases might include alerting a system administrator that a new revision has been downloaded, or converting the original raw format to an optimized format ready for consumption. For testing purposes, a simple *Hello World* example runs successfully.

Find the Lambda function ARN in the AWS Lambda console. Choose the function and the ARN appears at the top right of the page.

AWS Data Exchange product subscription

To download file-based data products through the AWS Marketplace using AWS Data Exchange, you must subscribe to a product. For details, refer to [Subscribing to Data Products on AWS Data Exchange](#) in the *AWS Data Exchange User Guide*. During deployment, provide the dataset ID in the required parameter.

Permissions

To deploy this solution, you must have an IAM user role with permissions to launch CloudFormation templates that create Amazon CloudWatch Events rules, AWS Step Functions, Lambda functions, and set the associated permissions.

Security

When you build systems on AWS infrastructure, security responsibilities are shared between you and AWS. This shared model can reduce your operational burden as AWS operates, manages, and controls the components from the host operating system and virtualization layer down to the physical security of the facilities in which the services operate. For more information about security on AWS, visit [AWS Cloud Security](#).

AWS CloudFormation template

To automate the deployment, this solution uses AWS CloudFormation. It includes the following CloudFormation template, which you can download before deployment:

A rectangular button with a light orange background and a thin black border. The text "View Template" is centered in a dark blue, sans-serif font. The word "View" is on the top line and "Template" is on the bottom line.

View
Template

aws-data-exchange-subscriber-coordinator.template: Use this template to launch AWS Data Exchange Subscriber Coordinator and all associated components. The template deploys an Amazon CloudWatch Events rule, AWS Lambda functions, AWS Step Functions, and associated AWS Identity and Access Management (IAM) roles. You can customize the template based on your specific needs.

Automated deployment

Before you launch the solution, review the architecture, configuration, network security, and other considerations discussed in this guide. Follow the step-by-step instructions in this section to configure and deploy the AWS Data Exchange Subscriber Coordinator into your account.

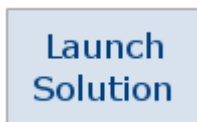
Time to deploy: Approximately five minutes

Prerequisites

To launch, the template requires an Amazon S3 bucket to store retrieved assets, a subscription prefix, and a downstream Lambda ARN for processing hooks when assets have been downloaded. Refer to [Design considerations \(p. 5\)](#) for details.

Launch the stack

1. Sign in to the AWS Management Console and select the button below to launch the in this guide to launch the aws-data-exchange-subscriber-coordinator AWS CloudFormation template.



Alternatively, you can [download the template](#) as a starting point for your own implementation.

2. The template launches in the US East (N. Virginia) Region by default. To launch the solution in a different AWS Region, use the Region selector in the console navigation bar.
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 solution stack. For information about naming character limitations, refer to [IAM and STS quotas](#) in the *AWS Identity and Access Management User Guide*.
5. Under **Parameters**, review the parameters for this solution template and modify them as necessary. This solution uses the following default values.

Parameter	Default	Description
SubscriptionBucket	<Requires input>	An existing Amazon S3 bucket used to store datasets retrieved by the AWS Data Exchange Subscriber Coordinator CloudFormation template.
DataSetId	<Requires input>	The dataset ID of the AWS Data Exchange product that you want to automatically retrieve.

AWS Data Exchange Subscriber
Coordinator Implementation Guide
Launch the stack

Parameter	Default	Description
SubscriptionPrefix	<Requires input>	A prefix to add to the retrieved datasets to enable multiple AWS Data Exchange Subscriber Coordinator deployments with a single Amazon S3 bucket output.
DownstreamProcessingLambda	<Requires input>	The ARN of an existing Lambda function used as a downstream hook for processing the retrieved datasets or notification of download.
LoggingLevel	INFO	Specifies the level of information detail to log in CloudWatch Logs when running. Other options include: DEBUG, WARNING, ERROR, and CRITICAL.

6. Choose **Next**.
7. On the **Configure stack options** page choose **Next**.
8. On the **Review** page, review and confirm the settings. Check the boxes acknowledging that the template will create AWS Identity and Access Management (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 five minutes.

New data set revision

When an AWS Data Exchange publisher creates a new revision for the specified dataset, the stack runs the subscription automation steps. This includes copying all of the new revision assets to the subscription bucket, then running the downstream processing Lambda function.

Additional resources

AWS services

- | | |
|--|--|
| <ul style="list-style-type: none">• Amazon Simple Storage Service• AWS CloudFormation• Amazon CloudWatch• AWS Data Exchange | <ul style="list-style-type: none">• AWS Identity and Access Management• AWS Lambda• AWS Step Functions |
|--|--|

Related AWS Solutions Implementation

Publishers can use the [AWS Data Exchange Publisher Coordinator](#) solution to automate the publication of new revisions from their Amazon S3 buckets to AWS Data Exchange.

Uninstall the solution

To uninstall the AWS Data Exchange Subscriber Coordinator solution, you must delete the CloudFormation stack. Deleting the solution does not delete any assets (Amazon S3 objects) received from AWS Data Exchange as a result of the solution.

Using the AWS Management Console

1. Sign in to the [AWS CloudFormation console](#).
2. Select this solution'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, refer to [What Is the AWS Command Line Interface](#) in the *AWS CLI User Guide*. After confirming that the AWS CLI is available, run the following command substituting `<installation-stack-name>` with the name of the solution you used at installation.

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

Collection of operational metrics

This solution includes an option to send anonymous operational metrics to AWS. We use this data to better understand how customers use this solution and related services and products. When enabled, the following information is collected and sent to AWS when the CloudFormation stack is created, runs, or deleted:

1. **Solution ID:** The AWS solution identifier
2. **Unique ID (UUID):** Randomly generated, unique identifier for each AWS Data Exchange Subscriber Coordinator deployment
3. **Timestamp:** Data-collection timestamp
4. **Version:** The version number of the solution
5. **AssetCount:** The number of assets received

AWS owns the data gathered through this survey. Data collection is subject to the [AWS Privacy Policy](#). To opt out of this feature, complete the following task.

Modify the AWS CloudFormation template mapping section from:

```
"Send" : {  
  "AnonymousUsage" : { "Data" : "Yes" }  
},
```

to:

```
"Send" : {  
  "AnonymousUsage" : { "Data" : "No" }  
},
```


Source code

Visit our [GitHub repository](#) to download the templates and scripts for this solution, and to share your customizations with others.

Contributors

The following individuals contributed to this document:

- Chris Marshall
- Colin Marden

Revisions

Date	Change
December 2020	<i>Initial release</i>

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 current AWS 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. The responsibilities and liabilities of AWS 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.

AWS Data Exchange Subscriber Coordinator is licensed under the terms of the MIT No Attribution at [The Software Package Data Exchange](#).