Cross-Region Replication Monitor

AWS Implementation Guide

Vijay Satish

Mike O'Brien

June 2017

Contents

About This Guide

This implementation guide discusses architectural considerations and configuration steps for deploying the Cross-Region Replication Monitor (CRR Monitor) solution on the Amazon Web Services (AWS) Cloud. It includes links to AWS CloudFormation templates that launch, configure, and run the AWS compute, network, storage, and other services required to deploy this solution on AWS, using AWS best practices for security and availability.

The guide is intended for IT infrastructure architects who have a working knowledge of data replication and practical experience architecting on the AWS Cloud.

Overview

Amazon Simple Storage Service (Amazon S3) offers cross-region replication, a bucket-level feature that enables automatic, asynchronous copying of objects across buckets in different AWS Regions. This feature can help companies minimize latency when accessing objects in

different geographic regions, meet compliance requirements, and for operational purposes. Amazon S3 encrypts all data in transit across AWS Regions using SSL, and objects in the destination bucket are exact replicas of objects in the source bucket. For more information on cross-region replication, see the Amazon S3 Developer Guide.

Currently, AWS customers can retrieve the replication status of their objects manually or use an Amazon S3 inventory to generate metrics on a daily or weekly basis. To help customers more proactively monitor the replication status of their Amazon S3 objects, AWS offers the Cross-Region Replication Monitor (CRR Monitor) solution. The CRR Monitor automatically checks the replication status of Amazon S3 objects across all AWS Regions in a customers' account, providing near real-time metrics as well as failure notifications to help customers proactively identify failures and troubleshoot problems. The solution automatically provisions the necessary AWS services to monitor and view replication status, including AWS Lambda, Amazon CloudWatch, Amazon Simple Notification Service (Amazon SNS), AWS CloudTrail, Amazon Simple Queue Service (Amazon SQS), and Amazon DynamoDB, and offers an option to use Amazon Kinesis Firehose to archive replication metadata in Amazon S3.

This guide assumes basic knowledge of Amazon S3 cross-region replication. It is also helpful to have working knowledge of Amazon S3, AWS Lambda, Amazon CloudWatch, Amazon SQS, Amazon SNS, and Amazon DynamoDB.

Cost

You are responsible for the cost of the AWS services used while running the CRR Monitor. The total cost for running this solution depends on the interval at which you run the AWS Lambda functions.

As of the date of publication, the cost for running this solution with default settings in the US East (N. Virginia) Region is approximately **$0.35 an hour**, or less if you have AWS Lambda free tier[1] monthly usage credit. This cost estimate assumes one million Lambda requests, and 100,000 AWS CloudTrail events.

> **Note:** Pricing does not include the cost for Amazon DynamoDB. These charges will vary depending on the data volume being processed. (e.g. for a little less than $0.25/day ($7.50/month), you could support an application that performs 1 million writes and reads per day, 100K read requests from Streams, and stores 1 GB of data). For more information, see Amazon DynamoDB pricing.

[1] https://aws.amazon.com/lambda/pricing/

This pricing is subject to change and does not reflect variable charges for Amazon S3 data storage and replication, and the Amazon Athena query service. For full details, see the pricing webpage for each AWS service you will be using in this solution.

Architecture Overview

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

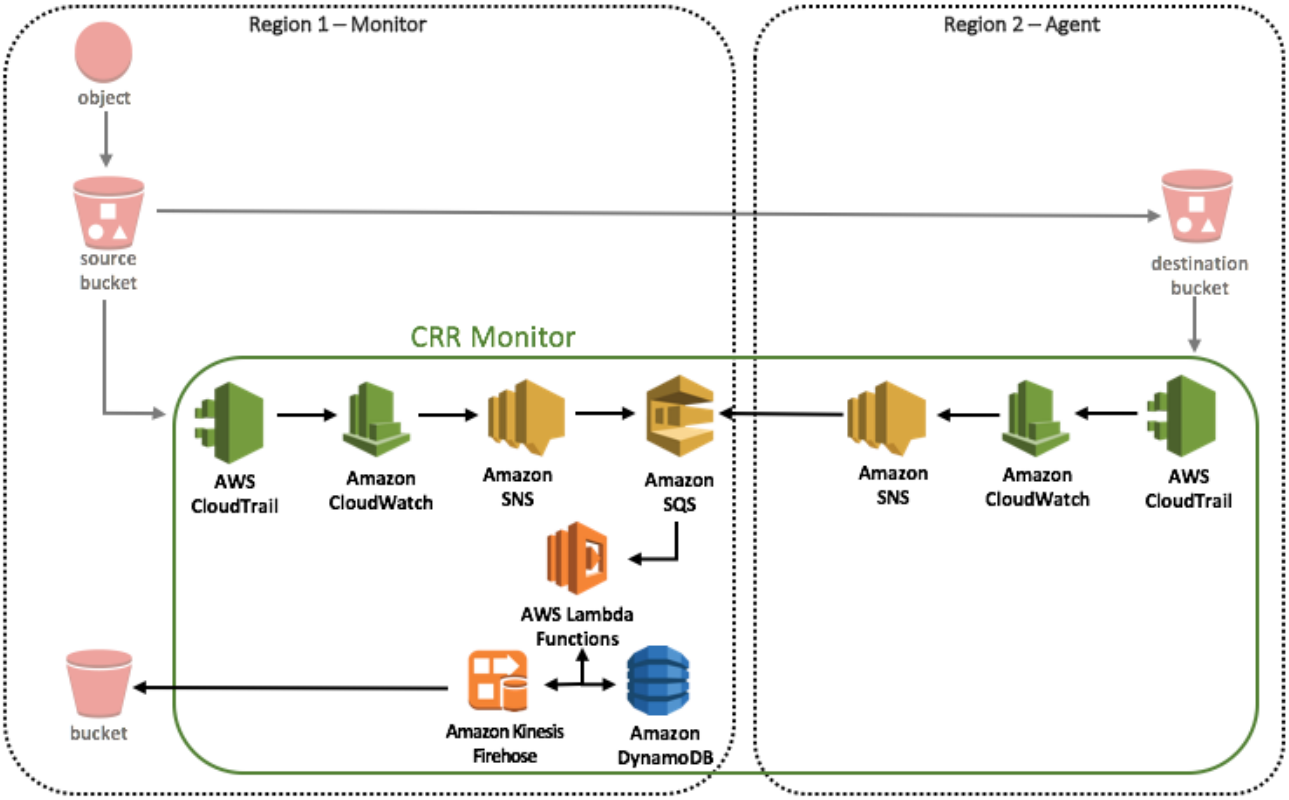

Figure 1: Cross-Region Replication Monitor architecture on AWS

The primary AWS CloudFormation template deploys all solution components in the AWS Region that hosts the source Amazon S3 bucket(s) (labeled the *Monitor* region in the diagram). These include AWS Identity and Access Management (IAM) roles, AWS Lambda functions, an AWS CloudTrail trail, an Amazon CloudWatch event, an Amazon SNS topic, and an Amazon DynamoDB table. The solution turns on AWS CloudTrail and automatically enables the data events for the source and destination buckets that have CRR enabled.

When an object is added to a source S3 bucket in the Monitor region, AWS CloudTrail logs the data event; this activity triggers an Amazon CloudWatch alarm that delivers the status

information to an Amazon SNS topic. Amazon SNS then sends the data to an Amazon SQS queue. When a replicated object is added to a destination bucket, it triggers a similar event in the Agent region, sending status information back to the Amazon SQS queue in the Monitor region. The AWS Lambda functions process the data in the queue to verify that the object was replicated successfully. If the object replication fails, an AWS Lambda function logs the event in CloudWatch, triggering a CloudWatch alarm that sends an Amazon SNS notification to the subscriber.

Once the solution's Lambda function verifies an object was successfully replicated, it stores the data in an Amazon DynamoDB table for immediate access. Status data in the DynamoDB table is deleted and replaced every 24 hours. Customers who deploy this solution in an AWS Region that offers Amazon Kinesis Firehose can choose to archive solution data to Amazon S3. If you enable this feature, the solution uses a Firehose delivery stream to upload data to one of your existing S3 buckets for later analysis. You can use Amazon Athena, a serverless, interactive query service, to easily analyze historical data in Amazon S3.

Implementation Considerations

Regional Deployments

Customers can deploy the CRR Monitor in any AWS Region that supports AWS Lambda and Amazon Kinesis Firehose.[2] The solution uses a Firehose delivery stream as part of an optional solution feature to archive replication data in Amazon Simple Storage Service (Amazon S3). You can choose to disable this feature (due to regional availability or other reasons) during initial configuration of the AWS CloudFormation template. If you disable data archiving, you can use the Amazon S3 inventory feature to get a daily or weekly replication status for your objects, but it will not include detailed metadata from this solution, such as CRR rate, elapsed time, etc.

Once deployed, the CRR Monitor applies the appropriate configuration for monitoring the replicated Amazon S3 buckets across all AWS Regions in a single account. For information, general requirements for cross-region replication, see the Amazon S3 Developer Guide.

[2] For the most current AWS Lambda, and Amazon Kinesis Firehouse availability by region, see https://aws.amazon.com/about-aws/global-infrastructure/regional-product-services/

Scaling

This solution will run multiple concurrent instances of the AWS Lambda CRR Monitor function, based on workload, up to a default maximum of 20 per minute. Each instance can process 1,800 S3 PUT records in its five-minute maximum run time. This allows CRR Monitor to achieve scale within predictable and controllable bounds to limit impact on other Lambda workloads in your account.

The solution provisions Amazon DynamoDB throughput capacity to support the maximum read and write requests from the default limit (20) of Lambda functions. If you plan to monitor a large number of objects and anticipate a high number of concurrent requests, consider increasing both the Lambda and Amazon DynamoDB settings in this solution.

AWS CloudFormation Template

This solution uses AWS CloudFormation to automate the deployment of the Cross-Region Replication Monitor. It includes the following AWS CloudFormation template, which you can download before deployment:

View template

crr-monitor.template: Use this template to launch the CRR Monitor solution. This template launches the following components in the Monitor region: AWS Identity and Access Management (IAM) roles, AWS Lambda functions, an Amazon CloudWatch event, an Amazon SNS topic, Amazon SQS, an Amazon DynamoDB table, and an Amazon S3 bucket to host AWS CloudTrail logs in the AWS Region that hosts *source* Amazon S3 buckets. The template also creates an Amazon CloudWatch event, and an Amazon SNS topic in the Agent region. You can customize the template based on your specific needs.

Automated Deployment

This solution is intended for customers who have already configured Amazon S3 cross-region replication in their account. Before you launch the automated deployment, please review the architecture, configuration, and other considerations discussed in this guide. Follow the step-by-step instructions in this section to configure and deploy the Cross-Region Replication Monitor into your account.

Time to deploy: Approximately five minutes

What We'll Cover

The procedure for deploying this architecture on AWS consists of the following steps. For detailed instructions, follow the links for each step.

Step 1. Launch the Stack

- Launch the AWS CloudFormation template into your AWS account.
- Enter values for optional parameters: **Archive to S3**, **S3 Archive Bucket**
- Review the other template parameters, and adjust if necessary.

Step 2. Subscribe to Amazon SNS Topic

- Subscribe to the custom Amazon SNS Topic to receive failure notifications.

Step 1. Launch the Stack

Note: You are responsible for the cost of the AWS services used while running this solution. See the Cost section for more details. For full details, see the pricing webpage for each AWS service you will be using in this solution.

1. Sign in to the AWS Management Console and click the button to the right to launch the `crr-monitor` AWS CloudFormation template.
   You can also download the template as a starting point for your own implementation.

2. The template is launched in the US East (N. Virginia) Region by default. To launch the CRR Monitor in a different AWS Region, use the region selector in the console navigation bar.

Note: This solution uses the AWS Lambda service and Amazon Kinesis Firehose, which is currently available in specific AWS Regions only. Therefore, you must launch this solution an AWS Region where Lambda and Kinesis are available. [3]

3. On the **Select Template** page, verify that you selected the correct template and choose **Next**.
4. On the **Specify Details** page, assign a name to your CRR Monitor stack.
5. Under **Parameters**, review the parameters for the template and modify them as necessary. This solution uses the following default values.

[3] For the most current AWS Lambda and Amazon Kinesis Firehose availability by region, see https://aws.amazon.com/about-aws/global-infrastructure/regional-product-services/

Parameter	Default	Description
Archive to S3	`No`	This solution has the option to archive status data from DynamoDB to Amazon S3 for later analysis. To enable this feature, select `Yes`. **Note:** If you use this feature, you must deploy this template in an AWS Region that supports Amazon Kinesis Firehose.
S3 Archive Bucket	<Requires input>	If you chose to enable data archiving to Amazon S3, enter the name of an existing S3 bucket. **Note**: To use this feature, you must select `Yes` for the **Archive to S3** parameter. You must specify an existing S3 bucket. If you plan to use multiple implementations of the solution in different AWS Regions, we recommend that you use the same bucket to collect all solution data.
Send Anonymous Usage Data	`Yes`	Send anonymous data to AWS to help us understand CRR Monitor usage across our customer base as a whole. To opt out of this feature, select `No`. For more information, see the appendix.

6. Choose **Next.**
7. On the **Options** page, choose **Next.**
8. On the **Review** page, review and confirm the settings. Be sure to check the box acknowledging that the template will create AWS Identity and Access Management (IAM) resources.
9. Choose **Create** to deploy the stack.

   You can view the status of the stack in the AWS CloudFormation Console in the **Status** column. You should see a status of CREATE_COMPLETE in approximately five minutes.

Note: In addition to the solution's AWS Lambda functions, this solution includes the `solution-helper` Lambda function, which runs only during initial configuration or when resources are updated or deleted.

When running this solution, you will see the regularly active Lambda functions whose names contain `crr`. However, do not delete the `solution-helper` function as it is necessary to manage associated resources.

Step 2. Subscribe to Amazon SNS Topic

This solution uses the FailedReplication Amazon CloudWatch metric to trigger a CloudWatch alarm. If an object fails to replicate across AWS Regions, the CloudWatch alarm will trigger an Amazon SNS notification about the failure. This enables customers to identify failures in near real-time and troubleshoot them immediately.

To receive this notification, you must subscribe to the solution's custom Amazon SNS topic: `CRRMonitorMetricsTopic`. For detailed instructions, see Subscribe to a Topic in the Amazon SNS Developer Guide.

Additional Resources

AWS services

- Amazon S3 Cross-Region Replication
- Amazon SQS
- AWS CloudTrail
- Amazon CloudWatch
- AWS CloudFormation
- AWS Lambda
- Amazon Dynamic DynamoDB
- Amazon S3
- Amazon SNS
- Amazon Athena

Appendix: Collection of Anonymous Data

This solution includes an option to send anonymous usage data to AWS. We use this data to better understand how customers use this solution to improve the services and products that we offer. When enabled, the following information is collected and sent to AWS each time CRR Monitor Lambda function runs:

- **Solution ID:** The AWS solution identifier
- **Unique ID (UUID):** Randomly generated, unique identifier for each CRR Monitor deployment
- **Objects:** Total number of objects processed by CRR Monitor
- **Size:** Total size of objects processed by CRR Monitor

Note that AWS will own the data gathered via this survey. Data collection will be subject to the AWS Privacy Policy. To opt out of this feature, set the **Send Anonymous Usage Data** parameter to No.

Send Us Feedback

We welcome your questions and comments. Please post your feedback on the AWS Solutions Discussion Forum.

You can visit our GitHub repository to download the templates and scripts for this solution, and to share your customizations with others.

Document Revisions

Date	Change	In sections
June 2017	Initial Release	--