

使用DataSync 同步GCS数据到Amazon S3

背景

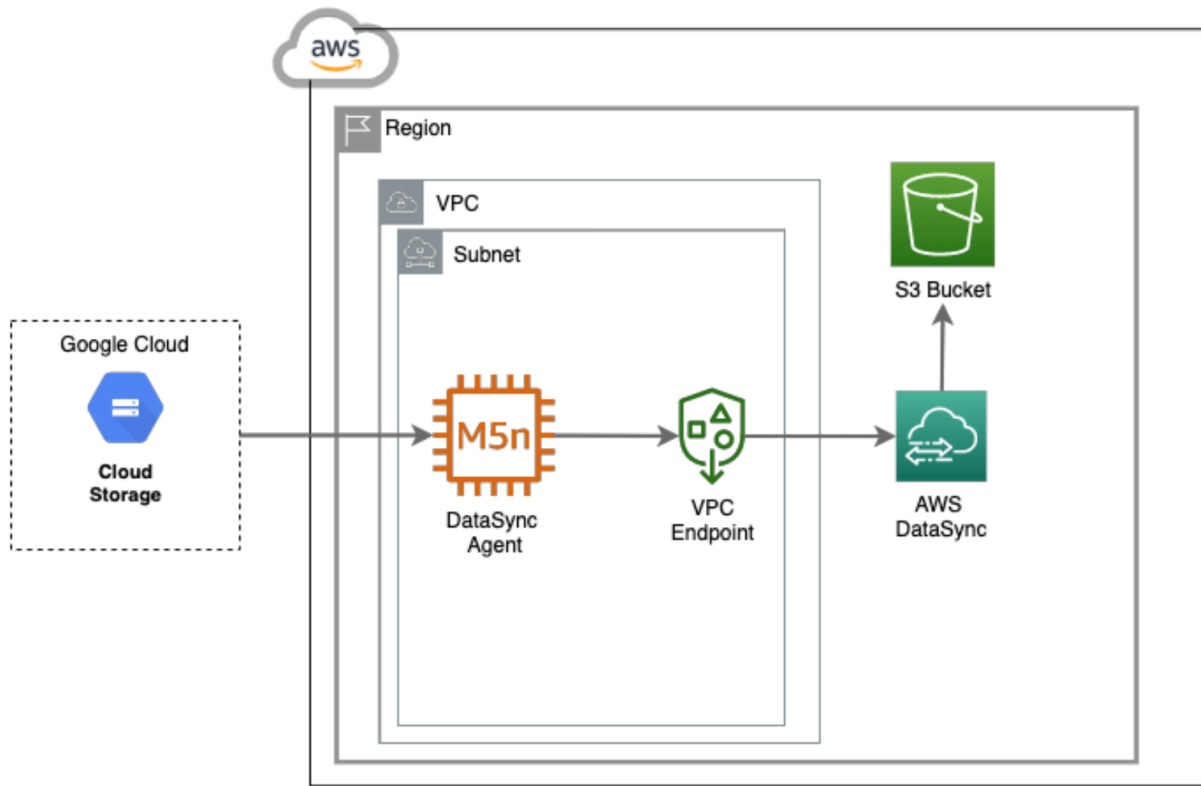
客户需要将GCP的Cloud Storage 的存量和增量数据和同步到Amazon S3做数据分析,但没有一个好用的数据迁移工具,rsync需要写脚本和定期人工检查,工作量大,且无法保证质量,因此想测试一下DataSync。

AWS DataSync 能够与 S3 兼容的对象存储传输文件。Google Cloud (GCP) 使用其互操作性 XML API 支持其 Google Cloud Storage (GCS) 产品的 S3 API。本文档中的过程将指导您完成将数据从 GCS 存储桶传输到 AWS 上的 S3 存储桶的过程。

需求

- 全量数据的迁移;
- 每天在规定时间内保证增量数据的迁移完成;
- 可检查和监控任务的成功或失败;
- 保证数据的正确性:增量数据有可能是新文件,原旧文件可能已经删除,也有可能是同名文件,但是会更新数据;这两种情况都需要处理;
- 按时完成数据的同步;

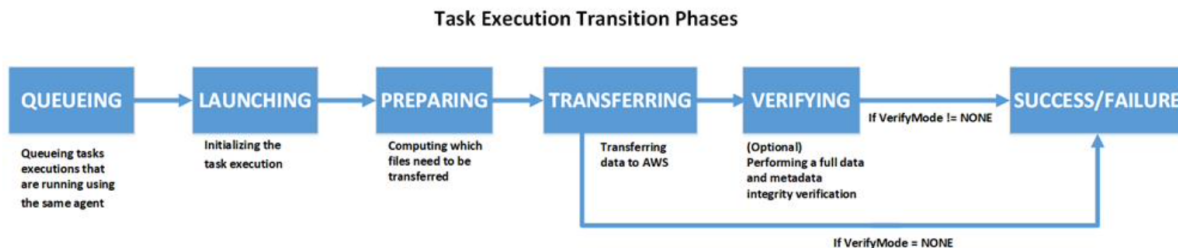
架构



AWS DataSync 使用 S3 互操作性 XML API 连接到您的 GCS 存储桶。它使用 HMAC AccessKey 和 SecretKey 进行身份验证，并同步任务中定义的数据。在 AWS 区域中，AWS DataSync 代理部署为 Amazon EC2 实例，代理使用 VPC Endpoint 连接到 AWS DataSync 服务以在同一子网内移动数据，从而避免任何跨 AZ 或跨区域费用在 AWS 内。代理从 GCS 读取数据，将其发送到 AWS DataSync 服务，然后将数据推送到 S3 存储桶。

DataSync的组件介绍:

- **代理 (agent)** – 用于从自我管理位置读取数据或将数据写入到自我管理位置的虚拟机 (VM)。在同一账户中的 AWS 存储服务之间传输时不需要代理。
- **位置(Location)**– 数据传输中使用的任何源或目标位置（例如，Amazon S3、Amazon EFS、Amazon FSx for Windows File Server、NFS、SMB 或自我管理的对象存储）。
- **任务(Task)**– 由源位置和目标位置以及定义数据传输方式的配置组成。任务总是将数据从源传输到目标。配置可以包括任务计划、带宽限制等选项。任务是数据传输的完整定义。
- **任务执行(Task Execution)**– 任务的单独运行，包括开始时间、结束时间、写入的字节数和状态等信息。



前提条件

- [AWS CLI](#)
- Create [HMAC keys](#) from a GCP account that has access to the GCS bucket to transfer data from. You will use these credentials to configure an Object Storage location for DataSync.

过程

GCP

1. GCP中创建Service account

 IAM 和管理

 IAM

 身份和组织

 政策问题排查工具

 政策分析器

 组织政策

 服务帐号

 工作负载身份联合

 标签

 标记

 设置

 管理资源

 版本说明

<|

创建服务帐号

1 服务帐号详情

服务帐号名称

此服务帐户的显示名

服务帐号 ID *

✕ ↺

电子邮件地址: <id>@gcp-347713.iam.gserviceaccount.com 

服务帐号说明

说明此服务帐号的用途

创建并继续

2 向此服务帐号授予对项目的访问权限 (可选)

3 向用户授予访问此服务帐号的权限 (可选)

完成

取消

2. 为Cloud Storage创建服务账号的密钥

<div>Cloud Storage</div> <div> <div>浏览器</div> <div>监控</div> <div>设置</div> </div> <div>管理资源</div> <div>Marketplace</div> <div>版本说明</div> <div><1</div> <div>Cloud Storage</div> <div> <div>浏览器</div> <div>监控</div> <div>设置</div> </div> <div>管理资源</div>	<div>设置</div> <div>项目访问权限 互操作性</div> <div> <p>Interoperability API 可让 Google Cloud Storage 与专为其他云端存储系统开发的工具实现互操作。这可让您运行以 Cloud Storage 为目标的迁移，并使用带有密钥的哈希消息身份验证代码 (HMAC) 对用户和服务帐号进行身份验证。了解详情</p> <p>请求端点</p> <p>在搭配其他云端存储系统（例如 Amazon S3）使用的工具或库中，确保请求端点使用 Cloud Storage URI。</p> <div>Storage URI <code>https://storage.googleapis.com</code></div> <p>服务帐号 HMAC</p> <p>如果不希望将 HMAC 身份验证与特定的用户帐号绑定，可以为组织的 Cloud Platform 服务帐号使用访问密钥。了解详情</p> <ul style="list-style-type: none"> 每个服务帐号最多可使用 5 个密钥。 请注意，必须先停用密钥才能将其删除。 向您的服务帐号授予执行意向操作所需的权限 - 通常是 IAM Storage Object Admin 角色。 <p>服务帐号的访问密钥</p> </div> <div> <div>设置</div> <div>Storage URI <code>https://storage.googleapis.com</code></div> <p>服务帐号 HMAC</p> <p>如果不希望将 HMAC 身份验证与特定的用户帐号绑定，可以为组织的 Cloud Platform 服务帐号使用访问密钥。了解详情</p> <ul style="list-style-type: none"> 每个服务帐号最多可使用 5 个密钥。 请注意，必须先停用密钥才能将其删除。 向您的服务帐号授予执行意向操作所需的权限 - 通常是 IAM Storage Object Admin 角色。 <p>服务帐号的访问密钥</p> <p>此项目没有任何服务帐号 HMAC 密钥。</p> <div>+ 为服务帐号创建密钥</div> </div>
---	--

选择之前创建的Service account:

选择服务帐号

≡ 过滤条件

输入属性名称或值

电子邮件 ↑	名称	密钥
<input type="radio"/> 623563503379-compute@developer.gserviceaccount....	Compute Engine default service ...	0
<input checked="" type="radio"/> datasyncserviceaccount@gcp-347713.iam.gserviceac...	datasyncserviceaccount	0

取消

创建密钥

创建新帐号

3. 注意保留生成的密钥：

访问密钥：GOOG1EC3BLJN4DC3KRWD7B2TMLK23K3T2NZBTSC6QZAXWGBZNOVOL75I
密钥：KoWx1l+oFQ2XCvWUK47fEisCRumqhW2sbB/

新服务帐号 HMAC 密钥

datasyncserviceaccount@gcp-347713.iam.gserviceaccount.com	
访问密钥	GOOG1ECGF4N3BLJN4DC3KRWD7B2TMLK23K3T2NZBTSC6QZAXWGBZNOVOL75I
密钥	KoWx1l+oFQ2XCvWUK47fEisCuBoHXRunqhW2sbB/

如果将来会用到此密钥的密文，请复制该密文。关闭此对话框后就无法恢复该密文了。

1. 创建VPC endpoint:

New VPC Experience

Tell us what you think

Subnets

Route Tables

Internet Gateways

Egress Only Internet Gateways

Carrier Gateways

DHCP Option Sets

Elastic IPs

Managed Prefix Lists

Endpoints

Endpoint Services

NAT Gateways

Endpoints (2)

Info

↺

↻

Actions

Create endpoint

Filter endpoints

< 1 > ⚙

<input type="checkbox"/>	Name	VPC endpoint ID	VPC ID	Service name
<input type="checkbox"/>	s3enpoints	vpce-0bf8a09d13532671a	vpce-88a9d2f3	com.amazonaws.us-east-1.s3
<input type="checkbox"/>	datasyncendpoint	vpce-0162be71cc1983ae1	vpce-88a9d2f3	com.amazonaws.us-east-1.datasync

Select an endpoint

⏏

⏏

⏏

Create endpoint [Info](#)

There are three types of VPC endpoints – Interface endpoints, Gateway Load Balancer endpoints, and Gateway endpoints. Interface endpoints and Gateway Load Balancer endpoints are powered by AWS PrivateLink, and use an Elastic Network Interface (ENI) as an entry point for traffic destined to the service. Interface endpoints are typically accessed using the public or private DNS name associated with the service, while Gateway endpoints and Gateway Load Balancer endpoints serve as a target for a route in your route table for traffic destined for the service.

Endpoint settings

Name tag - *optional*

Creates a tag with a key of 'Name' and a value that you specify.

Service category

Select the service category

☒ **AWS services**

Services provided by Amazon

☐ **PrivateLink Ready partner services**

Services with an AWS Service Ready designation

☐ **AWS Marketplace services**

Services that you've purchased through AWS Marketplace

☐ **Other endpoint services**

Find services shared with you by service name

Services (1/1)



1



Service Name: com.amazonaws.s3-global.accesspoint ✕

Clear filters

VPC
Select the VPC in which to create the endpoint

VPC
The VPC in which to create your endpoint.

vpc-88a9d2f3

Additional settings

DNS name

☒ **Enable DNS name** [Info](#)
Associates a private hosted zone with the VPC that contains a record set that enables you to leverage Amazon's private network connectivity to the service while making requests to the service's default public endpoint DNS name. To use this feature, ensure that the attributes 'Enable DNS hostnames' and 'Enable DNS support' are enabled for your VPC.

Subnets (2/6) [Info](#)

	Availability Zone	Subnet ID
<input checked="" type="checkbox"/>	us-east-1a (use1-az6)	subnet-4bd7c816
<input checked="" type="checkbox"/>	us-east-1b (use1-az1)	subnet-9a1af0fd
<input type="checkbox"/>	us-east-1c (use1-az2)	
<input type="checkbox"/>	us-east-1d (use1-az4)	
<input type="checkbox"/>	us-east-1e (use1-az3)	
<input type="checkbox"/>	us-east-1f (use1-az5)	

Security groups (1) [Info](#)

Filter security groups

Group name: launch-wizard-4 X Clear filters

	Group ID	Group name	VPC ID
<input type="checkbox"/>	sg-04b9ed7c13d1ac1c9	launch-wizard-4	vpc-88a9d2f3

Policy [Info](#)
VPC endpoint policy controls access to the service.

☒ **Full access**
Allow access by any user or service within the VPC using credentials from any Amazon Web Services accounts to any resources in this Amazon Web Services service. All policies — IAM user policies, VPC endpoint policies, and Amazon Web Services service-specific policies (e.g. Amazon S3 bucket policies, any S3 ACL policies) — must grant the necessary permissions for access to succeed.

☐ **Custom**
Use the [policy creation tool](#) to generate a policy, then paste the generated policy below.

1

注意，安全组的端口要记得对互联网开放80和22，同时对安全组自己要开放：

VPC > Security Groups > sg-04b9ed7c13d1ac1c9 - launch-wizard-4 > Edit inbound rules

Edit inbound rules [Info](#)

Inbound rules control the incoming traffic that's allowed to reach the instance.

Inbound rules [Info](#)

Security group rule ID	Type Info	Protocol Info	Port range Info	Source Info	Description - optional Info		
sg-0125c34af0660630e	SSH	TCP	22	Custom	<input type="text" value="Q"/>	<input type="text" value="0.0.0.0/0"/> <input type="button" value="X"/>	<input type="button" value="Delete"/>
sg-0be326d0dda1992ed	HTTP	TCP	80	Custom	<input type="text" value="Q"/>	<input type="text" value="0.0.0.0/0"/> <input type="button" value="X"/>	<input type="button" value="Delete"/>
sg-0860be477a75e32a9	All TCP	TCP	0 - 65535	Custom	<input type="text" value="Q"/>	<input type="text" value="sg-04b9ed7c13d1ac1c9"/> <input type="button" value="X"/>	<input type="button" value="Delete"/>

Tags

Key

Value - optional

You can add 49 more tags.

2. 安装agent所需要的EC2:

在AWS的EC2上或具备aws cli执行的环境中，执行如下命令（注意替换红字）：

```
aws ssm get-parameter --name /aws/service/datasync/ami --region $region
```

得到如下内容，重点关注ami-id：

```
aws ssm get-parameter --name /aws/service/datasync/ami --region us-east-1

{
  "Parameter": {
    "Name": "/aws/service/datasync/ami",
    "Type": "String",
    "Value": "ami-id",
    "Version": 6,
    "LastModifiedDate": 1569946277.996,
    "ARN": "arn:aws:ssm:us-east-1::parameter/aws/service/datasync/ami"
  }
}
```

替换source-file-system-region为实际的region代码，并用上面的实际红字来替换下面命令中的和ami-id：

https://console.aws.amazon.com/ec2/v2/home?region=**source-file-system-region**#LaunchInsta

Launch an instance [Info](#)

Amazon EC2 allows you to create virtual machines, or instances, that run on the AWS Cloud. Quickly get started by following the simple steps below.

Name and tags

Name

datasyncagent

[Add additional tags](#)

▼ Application and OS Images (Amazon Machine Image) [Info](#)

An AMI is a template that contains the software configuration (operating system, application server, and applications) required to launch your instance. Search or Browse for AMIs if you don't see what you are looking for below

Q Search our full catalog including 1000s of application and OS images

AMI from catalog

Recents

My AMIs

Quick Start

Amazon Machine Image (AMI)

aws-datasync-2.0.1650659999.1-x86_64-gp2
ami-0913ce02ce00eecee



[Browse more AMIs](#)

Including AMIs from
AWS, Marketplace and
the Community

Published	Architecture	Virtualization	Root device type	ENA Enabled
2022-04-22T21:40:04.000Z	x86_64	hvm	ebs	Yes

记得不要选小于M5.2xlarge的机型。

▼ Instance type [Info](#)

Instance type

m5n.2xlarge

Family: m5n 8 vCPU 32 GiB Memory

On-Demand Linux pricing: 0.476 USD per Hour

On-Demand Windows pricing: 0.844 USD per Hour

[Compare instance types](#)

▼ Key pair (login) [Info](#)

You can use a key pair to securely connect to your instance. Ensure that you have access to the selected key pair before you launch the instance.

Key pair name - *required*

virginia



[Create new key pair](#)

Tags

Key

Q Name

X

Value - optional

Q datasyncendpoint

X

Remove

Add new tag

You can add 49 more tags.

Cancel

Create endpoint

注意安全组的选择：

要22和80端口可以让公网访问，同时endpoint 和你的ec2的安全组最好配置为一个，这样减少配置，里面把本安全组自己配置进去。

▼ Network settings

VPC - required [Info](#)

vpc-88a9d2f3

172.31.0.0/16

(default) ▼

↺

Subnet [Info](#)

subnet-4bd7c816

VPC: vpc-88a9d2f3 Owner: 154998728054 Availability Zone: us-east-1a

IP addresses available: 4050

▼

↺

Create new subnet [↗](#)

Auto-assign public IP [Info](#)

Enable

▼

Firewall (security groups) [Info](#)

A security group is a set of firewall rules that control the traffic for your instance. Add rules to allow specific traffic to reach your instance.

☐ Create security group

☒ Select existing security group

Common security groups [Info](#)

Select security groups ▼

↺ Compare security group rules

launch-wizard-4 sg-04b9ed7c13d1ac1c9 X

VPC: vpc-88a9d2f3

Security groups that you add or remove here will be added to or removed from all your network interfaces.

▶ Advanced network configuration

在DataSync中创建agent：

DataSync

Tasks

Locations

Agents

DataSync > Agents

Agents (1)

Agents read data from or write to your on-premises locations.

Q Find agents

Edit

Delete

Create agent

< 1 >

Create agent

Deploy agent

To transfer data between self-managed storage and AWS Storage services, ensure you have deployed and started the agent VM. Your agent can run on either an on-premises hypervisor or Amazon EC2.

Hypervisor

Amazon EC2

Create an EC2 instance using the AMIs provided in our User Guide. [Learn more](#)

Service endpoint

Endpoint type

Choose the service endpoints that this agent will connect to. [Learn more](#)

VPC endpoints using AWS PrivateLink

VPC endpoint

The Virtual Private Cloud endpoint that the agent has access to.

vpce-0162be71cc1983ae1 (com.amazonaws.us-east-1.datasync)



[Create VPC endpoint](#)

Subnet

The subnet in which elastic network interfaces will be created for tasks that run on this agent.

subnet-4bd7c816
us-east-1a

172.31.42.190



[Create subnet](#)

[Manage subnets](#)

Security group

The security group that will protect the elastic network interfaces for tasks that run on this agent.

launch-wizard-4 (sg-04b9ed7c13d1ac1c9)



[Create security group](#)

[Manage security groups](#)

Activation key

Activation securely associates the agent that you deployed with your AWS account. [Learn more](#)

Choose a method for obtaining your agent's activation key. [Learn more](#)

☒ Automatically get the activation key from your agent

Your browser will connect to your agent over HTTP to get a unique activation key.

☐ Manually enter your agent's activation key

Choose this option if you already have an activation key and need to enter it manually.

Agent address

When your DataSync agent has been deployed, enter its domain name or IP address. When you click **Get key**, your browser will connect to this address to get a unique activation key.

http:// 15.153.179.72

Make sure this address is accessible from your browser.

Cancel

[Get key](#)

激活agent, 得到下面这样的key,然后创建agent:

Activation key

Activation securely associates the agent that you deployed with your AWS account. [Learn more](#)



Successfully retrieved activation key from agent

Activation key

SH1N9-G2TNO-ILUHN-RMUNC-3JIS

Agent address

35.153.1.73

Agent name - optional

Agent name

Give your agent a name so you can easily identify it.

My agent

Tags - optional

Optionally add tags to your agent

No tags associated with the resource.

Add tag

You can add up to 50 more tags.

Cancel

Create agent

3. 创建源location :

Configuration

Location type

Object storage ▼

Agents

The agents you choose should have access to the object storage server.

Choose one or more agents ▼

gcpagent (agent-00de016f5def15782) VPC Endpoint - 172.31.42.190 ✕
vpce-0162be71cc1983ae1 subnet-4bd7c816 sg-04b9ed7c13d1ac1c9

Server

Domain name or IP address of the object storage server.

storage.googleapis.com

Bucket name

Name of the bucket that will be used for the data transfer.

testbucketname

Folder

Choose a folder prefix that will be used for the data transfer.

/test

Additional settings

Server protocol

☐ HTTP

☒ HTTPS

Server port

443

Authentication

☒ Requires credentials

Access key

Access key needed to access the bucket.

写入得到的gcp service account access key

Secret key

Secret key needed to access the bucket.

.....

Tags

No tags associated with the resource.

Add tag

You can add up to 50 more tags.

Cancel

Create location

创建destination location:y

Create location

Configuration

Location type

Amazon S3

S3 bucket

albingress

us-east-1

S3 storage class

Choose a storage class you want your objects to be transferred into.

Standard

Folder

Choose a folder prefix that will be used for the data transfer.

/target

IAM role

Used to access the selected S3 bucket. [Learn more](#)

AWSDataSyncS3BucketAccess-albingress



Autogenerate

The selected bucket already has an autogenerated role: AWSDataSyncS3BucketAccess-albingress

4. 创建task

DataSync

Tasks

History

Locations

Agents

DataSync > Tasks

Tasks (3)

Tasks are configurations for data transfer and synchronization between two locations.

< 1 >

<input type="checkbox"/>	Task ID	Name	Status	Source	Destination	Created
--------------------------	---------	------	--------	--------	-------------	---------

配置源location:

DataSync > Tasks > Create task

Step 1

Configure source location

Step 2

Configure destination location

Step 3

Configure settings

Step 4

Review

Configure source location

Source location options
Create or select a source location

Choose the location that you want to transfer data from

Create a new location

Choose an existing location

Choose existing location

Region

US East (N. Virginia)us-east-1

Existing locations

object-storage://storage.googleapis.com/learnings_soli_test/
loc-0529f5cf3f976fee5

VPC Endpoint - 172.31.42.190

Cancel

Next

配置目标location:

DataSync > Tasks > Create task

Step 1

Configure source location

Step 2

Configure destination location

Step 3

Configure settings

Step 4

Review

Configure destination location

Destination location options
Create or select a destination location

Choose the location that you want to transfer data from

Create a new location

Choose an existing location

Choose existing location

Region

US East (N. Virginia)
Transfers from Object storage must stay in the current region.

Existing locations

s3://albingress/
Standardloc-06520361a7cceab8c

Transfers from Object storage can go to Amazon EFS, Amazon FSx, or Amazon S3.

Cancel

Previous

Next

5. 设置其它配置

Step 1

[Configure source location](#)

Step 2

[Configure destination location](#)

Step 3

Configure settings

Step 4

[Review](#)

Configure settings

Task Name

datasynctest

Task execution configuration

Verify data

Verify the data in your destination after transfer. DataSync always performs data integrity checks while transferring and writing data.

Verify only the data transferred

Set bandwidth limit

Allocate maximum bandwidth to be utilized by this task

☒ Use available

☐ Set bandwidth limit (MiB/s)

Additional settings

Queueing

☒ Enabled

Allow this task to be started and queued until one of its agents becomes available

Data transfer configuration

Data to scan

☒ Entire source location

Scan all source data.

☐ Specific files and folders


Scan specific files, folders, and objects.

Transfer mode

Transfer only data that has changed

☐ Keep deleted files

Keep files in destination even when deleted from source

 Files not present on the source will be deleted from the destination when this task executes.

☒ Overwrite files

Overwrite files in destination with files from source

► Exclude patterns (0)

Additional settings

☒ Copy object tags

Maintain tags on objects

Schedule

Execute this task on a schedule by specifying the time and frequency.

Frequency

Hourly ▼

at minute

00 ▼

Tags

No tags associated with the resource.

Add tag

You can add up to 50 more tags.

Task logging

Log level
Defines the level of detail that is logged to CloudWatch Logs. CloudWatch Logs pricing applies for logs generated by DataSync.

Log basic information such as transfer errors ▼

CloudWatch log group
Log task execution activity such as error or other failures.

Choose a CloudWatch Log group for task execution activity ▼

↺

Autogenerate

Cancel

Previous

Next

6. 最后创建并启动任务

DataSync

×

Tasks

History

Locations

Agents

DataSync > Tasks

Tasks (1/3)

Tasks are configurations for data transfer and synchronization between two locations.

Find tasks

Task ID	Name	Status	Source	Destination	
<input checked="" type="checkbox"/> task-00d31e59632acb2af	idailybread	Available	object-storage://storage.goo...	s3://albingress/idailybread/	
<input type="checkbox"/> task-017d3e11c4bfd291f	second	Available	object-storage://storage.goo...	s3://albingress/	May 7, 2022, 8:16:42 PM GMT+8
<input type="checkbox"/> task-05b85a25b0b395842	syncgcptos3	Available	object-storage://storage.goo...	s3://albingress/	May 7, 2022, 7:55:00 PM GMT+8

Actions

Create task

Start

Start with overriding options

Stop

Edit

Delete

7. 监控任务执行

LocationsOptionsScheduleTagsTask loggingHistory

History (6)

History shows each task execution and contains information about the execution results.

Find task executions

1

	Execution ID	Start time	Duration	Status	MIB/s	Files/sec
<input type="radio"/>	exec-0ceef51e5e10976e0	May 11, 2022, 5:00:47 PM GMT+8	10 seconds	Success	0	0
<input type="radio"/>	exec-0e233e35d57fe32f0	May 11, 2022, 4:00:47 PM GMT+8	9 seconds	Success	0	0
<input type="radio"/>	exec-0a55ff9b22365ca7a	May 11, 2022, 3:00:47 PM GMT+8	10 seconds	Success	0	0
<input type="radio"/>	exec-06bfd497042ee96da	May 11, 2022, 2:00:47 PM GMT+8	9 seconds	Success	0	0
<input type="radio"/>	exec-0f900627463ce95e1	May 11, 2022, 1:00:47 PM GMT+8	10 seconds	Success	0	0
<input type="radio"/>	exec-05de4d87439d4e479	May 11, 2022, 12:35:47 PM GMT+8	18 seconds	Success	38.3	0.54

8. 结果

Objects (5)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Copy S3 URICopy URLOpenDeleteActionsCreate folder

Upload

Find objects by prefix

1

	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	gcs-to-s3-2.0.0.jar	jar	May 11, 2022, 12:38:39 (UTC+08:00)	70.6 MB	Standard
<input type="checkbox"/>	gcs-to-s3-2.2.1.jar	jar	May 11, 2022, 12:38:39 (UTC+08:00)	70.6 MB	Standard
<input type="checkbox"/>	gcs-to-s3-2.3.2.jar	jar	May 11, 2022, 12:38:39 (UTC+08:00)	70.6 MB	Standard
<input type="checkbox"/>	gcs-to-s3-2.3.3.jar	jar	May 11, 2022, 12:38:39 (UTC+08:00)	70.6 MB	Standard
<input type="checkbox"/>	gcs-to-s3-2.3.4.jar	jar	May 11, 2022, 12:38:40 (UTC+08:00)	70.6 MB	Standard

结论

Datasync可以满足客户对于同步存量数据和增量数据从GCS到S3的功能，并有如下优势：

- 数据加密和验证
您的所有数据在传输过程中都使用传输层安全性 (TLS) 进行加密。DataSync 支持对 S3 存储桶使用默认加密、静态数据的 Amazon EFS 文件系统加密以及静态和传输中数据的 Amazon FSx for Windows File Server 加密。AWS DataSync 可确保您的数据完好无损地到达。对于每次传输，服务都会在传输中和静止时执行完整性检查。这些检查确保写入目标的数据与从源读取的数据相匹配，从而验证一致性。
- 数据传输调度
AWS DataSync 带有内置的调度机制，使您能够定期执行数据传输任务以检测更改从源存储系统复制到目标。您可以使用 AWS DataSync 控制台或 AWS 命令行界面 (CLI) 安排您的任务，而无需编写脚本来管理重复传输。任务计划自动按照您配置的计划运行任务，控制台中直接提供每小时、每天或每周选项。

- 使用 Amazon CloudWatch 和 AWS CloudTrail 进行监控和审计

使用 Amazon CloudWatch，您可以监控当前正在进行的任何 DataSync 传输的状态，并检查以前数据传输的历史记录。使用 CloudWatch 指标，您可以查看已复制的文件数量和数据量。您可以查阅 CloudWatch Logs 以了解有关在给定时间传输的单个文件的信息，以及 DataSync 执行的完整性验证的结果。这简化了监控、报告和故障排除，并使您能够及时向利益相关者提供更新。此外，可以在您的传输任务完成时触发 CloudWatch Events，从而实现依赖工作流的自动化。出于审计目的，您可以查阅 AWS CloudTrail，它记录了 DataSync 执行的所有操作。

- 即用即付定价

您只需为服务复制的数据付费，按每 GB 统一的费率付费，无需软件许可、合同、维护费用、开发周期或所需硬件。与手动构建、操作和优化您自己的高性能脚本传输相比，这提供了更低的总拥有成本。它还提供比购买和运行商业传输工具更低的总成本。

- 自动基础架构管理

AWS DataSync 消除了您在编写、优化和管理自己的复制脚本或部署和调整重量级商业传输工具时面临的许多基础设施和管理挑战。DataSync 带有内置的监控和重试机制，以及对用于传输数据的网络带宽部分的精细控制。