

Multi-Cloud Data Transfer with AWS and GCP

RU

Rubanpreet Singh

The screenshot shows the AWS S3 console interface. At the top, there is a breadcrumb navigation: 'nextwork-data-transfer-destination-gcp-ruban'. Below the navigation, there are tabs for 'Objects' (which is selected), 'Configuration', 'Permissions', 'Protection', 'Lifecycle', 'Observability', 'New', 'Inventory Reports', and 'Reports'. The 'Objects' tab is active. In the main content area, there is a 'Create folder' button, an 'Upload' button with a dropdown, an 'Other services' dropdown, and a 'Transfer data' dropdown. Below these buttons, there is a 'Filter by name prefix only' dropdown, a 'Filter' button, and a 'Filter objects and folders' input field. To the right of these filters, there is a 'Show' button with a dropdown labeled 'Live objects only' and a 'More' button. A table below the filters lists three objects:

<input type="checkbox"/>	Name	Size	Type	Created	Storage class	Actions
<input type="checkbox"/>	Automate Your Browser with AI Ag...	18.2 MB	application/pdf	Apr 12, 2025, 5:31:39 PM	Standard	
<input type="checkbox"/>	Build a Three-Tier Web App.pdf	17.4 MB	application/pdf	Apr 12, 2025, 5:31:39 PM	Standard	
<input type="checkbox"/>	Building an AI Workflow.pdf	17.2 MB	application/pdf	Apr 12, 2025, 5:31:39 PM	Standard	

Introducing Today's Project!

In this project, I will demonstrate how to set up a multi-cloud storage system using AWS and GCP. I'm doing this project to learn how to move data between AWS S3 and Google Cloud Storage.

Tools and concepts

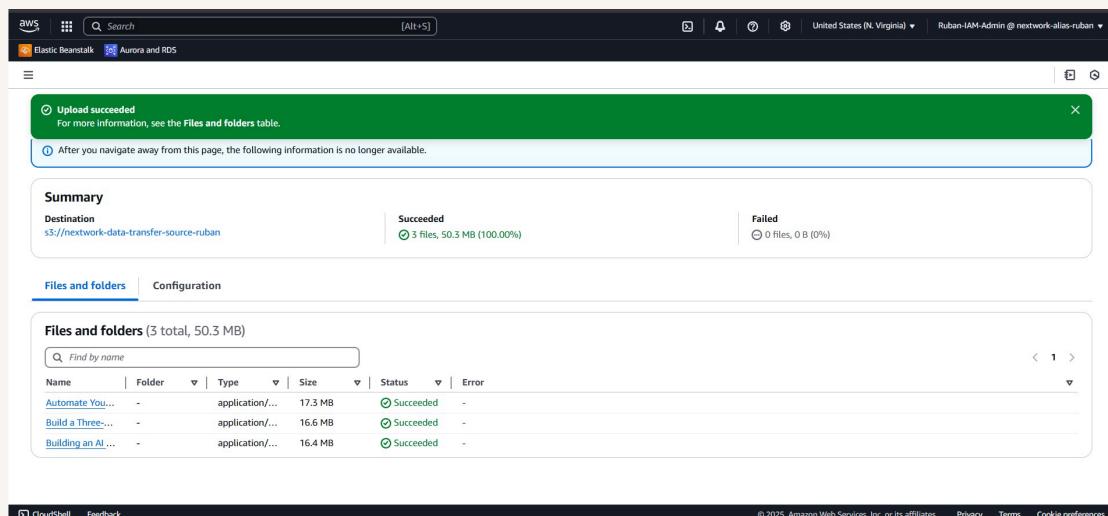
Services I used were Amazon S3, IAM, Google Cloud Storage, Storage Transfer Service. Key concepts I learnt include IAM Roles and permissions and multi-cloud data transfer.

Project reflection

This project took me approximately 1hr 30mins. The most challenging part was to make sure the IAM role had the required permissions. It was most rewarding to navigate the GCP bucket and open the files after the transfer. was complete.

Setting up Data in S3

I started this project by setting up a S3 bucket with a unique name and I uploaded three pdf files to the bucket.

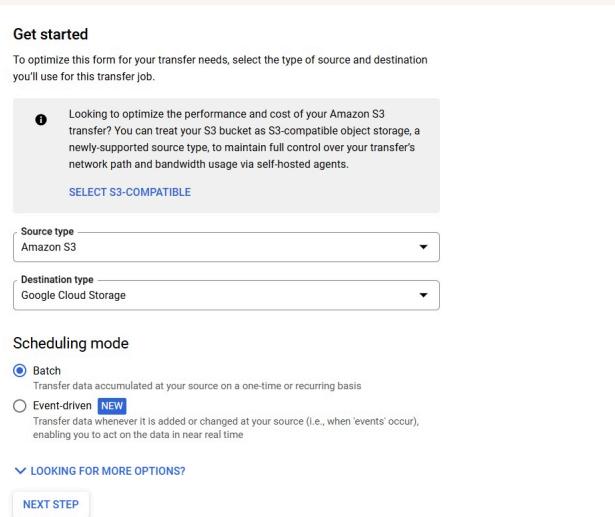


Storage Transfer

Data transfers are important for reducing dependency on single vendor. Using multiple cloud providers for the transfer has benefits such as cost optimization and ensuring high availability of data.

The transfer is set up using Storage Transfer Service, which is a service that helps in transferring data in and out of the GCP Cloud Storage. We need this service because cloud providers don't offer native ways to connect to their competitors.

There are two different types of transfers you could set up one is Batch and the other is Event-driven transfer. The Batch option transfers data at once or schedules the transfer while Event-driven transfers data when new data is available.



The screenshot shows the 'Get started' page for setting up a data transfer. The page has a light gray background with a white central form area. At the top left of the form, the text 'Get started' is in bold. Below it is a sub-instruction: 'To optimize this form for your transfer needs, select the type of source and destination you'll use for this transfer job.' A callout box with a blue border and a small info icon contains the text: 'Looking to optimize the performance and cost of your Amazon S3 transfer? You can treat your S3 bucket as S3-compatible object storage, a newly-supported source type, to maintain full control over your transfer's network path and bandwidth usage via self-hosted agents.' Below this box is a blue button labeled 'SELECT S3-COMPATIBLE'. The 'Source type' dropdown is set to 'Amazon S3'. The 'Destination type' dropdown is set to 'Google Cloud Storage'. Under 'Scheduling mode', the radio button for 'Batch' is selected, with the sub-instruction: 'Transfer data accumulated at your source on a one-time or recurring basis'. The 'Event-driven' option is also present with its own sub-instruction: 'Transfer data whenever it is added or changed at your source (i.e., when "events" occur), enabling you to act on the data in near real time'. At the bottom of the form, there is a link 'LOOKING FOR MORE OPTIONS?' and a blue 'NEXT STEP' button.

Granting GCP Access to AWS

To connect AWS and GCP, I'm using identity federation, which works by providing temporary access to AWS resources without permanent credentials. This is more secure than other methods because these credentials expire after use typically 15 - 60 mins.

I created a custom IAM role for granting GCP Storage Service access to AWS S3. This role only lets GCP to assume the role and the permissions only allow read access to S3 buckets following the least privilege principle.

Within the IAM role, I needed to write a custom trust policy because it allows us to define who can assume the role. The policy identifies GCS based on a subject ID, which is generated by Google for every GCP service for every account.

Custom trust policy

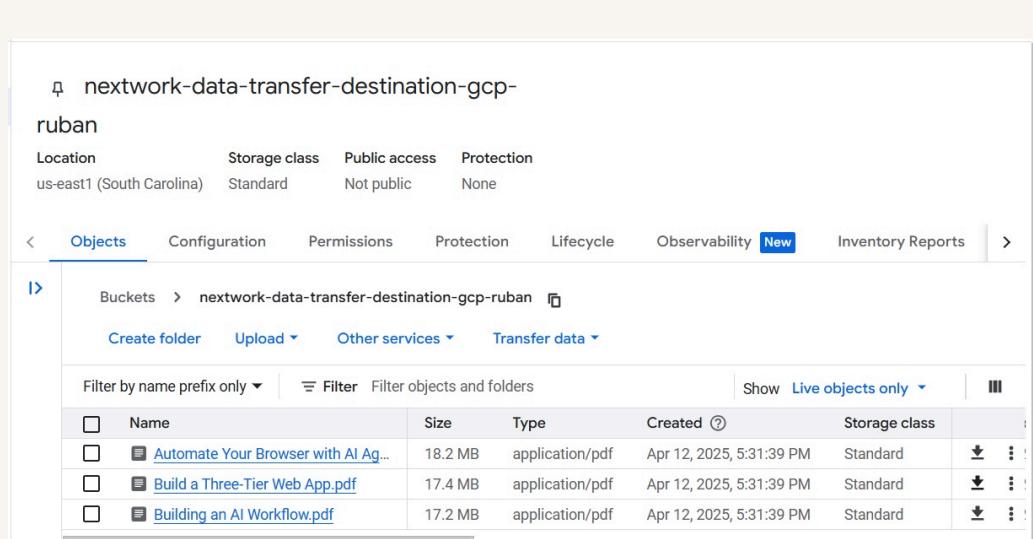
Create a custom trust policy to enable others to perform actions in this account.

```
1 ▼ {
2     "Version": "2012-10-17",
3     "Statement": [
4         {
5             "Effect": "Allow",
6             "Principal": {
7                 "Federated": "accounts.google.com"
8             },
9             "Action": "sts:AssumeRoleWithWebIdentity",
10            "Condition": {
11                "StringEquals": {
12                    "accounts.google.com:sub": "113341202734135078331"
13                }
14            }
15        }
16    ]
17 }
```

Transferring from S3 to GCS!

To set up my destination GCS bucket, I needed to set up its region, which means the location where the data will be stored in GCP and its storage class, which means the pricing plan for the bucket decided based on how frequently the data is accessed.

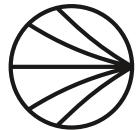
I verified my data transfer was successful by navigating to the bucket in the Cloud Storage Menu. A new bucket was created and the files that were in S3 bucket were now available in the GCS bucket.



Name	Size	Type	Created	Storage class	Actions
Automate Your Browser with AI App...	18.2 MB	application/pdf	Apr 12, 2025, 5:31:39 PM	Standard	
Build a Three-Tier Web App.pdf	17.4 MB	application/pdf	Apr 12, 2025, 5:31:39 PM	Standard	
Building an AI Workflow.pdf	17.2 MB	application/pdf	Apr 12, 2025, 5:31:39 PM	Standard	

Transfer with a Manifest

I verified my data transfer was successful by navigating to the bucket in the Cloud Storage Menu. A new bucket was created and the files that were in S3 bucket were now available in the GCS bucket.



NextWork.org

Everyone should be in a job they love.

Check out nextwork.org for
more projects

