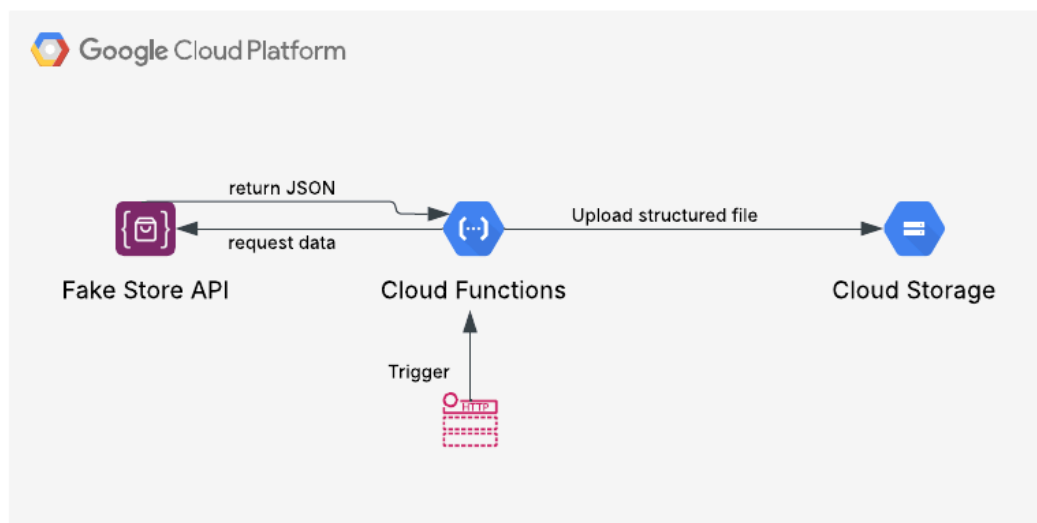


Data Ingestion: Fetching and Storing Product Data

I tackled this task using two different approaches. In the first approach, I manually created the required GCP resources via the console. In the second, I implemented a CI/CD pipeline to automate the provisioning of resources.

Architecture Diagram of the Data Ingestion Pipeline :



I chose a serverless solution using **Cloud Functions (2nd generation)**, which is well-suited for lightweight, event-driven processing and benefits from improved performance and scalability. The function is an **HTTP-triggered Gen 2 Cloud Function**, meaning it is deployed on **Cloud Run**, and it is invoked via an HTTP call using its generated URL.

Once triggered, the function sends a request to the **Fake Store API**, retrieves the **JSON response**, extracts and transforms the data into a **CSV file**, then uploads that structured file to a **Cloud Storage bucket**.

Steps I took to create and execute the pipeline :

Setting IAM roles :

Defined permissions to allow public invocation of the function and **Cloud Storage access** by ensuring the Cloud Function has the appropriate role to **upload the structured file to the target bucket**.

Free trial status: €286.10 credit and 91 days remaining. Activate your full account to get unlimited a

Google CloudMy First ProjectSearch (/) for resources, docs,

IAM & Admin / IAM

IAM

PAM

Principal Access Boun...

Organizations

Identity & Organization

Policy Troubleshooter

Policy Analyzer

Organization Policies

Service Accounts

Workload Identity Fede...

Workforce Identity Fed...

Labels

Manage Resources

Release Notes

IAM

AllowDenyRecommendations history

Permissions for project "My First Project"

These permissions affect this project and all of its resources. [Learn more](#)

View by principalsView by roles

Grant accessRemove access

Filter

Enter property name or value

Type	Principal	Name	Role	Security insights
<input type="checkbox"/>	84391613042-compute@developer.gserviceaccount.com	Default compute service account	Cloud Run Admin	
<input type="checkbox"/>	inssafcources@gmail.com	Inssaf Cherki	Owner	
<input type="checkbox"/>	nomadic-freedom-454920-v8@appspot.gserviceaccount.com	App Engine default service account	Editor	

Now viewing project "My organization"

Edit access to "My First Project"

Principal84391613042-compute@developer.gserviceaccount.com

ProjectMy First Project

Assign roles

Roles are composed of sets of permissions and determine what the principal can do with this resource. [Learn more](#)

RoleEditor

IAM condition (optional)

+ Add IAM condition

View, create, update, and delete most Google Cloud resources. See the list of included permissions.

RoleStorage Object Creator

IAM condition (optional)

+ Add IAM condition

Allows users to create objects. Does not give permission to view, delete, or replace objects.

+ Add another role

Save

Test changes

Cancel

Summary of changes

Roles removed

n/a

Role added

Storage Object Creator

Test changes

Google CloudMy First ProjectiamSearch

IAM & Admin / IAM

IAM

PAM

Principal Access Boun...

Organizations

Identity & Organization

Policy Troubleshooter

Policy Analyzer

Organization Policies

Service Accounts

Workload Identity Fede...

Workforce Identity Fed...

Labels

Manage Resources

Release Notes

IAM

AllowDenyRecommendations history

Permissions for project "My First Project"

These permissions affect this project and all of its resources. [Learn more](#)

View by principalsView by roles

Grant accessRemove access

Filter

Enter property name or value

Type	Principal	Name	Role	Security insights
<input type="checkbox"/>	84391613042-compute@developer.gserviceaccount.com	Default compute service account	Cloud Run Admin	
<input type="checkbox"/>	inssafcources@gmail.com	Inssaf Cherki	Owner	
<input type="checkbox"/>	nomadic-freedom-454920-v8@appspot.gserviceaccount.com	App Engine default service account	Editor	

Policy updated

Bucket creation :

Google Cloud My First Project cloud storage

Cloud Storage Create a bucket

- Get Started**
Pick a globally unique, permanent name. [Naming guidelines](#)

Tip: Don't include any sensitive information
Optimize storage for data-intensive workloads
☐ **Enable Hierarchical namespace on this bucket**
Optimize for AI/ML and analytics with a filesystem-like hierarchical structure. This permanent option enables atomic folder renames, faster folder listing, and other enhancements not available in Cloud Storage's standard buckets that uses flat namespace.
Labels (optional)
Labels are key:value pairs that allow you to group related buckets together or with other Cloud Platform resources. [Learn more](#)
[+ Add label](#)
[Continue](#)
- Choose where to store your data**
This choice defines the geographic placement of your data and affects cost, performance, and availability. Cannot be changed later. [Learn more](#)
Location type
☒ **Multi-region**
Highest availability across largest area

Creating and executing the function :

data-integration Region: us-central1 URL: <https://data-integration-84391613042.us-central1.run.app> Scaling: Auto (Min: 0)

Metrics SLOs Logs Revisions **Source** Triggers Networking Security YAML

Source Base image: Python 3.11 (Ubuntu 22) Function entry point: [Save and redeploy](#) [Discard changes](#)

us-central1-nomadic-freedom-454920-v8.cloudfunctions.net/data-ingestion

Successfully uploaded gs://fakestore-products-data/products.csv

Result :

product-structured-data

Location

us (multiple regions in United States)

Storage class

Standard

Public access

Not public

Protection

Soft Delete

Objects

Configuration

Permissions

Protection

Lifecycle

Observability

New

Inventory Reports

Operations

Folder browser

product-structured-data

Buckets > product-structured-data

Create folderUploadTransfer dataOther services

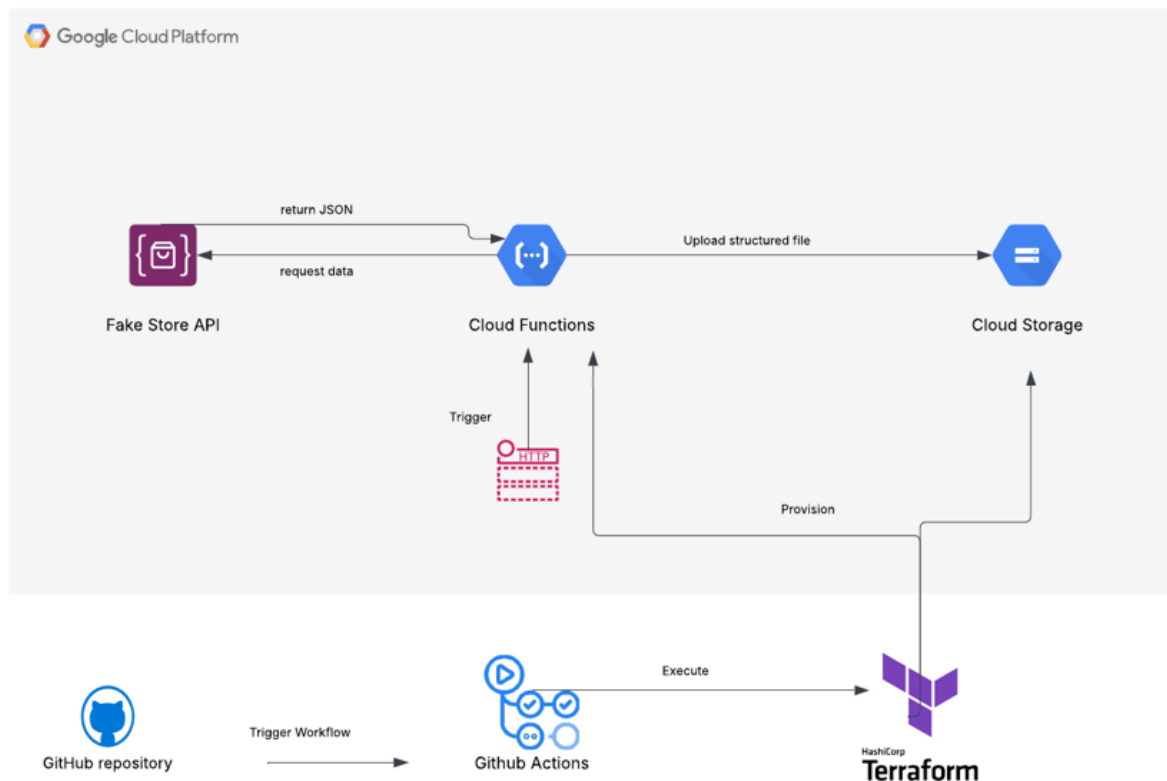
Filter by name prefix onlyFilterFilter objects and foldersShowLive objects only

<input type="checkbox"/>	Name	Size	Type	Created	Storage class	Li
<input type="checkbox"/>	products.csv	9.4 KB	text/csv	Mar 26, 2025, 10:58:12 PM	Standard	M

	A	B	C	D	E	F	G	H
1	id,title,price,description,category,image,rating							
2	1,"Fjallraven - Foldsack No. 1 Backpack, Fits 15 Laptops",109.95,"Your perfect pack for everyday use and travel in bad weather",bags,https://fakestoreapi.com/img/81F0jDnLnDL_000000.jpg,4.8							
3	2,Mens Casual Premium Slim Fit T-Shirts ,22.3,"Slim-fitting style, contrast raglan long sleeve, medium-wash, ultra soft material",t-shirts,https://fakestoreapi.com/img/71-3OjOpNwLk_000000.jpg,4.5							
4	3,Mens Cotton Jacket,55.99,"great outerwear jackets for Spring/Autumn/Winter, suitable for many occasions",jackets,https://fakestoreapi.com/img/21ckidgG8LK_000000.jpg,4.0							
5	4,Mens Casual Slim Fit,15.99,"The color could be slightly different between on the screen and actual product. Please see the picture for detail.",t-shirts,https://fakestoreapi.com/img/21ckidgG8LK_000000.jpg,4.5							
6	5,John Hardy Women's Legends Naga Gold & Silver Dragon Station Chain Bracelet,695,"This gorgeous piece is part of a collection of jewelry that honors India's beautiful myths and legends.",jewelry,https://fakestoreapi.com/img/51UjgUk4gMQ_000000.jpg,4.7							
7	6,Solid Gold Petite Micronave ,168,"Satisfaction Guaranteed. Return or exchange any order within 30 days.",jewelry,https://fakestoreapi.com/img/61UjgUk4gMQ_000000.jpg,4.6							

The function successfully fetched product data from the **Fake Store API**. It transformed the JSON response into a **CSV file** and automatically uploaded it to the designated Cloud Storage bucket.

Architecture Diagram of the CI/CD Data Ingestion Pipeline :



This pipeline automates the provisioning and execution of the serverless data ingestion process using **Terraform**, and **GitHub Actions**.

Services created via Terraform :

The screenshot shows the Google Cloud Platform console for a Cloud Function named 'data-ingestion'. The function is located in the 'us-central1' region and has a URL of <https://us-central1-nomadic-freedom-454920-v8.cloudfunctions.net/data-ingestion>. The source code is written in Python 3.11 (Ubuntu 22 Full) and the entry point is 'fetch_store_data'. The scaling is set to 'Auto (Min: 0)'.

Location	Storage class	Public access	Protection
us (multiple regions in United States)	Standard	Not public	Soft Delete

Result :

The screenshot shows the Google Cloud Storage console for a bucket named 'fakestore-products-data'. The bucket is located in the 'us (multiple regions in United States)' region and has a storage class of 'Standard'. The public access is 'Not public' and the protection is 'Soft Delete'. The bucket contains a single object named 'products.csv' with a size of 9.4 KB, type of 'text/csv', and created on Mar 27, 2025, at 12:31:25 PM.

Name	Size	Type	Created	Storage class	Location
products.csv	9.4 KB	text/csv	Mar 27, 2025, 12:31:25 PM	Standard	M