

# Bucket Creation

Ensure Hierarchical Namespace enabled

Ensure Public Access Prevention: off [For training only]

←

Create a bucket

✓

Get Started

Name: gk2-datalake

✓

Choose where to store your data

Location: europe-west3 (Frankfurt)

Location type: Region

✓

Choose how to store your data

Default storage class: Standard

Hierarchical namespace: Enabled

Anywhere Cache: Disabled

✓

Choose how to control access to objects

Public access prevention: Off

Access control: Uniform

•

Choose how to protect object data

Your data is always protected with Cloud Storage but you can also choose from these additional data protection options to add extra layers of security.

Data protection

❗

Object versioning and retention are not yet available for buckets with hierarchical namespaces.

✓

Soft delete policy (For data recovery)

When enabled, this bucket and its objects will be kept for a specified period after they're deleted and can be restored during this time. [Learn more](#)

●

Use default retention duration

All buckets have a 7 day soft delete duration by default, unless this default has been customized by your organization administrator.

○

Set custom retention duration

Specify how long this bucket and its objects should be retained after they're deleted. Setting a "0" duration disables soft delete, meaning any deleted objects will be permanently deleted.

☐

Object versioning (For version control)

For restoring deleted or overwritten objects. To minimize the cost of storing

Good to know

📄

Location pricing

Storage rates vary depending on the storage class of your data and location of your bucket. [Pricing details](#)

Current configuration: Region / Standard

Item	Cost
europe-west3 (Frankfurt)	\$0.023 per GB-month

Estimate your monthly cost

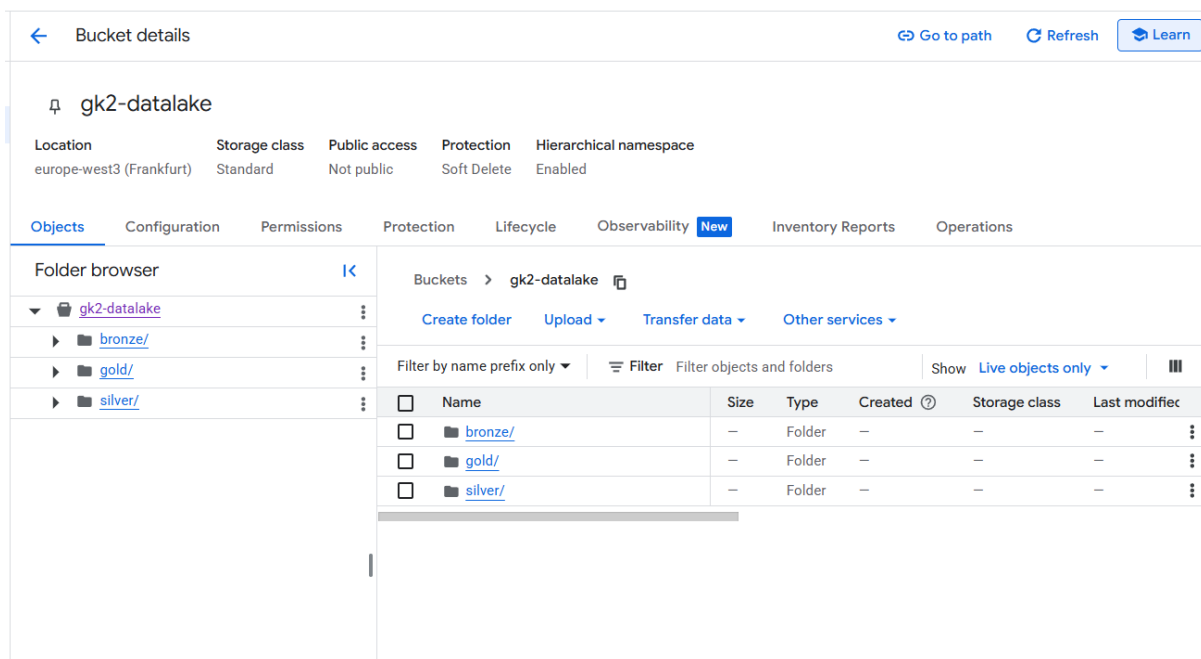
# Create Data Lake Folders

Three folders named bronze, silver, gold under data lake

bronze shall have csv, json, logs, unclear, inefficient storage data

silver shall have efficient, cleaned data like parquet, delta table or iceberg format

gold will have analytical data



## Download MovieLens dataset

MovieLens an open source dataset for movies, ratings, tags

<https://grouplens.org/datasets/movielens/>

The data set to download ml-latest-small.zip

<https://files.grouplens.org/datasets/movielens/ml-latest-small.zip>

Extract the zip files, explore the dataset by opening them in an editor without any modification

# Data Organizations

Create two directories named **movies**, **ratings**, all in lower case under **bronze** directory

gk2-datalake

Location

europa-west3 (Frankfurt)

Storage class

Standard

Public access

Not public

Protection

Soft Delete

Hierarchical namespace

Enabled

Objects

Configuration

Permissions

Protection

Lifecycle

Observability

New

Inventory Reports

Folder browser

gk2-datalake

bronze/

movies/

ratings/

gold/

silver/

Buckets > gk2-datalake > bronze

Create folder

Upload

Transfer data

Other services

Filter by name prefix only

Filter

Filter objects and folders

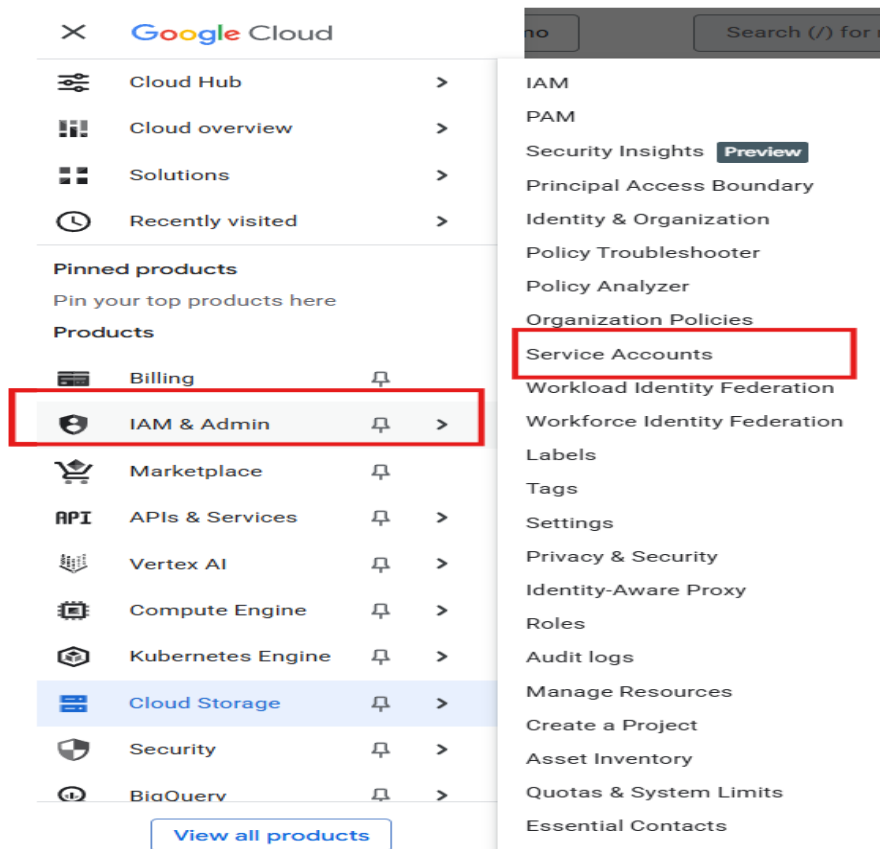
<input type="checkbox"/>	Name	Size	Type	Created
<input type="checkbox"/>	movies/	—	Folder	—
<input type="checkbox"/>	ratings/	—	Folder	—

Upload movies.csv into movies directory

Upload ratings.csv into ratings directory

# Permission to access Datalake from Flink

Go to IAM & Admin then Service account



Click on Create Service Account

← Create service account

---

**1 Create service account**

Service account name

Display name for this service account

Service account ID \*  ✕ ↺

Email address: gk-access@flink-demo-470113.iam.gserviceaccount.com

Service account description

[Create and continue](#)

Select editor permission

On Permission, ensure to use roles/storage.objectAdmin

## 2 Permissions (optional)

Grant this service account access to flink-demo so that it has permission to complete specific actions on the resources in your project. [Learn more](#)

Role

Storage Object Admin

Grants full control over objects, including listing, creating, viewing, and deleting objects.

+ Add another role

Help me choose roles

Continue

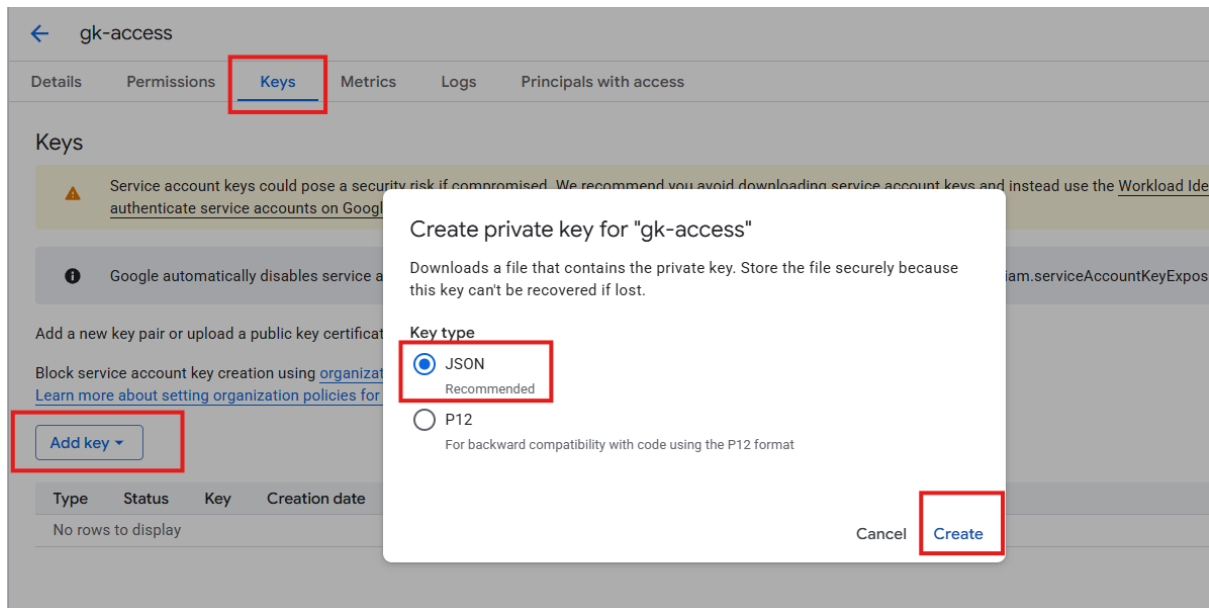
IAM condition (optional) ?

+ Add IAM condition



Click on the service account for generating key.json

Filter Enter property name or value								?	
<input type="checkbox"/>	Email	Status	Name ↑	Description	Key ID	Key creation date	OAuth 2 Client ID ?	Actions	
<input type="checkbox"/>	<a href="mailto:gk-access@flink-demo-470113.iam.gserviceaccount.com">gk-access@flink-demo-470113.iam.gserviceaccount.com</a>	✔ Enabled	gk-access		No keys		108265331223289577470	⋮	



This will download a json file.

Copy the downloaded json file, paste into **secret** directory of your project

Remove existing key.json if any

Ensure to rename the downloaded json file to key.json