

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, unclean, inefficient storage data

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

gold will have analytical data

The screenshot shows the AWS S3 Bucket details page for 'gk2-datalake'. The bucket is located in 'europe-west3 (Frankfurt)' with 'Standard' storage class, 'Not public' public access, 'Soft Delete' protection, and 'Enabled' hierarchical namespace. The 'Objects' tab is selected, displaying a folder browser. Inside 'gk2-datalake', there are three subfolders: 'bronze/', 'gold/', and 'silver/'. Each folder has a 'More options' menu icon (three dots). The object list table shows the following data:

Name	Type	Created	Storage class	Last modified
bronze/	Folder	—	—	—
gold/	Folder	—	—	—
silver/	Folder	—	—	—

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

The screenshot shows the AWS S3 console interface. At the top, there's a summary bar with details about the bucket: Location (europe-west3 (Frankfurt)), Storage class (Standard), Public access (Not public), Protection (Soft Delete), and Hierarchical namespace (Enabled). Below this is a navigation bar with tabs: Objects, Configuration, Permissions, Protection, Lifecycle, Observability, New (which is highlighted in blue), and Inventory Reports.

The main area displays a folder browser. On the left, a tree view shows the structure: gk2-datalake > bronze > movies/ and ratings/. The right side shows a list of objects under the bronze folder. There are two entries:

Name	Type	Size	Created
movies/	Folder	—	—
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

The screenshot shows the Google Cloud navigation bar with 'Google Cloud' and a search bar. Below the navigation bar is a sidebar with pinned products (Cloud Hub, Cloud overview, Solutions, Recently visited) and a list of products (Billing, IAM & Admin, Marketplace, APIs & Services, Vertex AI, Compute Engine, Kubernetes Engine, Cloud Storage, Security, BigQuery). The 'IAM & Admin' item is highlighted with a red box. To the right of the sidebar is a main content area with sections for IAM, PAM, Security Insights (Preview), Principal Access Boundary, Identity & Organization, Policy Troubleshooter, Policy Analyzer, Organization Policies, Service Accounts (which is also highlighted with a red box), Workload Identity Federation, Workforce Identity Federation, Labels, Tags, Settings, Privacy & Security, Identity-Aware Proxy, Roles, Audit logs, Manage Resources, Create a Project, Asset Inventory, Quotas & System Limits, and Essential Contacts.

Click on Create Service Account

[← Create service account](#)

1 Create service account

Service account name
gk-access

Display name for this service account

Service account ID *
gk-access

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



Service account description

Describe what this service account will do

[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 IAM condition (optional) [?](#) [+ Add IAM condition](#) [Delete](#)

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

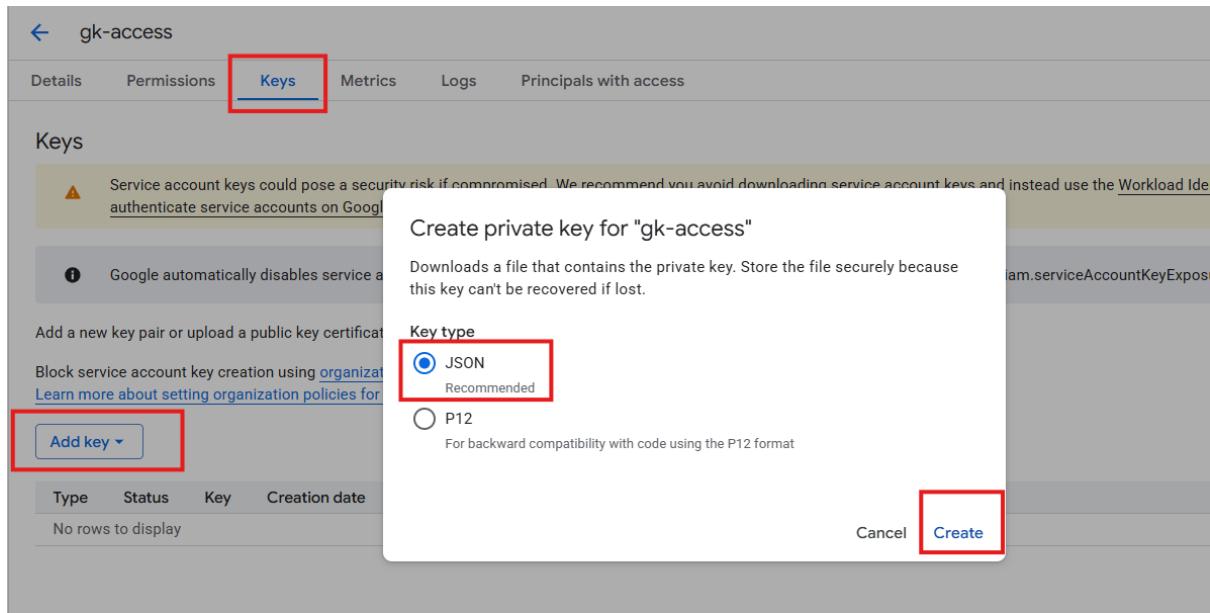
[+ Add another role](#)

[Help me choose roles](#)

[Continue](#)

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"/>	✉ gk-access@flink-demo-470113.iam.gserviceaccount.com	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