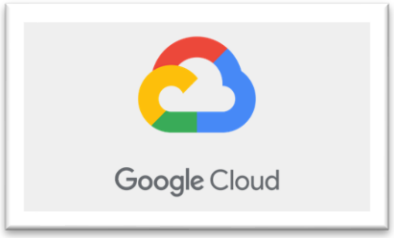
Analytics/Visuals on Employees from a data pipeline in GCP with BigQuery and Looker Studio



UNIVERSITY OF WOLLONGONG IN DUBAI

Lead Software Engineer

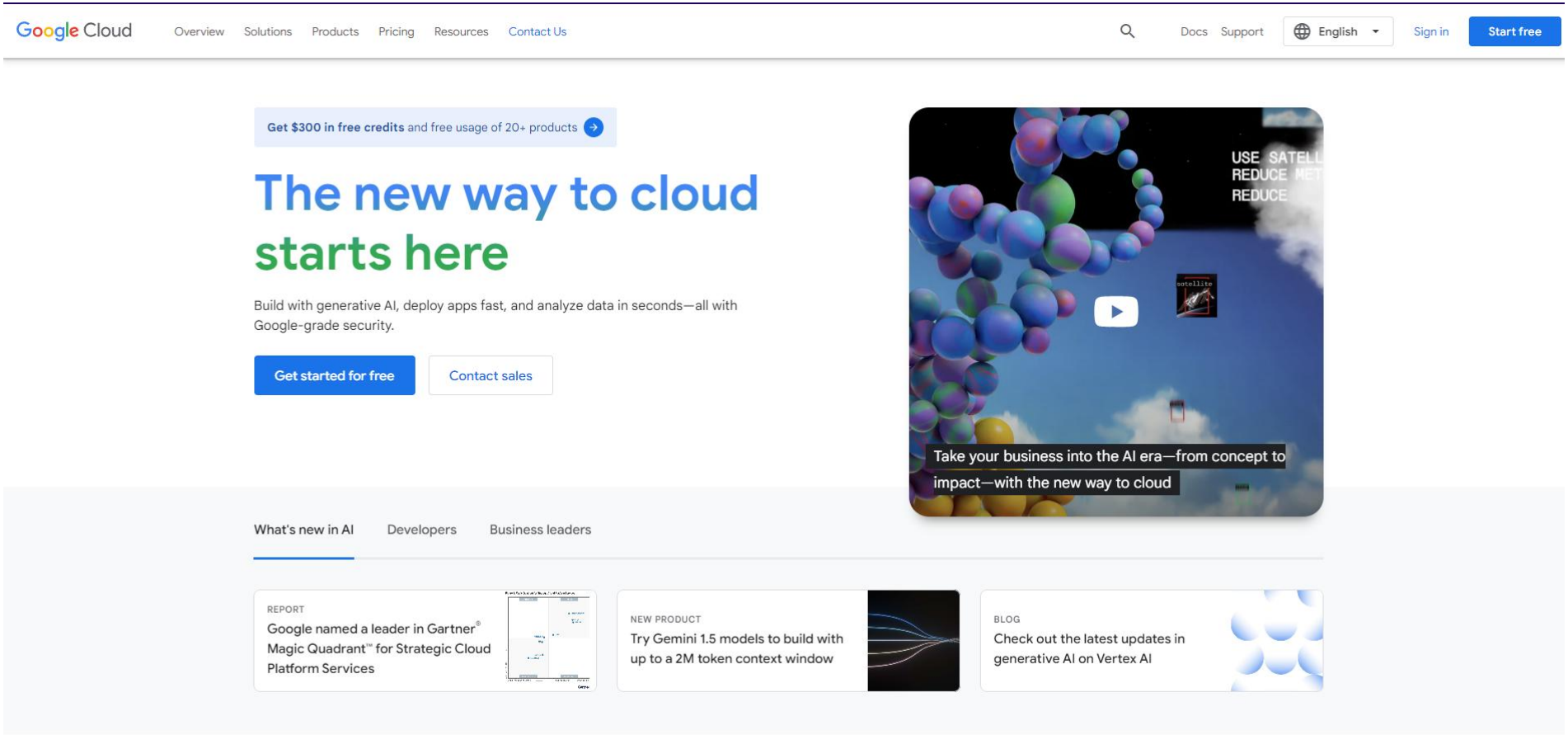
MD. Iqbal Mahmud

PurgeX

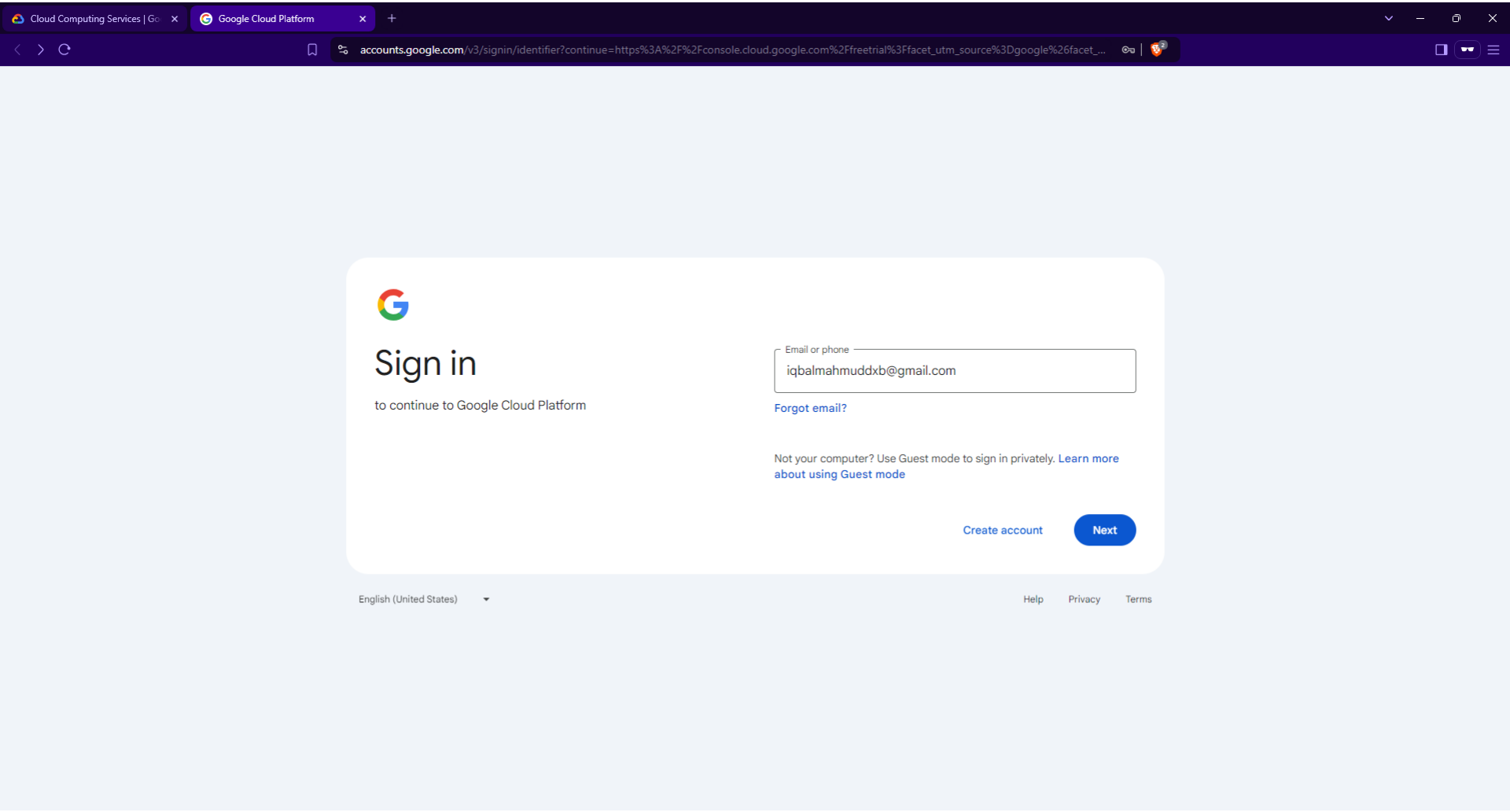
UNIVERSITY OF WOLLONGONG IN DUBAI

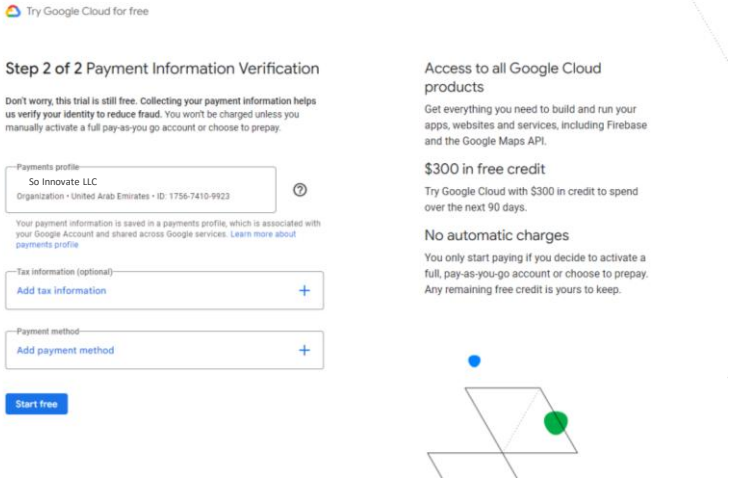
Agenda Overview

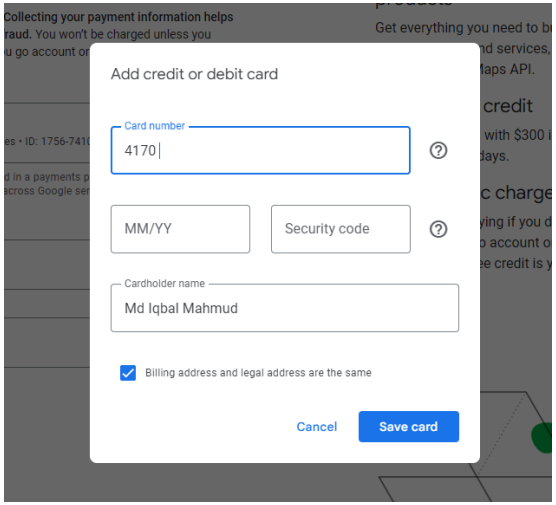
The objective of this project is to develop an automated data pipeline on Google Cloud Platform (GCP) that ingests data from a CSV file, processes it in BigQuery, and visualizes the results in Looker Studio. This pipeline will ensure that new data entries are added as new rows in BigQuery, and existing rows are updated whenever changes occur in the CSV file.

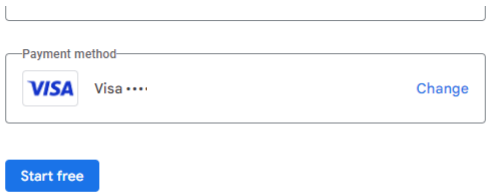

Google Cloud
Overview
Solutions
Products
Pricing
Resources
Contact Us
Docs
Support
English
Sign in
Start free
Get $300 in free credits and free usage of 20+ products
The new way to cloud starts here
Build with generative AI, deploy apps fast, and analyze data in seconds—all with Google-grade security.
Get started for free
Contact sales
Take your business into the AI era—from concept to impact—with the new way to cloud
What's new in AI
Developers
Business leaders
REPORT
Google named a leader in Gartner® Magic Quadrant™ for Strategic Cloud Platform Services
NEW PRODUCT
Try Gemini 1.5 models to build with up to a 2M token context window
BLOG
Check out the latest updates in generative AI on Vertex AI

UNIVERSITY OF WOLLONGONG IN DUBAI

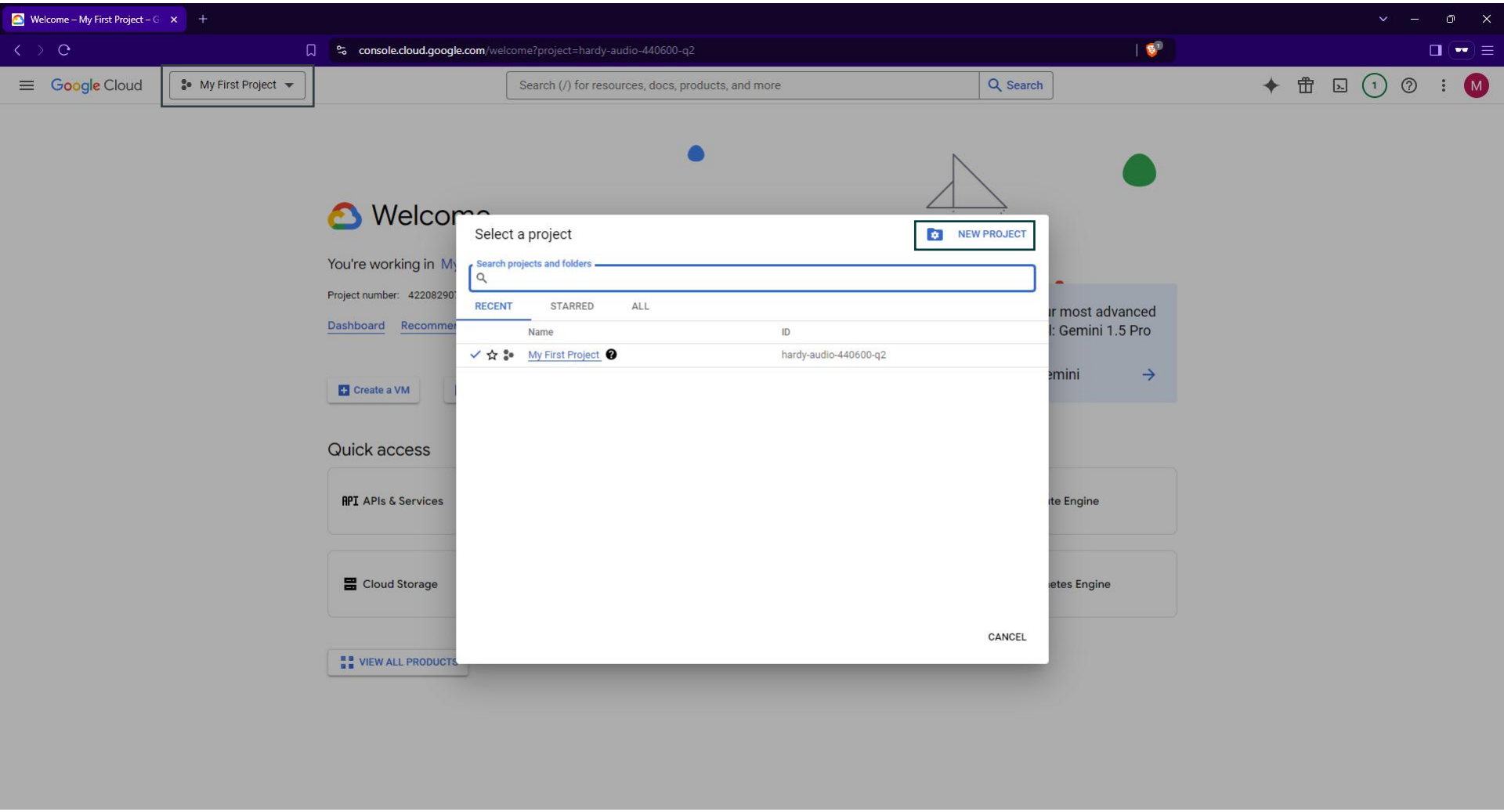



Try Google Cloud for free
Step 2 of 2 Payment Information Verification
Don't worry, this trial is still free. Collecting your payment information helps us verify your identity to reduce fraud. You won't be charged unless you manually activate a full pay-as-you go account or choose to prepay.
Payments profile
So Innovate LLC
Organization • United Arab Emirates • ID: 1756-7410-9923
Your payment information is saved in a payments profile, which is associated with your Google Account and shared across Google services. Learn more about payments profile
Tax information (optional)
Add tax information
Payment method
Add payment method
Start free
Access to all Google Cloud products
Get everything you need to build and run your apps, websites and services, including Firebase and the Google Maps API.
$300 in free credit
Try Google Cloud with $300 in credit to spend over the next 90 days.
No automatic charges
You only start paying if you decide to activate a full, pay-as-you-go account or choose to prepay. Any remaining free credit is yours to keep.

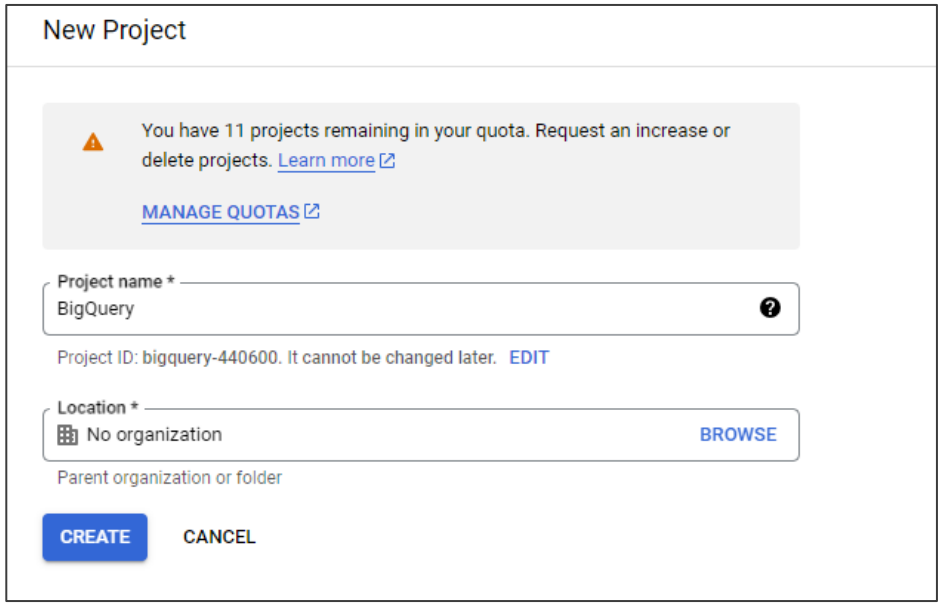

Add credit or debit card
Card number
4170
MM/YY
Security code
Cardholder name
Md Iqbal Mahmud
Billing address and legal address are the same
Cancel
Save card

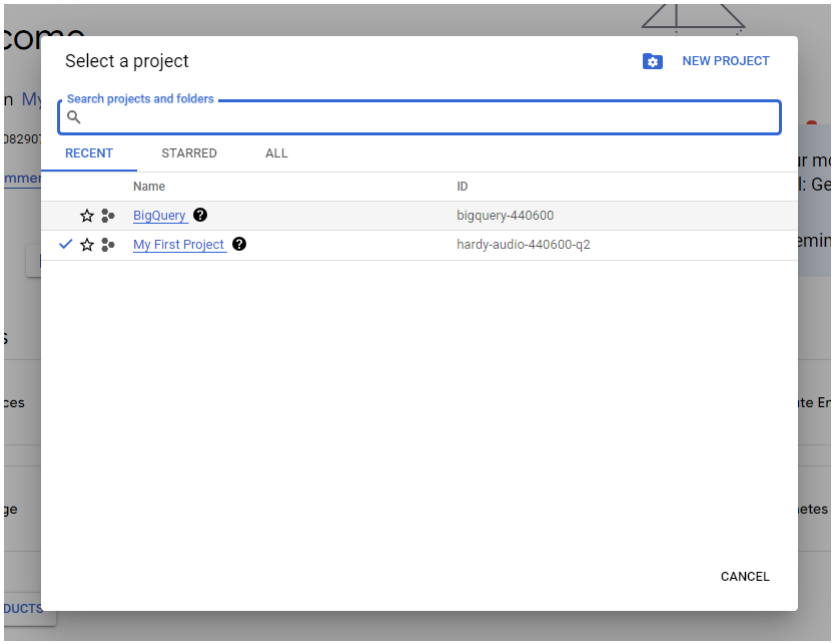

Payment method
VISA
Visa ••••
Change
Start free

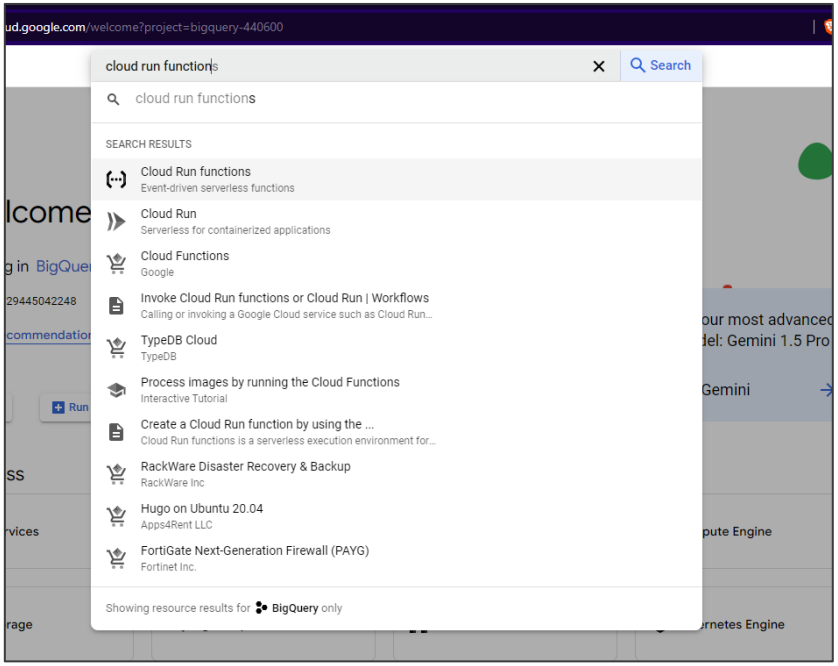
UNIVERSITY OF WOLLONGONG IN DUBAI

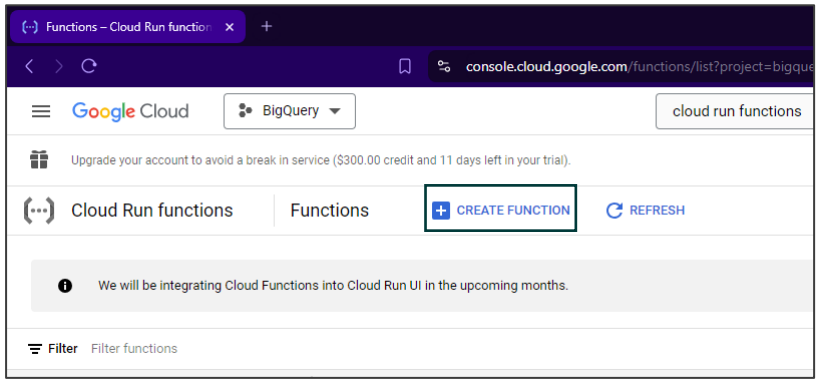


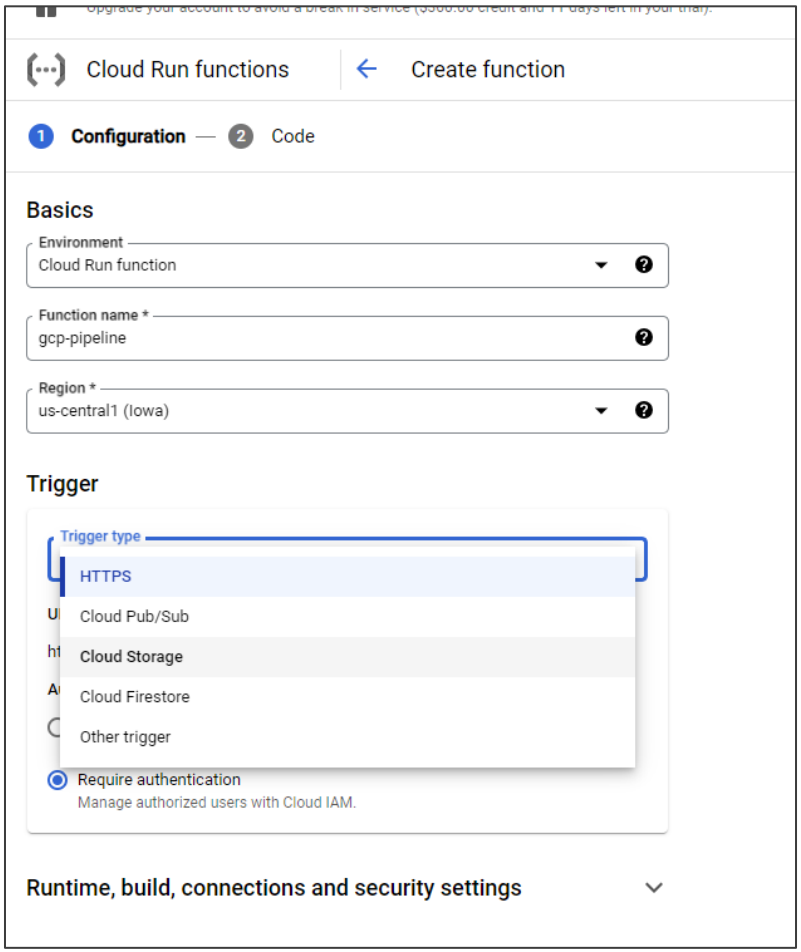
UNIVERSITY OF WOLLONGONG IN DUBAI

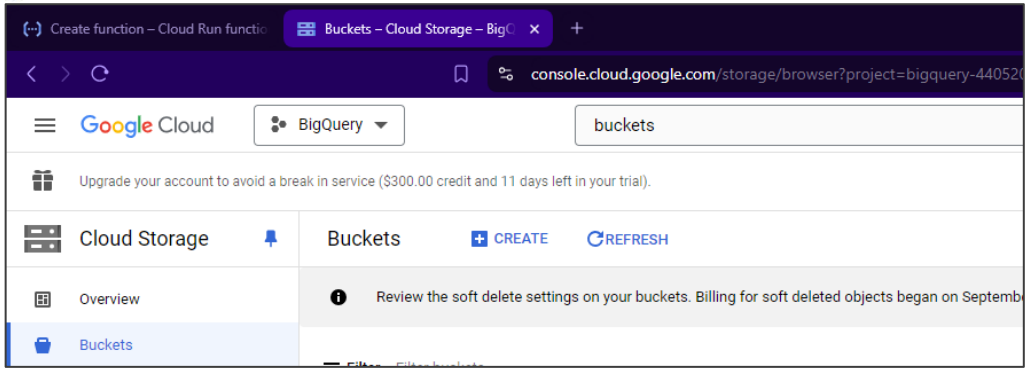


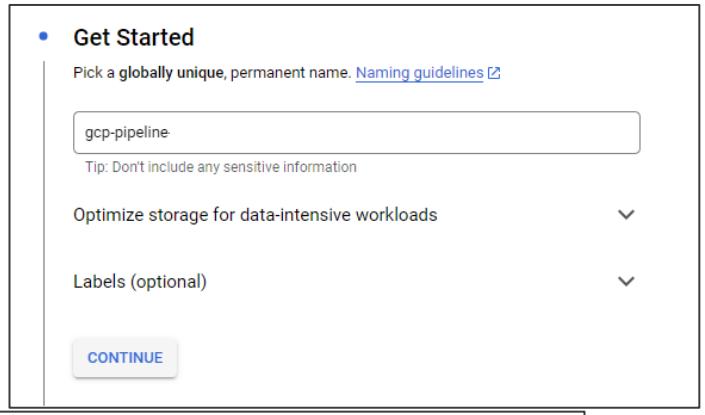


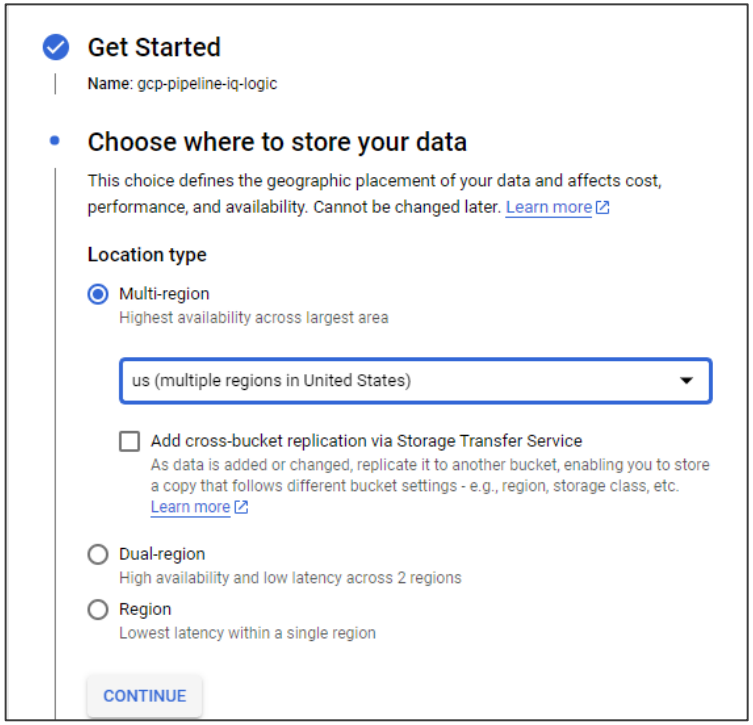


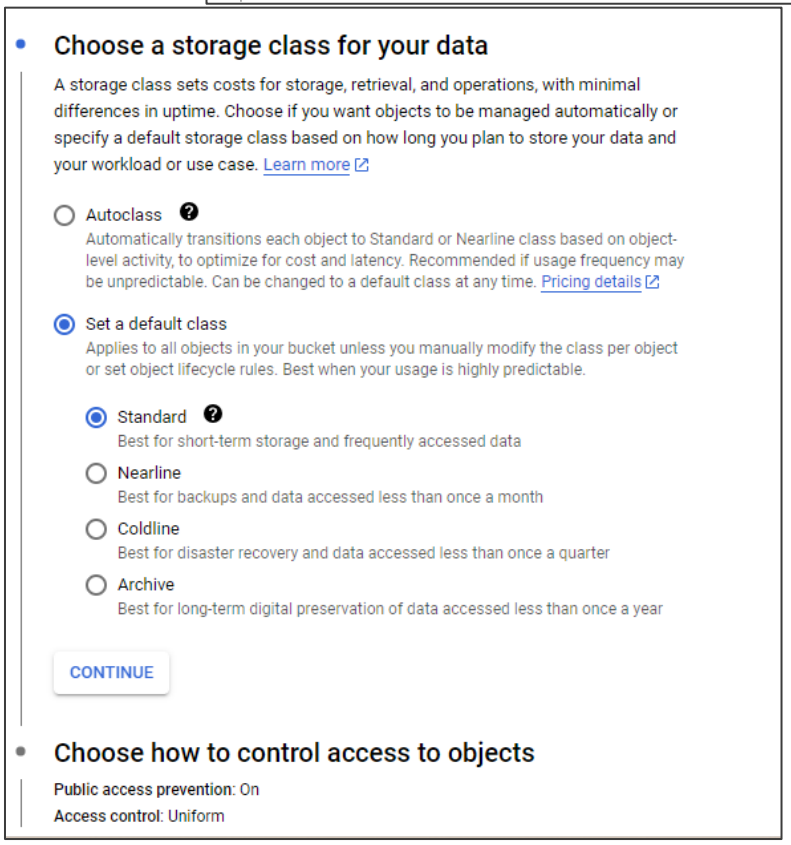

Functions – Cloud Run functio
console.cloud.google.com/functions/list?project=bigque
Google Cloud
BigQuery
cloud run functions
Upgrade your account to avoid a break in service ($300.00 credit and 11 days left in your trial).
Cloud Run functions
Functions
CREATE FUNCTION
REFRESH
We will be integrating Cloud Functions into Cloud Run UI in the upcoming months.
Filter
Filter functions

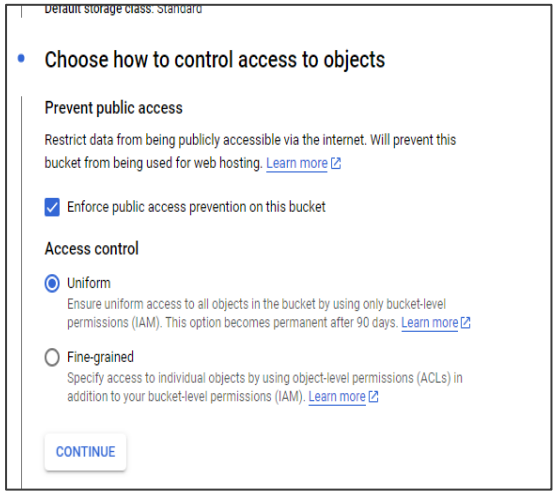

Cloud Run functions
Create function
1 Configuration — 2 Code
Basics
Environment
Cloud Run function
Function name *
gcp-pipeline
Region *
us-central1 (Iowa)
Trigger
Trigger type
HTTPS
Cloud Pub/Sub
Cloud Storage
Cloud Firestore
Other trigger
Require authentication
Manage authorized users with Cloud IAM.
Runtime, build, connections and security settings

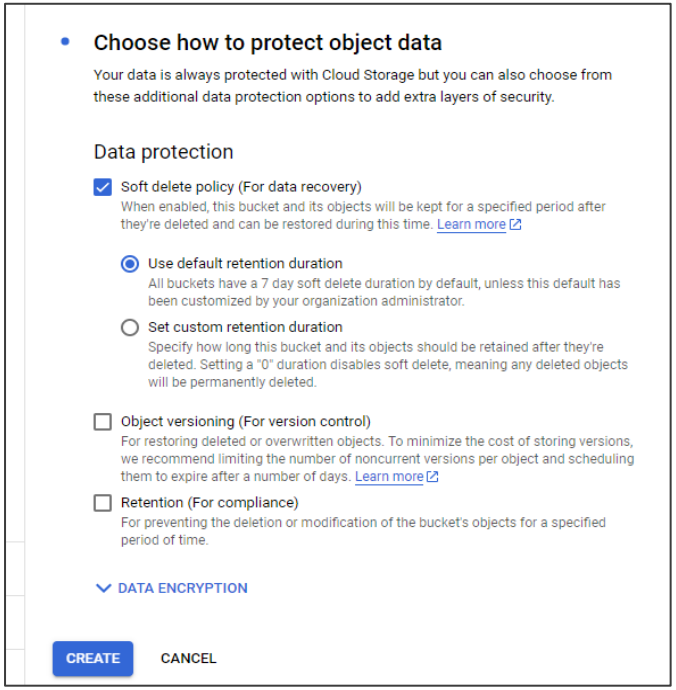

Create function – Cloud Run functio
Buckets – Cloud Storage – BigQ
console.cloud.google.com/storage/browser?project=bigquery-44052
Google Cloud
BigQuery
buckets
Upgrade your account to avoid a break in service ($300.00 credit and 11 days left in your trial).
Cloud Storage
Buckets
CREATE
REFRESH
Overview
Buckets
Review the soft delete settings on your buckets. Billing for soft deleted objects began on Septemb

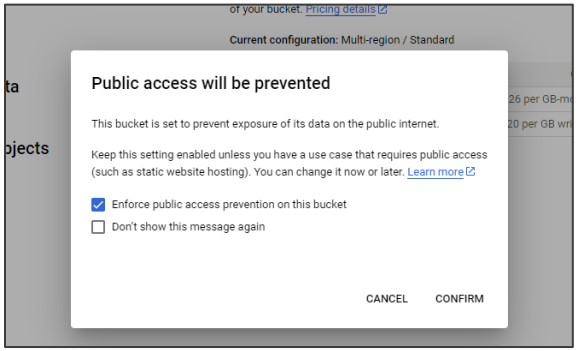

Get Started
Pick a globally unique, permanent name. Naming guidelines
gcp-pipeline
Tip: Don't include any sensitive information
Optimize storage for data-intensive workloads
Labels (optional)
CONTINUE

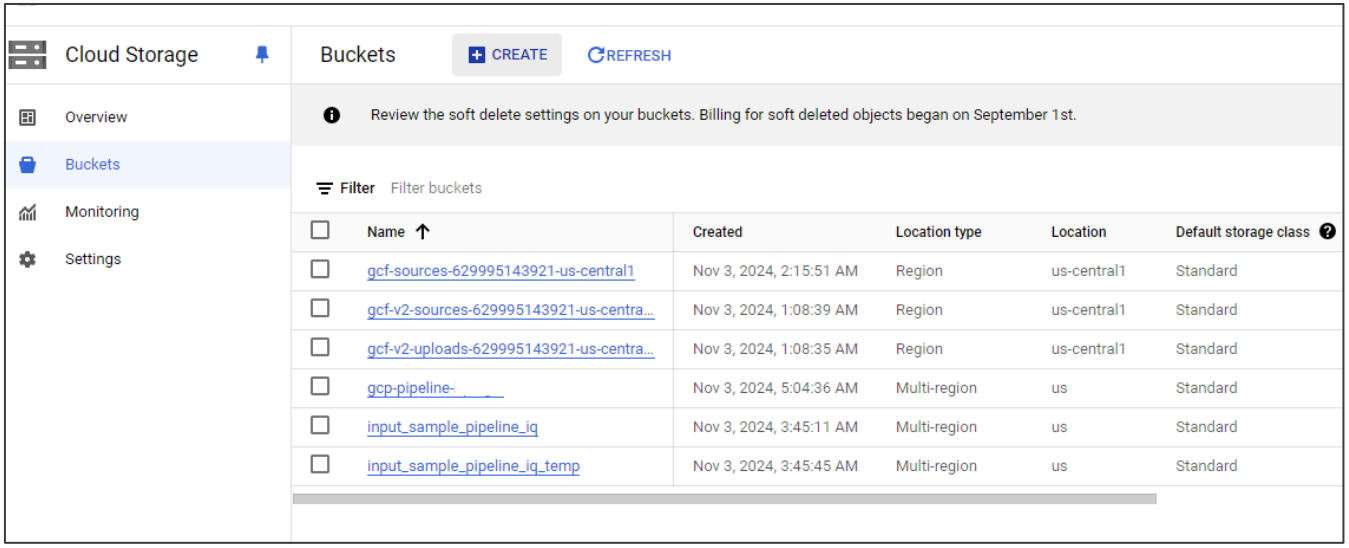

Get Started
Name: gcp-pipeline-iq-logic
Choose where to store your data
This choice defines the geographic placement of your data and affects cost, performance, and availability. Cannot be changed later. Learn more
Location type
Multi-region
Highest availability across largest area
us (multiple regions in United States)
Add cross-bucket replication via Storage Transfer Service
As data is added or changed, replicate it to another bucket, enabling you to store a copy that follows different bucket settings - e.g., region, storage class, etc.
Learn more
Dual-region
High availability and low latency across 2 regions
Region
Lowest latency within a single region
CONTINUE


Choose a storage class for your data
A storage class sets costs for storage, retrieval, and operations, with minimal differences in uptime. Choose if you want objects to be managed automatically or specify a default storage class based on how long you plan to store your data and your workload or use case. Learn more
Autoclass
Automatically transitions each object to Standard or Nearline class based on object-level activity, to optimize for cost and latency. Recommended if usage frequency may be unpredictable. Can be changed to a default class at any time. Pricing details
Set a default class
Applies to all objects in your bucket unless you manually modify the class per object or set object lifecycle rules. Best when your usage is highly predictable.
Standard
Best for short-term storage and frequently accessed data
Nearline
Best for backups and data accessed less than once a month
Coldline
Best for disaster recovery and data accessed less than once a quarter
Archive
Best for long-term digital preservation of data accessed less than once a year
CONTINUE
Choose how to control access to objects
Public access prevention: On
Access control: Uniform

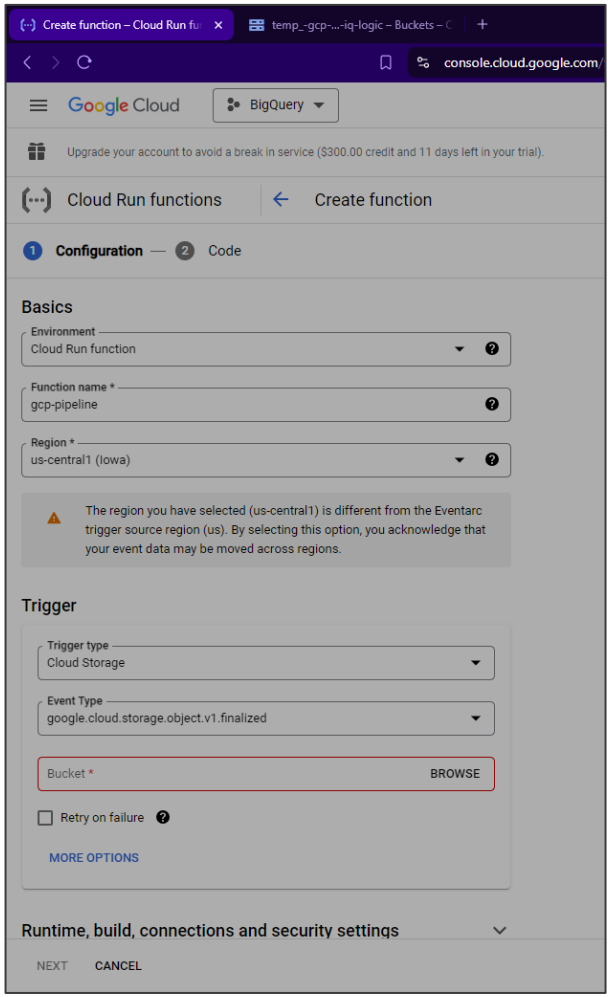

Default storage class: Standard
Choose how to control access to objects
Prevent public access
Restrict data from being publicly accessible via the internet. Will prevent this bucket from being used for web hosting. Learn more
Enforce public access prevention on this bucket
Access control
Uniform
Ensure uniform access to all objects in the bucket by using only bucket-level permissions (IAM). This option becomes permanent after 90 days. Learn more
Fine-grained
Specify access to individual objects by using object-level permissions (ACLs) in addition to your bucket-level permissions (IAM). Learn more
CONTINUE

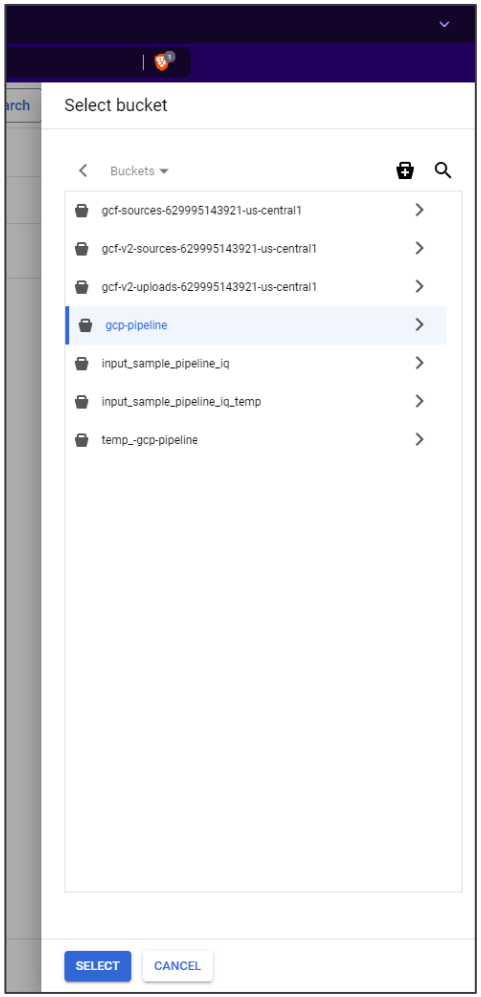

Choose how to protect object data
Your data is always protected with Cloud Storage but you can also choose from these additional data protection options to add extra layers of security.
Data protection
Soft delete policy (For data recovery)
When enabled, this bucket and its objects will be kept for a 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 versions, we recommend limiting the number of noncurrent versions per object and scheduling them to expire after a number of days. Learn more
Retention (For compliance)
For preventing the deletion or modification of the bucket's objects for a specified period of time.
DATA ENCRYPTION
CREATE
CANCEL

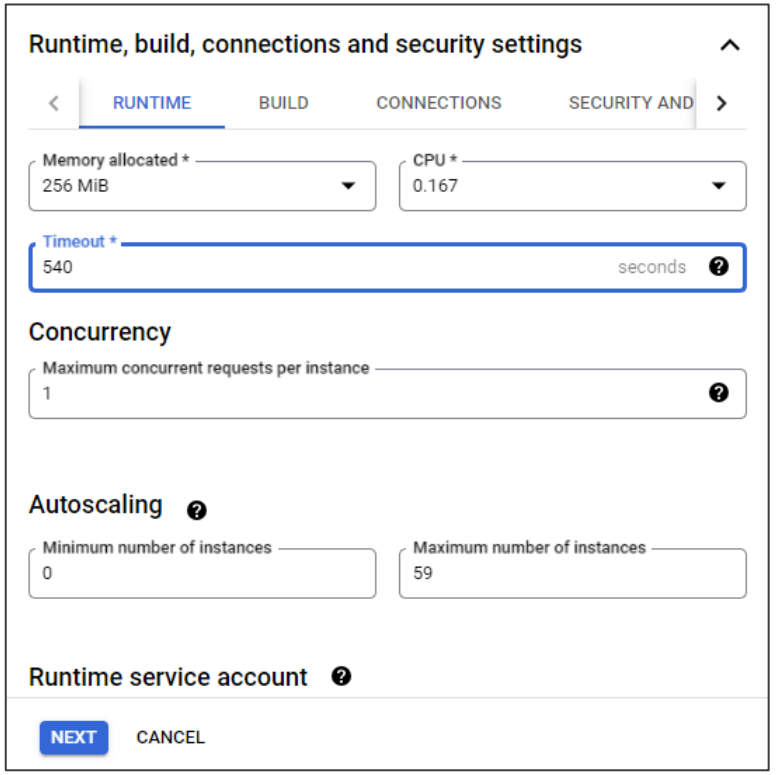

Current configuration: Multi-region / Standard
Public access will be prevented
This bucket is set to prevent exposure of its data on the public internet.
Keep this setting enabled unless you have a use case that requires public access (such as static website hosting). You can change it now or later. Learn more
Enforce public access prevention on this bucket
Don't show this message again
CANCEL
CONFIRM


Cloud Storage
Buckets
CREATE
REFRESH
Overview
Buckets
Monitoring
Settings
Review the soft delete settings on your buckets. Billing for soft deleted objects began on September 1st.
Filter
Filter buckets
Name
Created
Location type
Location
Default storage class
gcf-sources-629995143921-us-central1
Nov 3, 2024, 2:15:51 AM
Region
us-central1
Standard
gcf-v2-sources-629995143921-us-centra...
Nov 3, 2024, 1:08:39 AM
Region
us-central1
Standard
gcf-v2-uploads-629995143921-us-centra...
Nov 3, 2024, 1:08:35 AM
Region
us-central1
Standard
gcp-pipeline-
Nov 3, 2024, 5:04:36 AM
Multi-region
us
Standard
input_sample_pipeline_iq
Nov 3, 2024, 3:45:11 AM
Multi-region
us
Standard
input_sample_pipeline_iq_temp
Nov 3, 2024, 3:45:45 AM
Multi-region
us
Standard

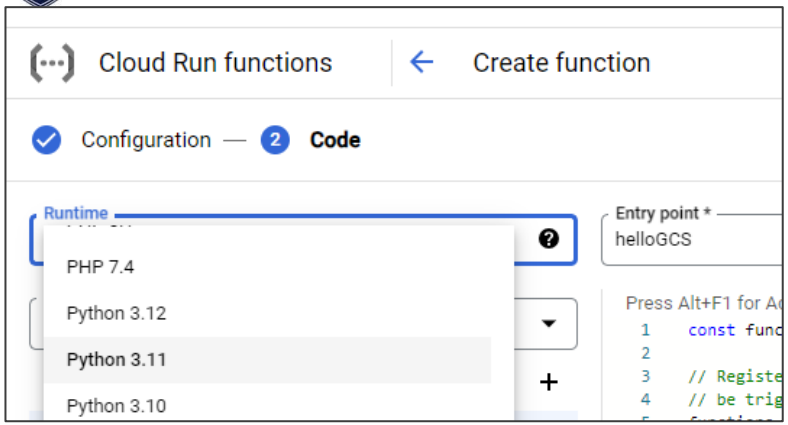
UNIVERSITY OF WOLLONGONG IN DUBAI

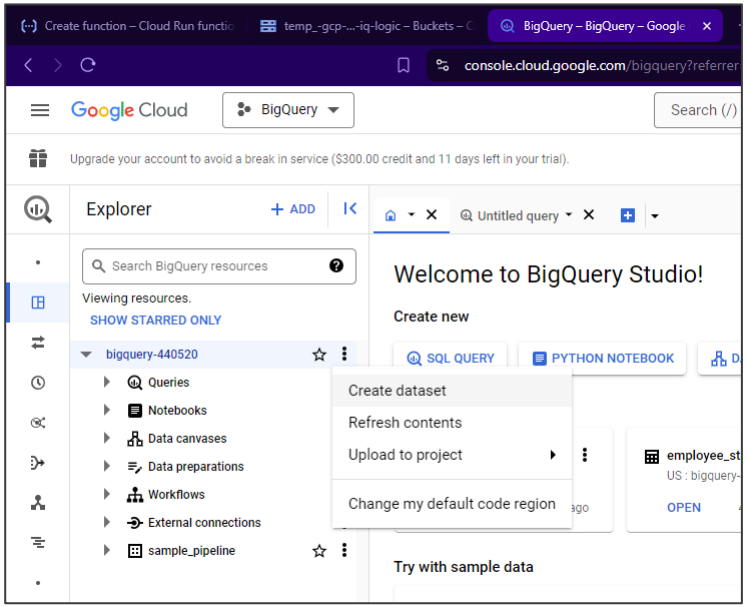


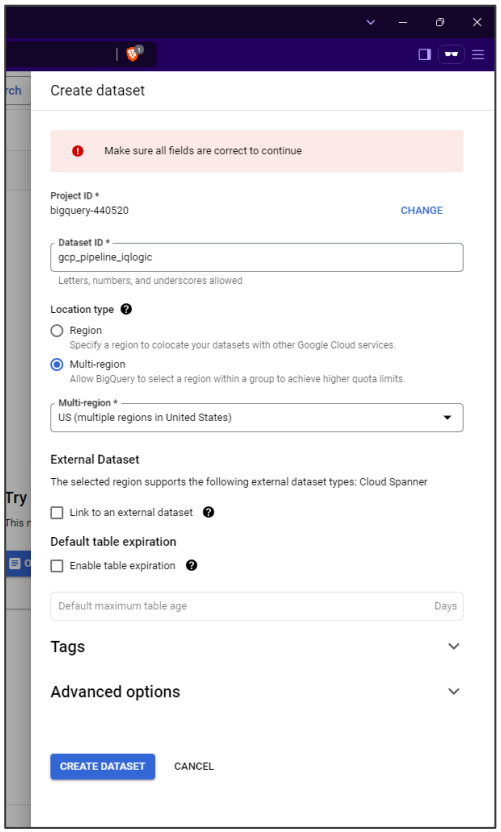


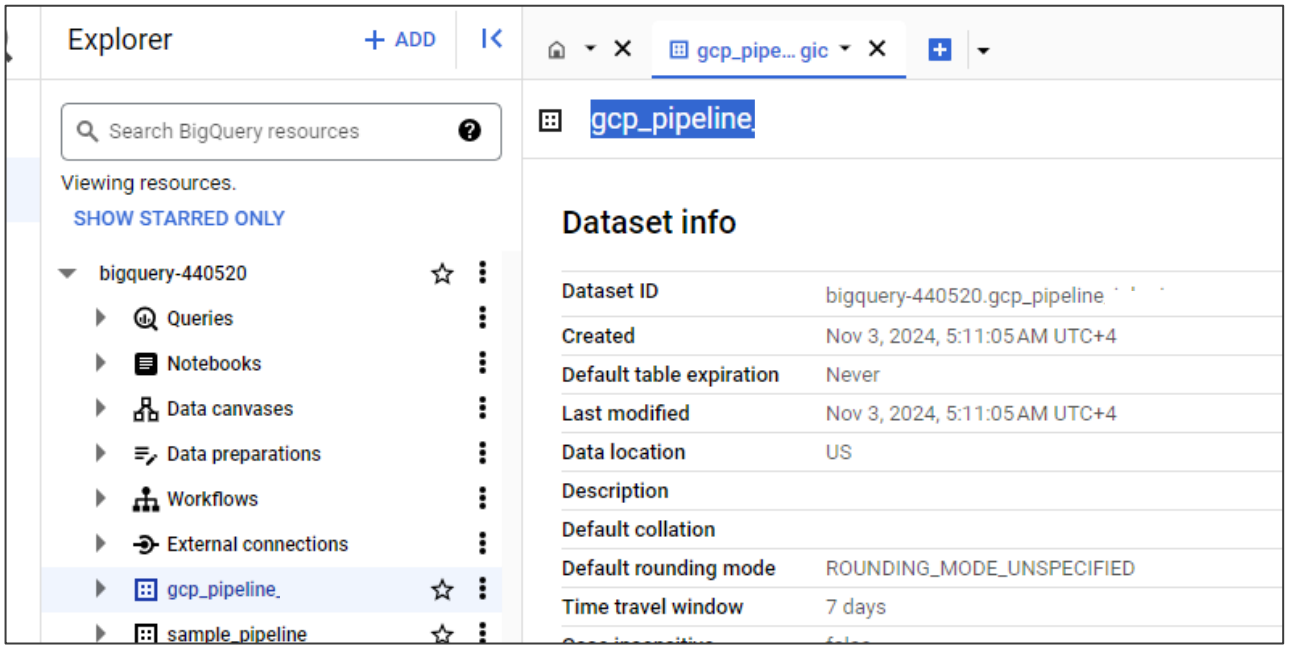


UNIVERSITY OF WOLLONGONG IN DUBAI









UNIVERSITY OF WOLLONGONG IN DUBAI

Google Cloud | BigQuery | cloud run functions | Search

Upgrade your account to avoid a break in service ($300.00 credit and 11 days left in your trial).

Cloud Run functions | Create function

Configuration — 2 **Code**

Runtime: Python 3.11

Entry point *: main

TEST FUNCTION

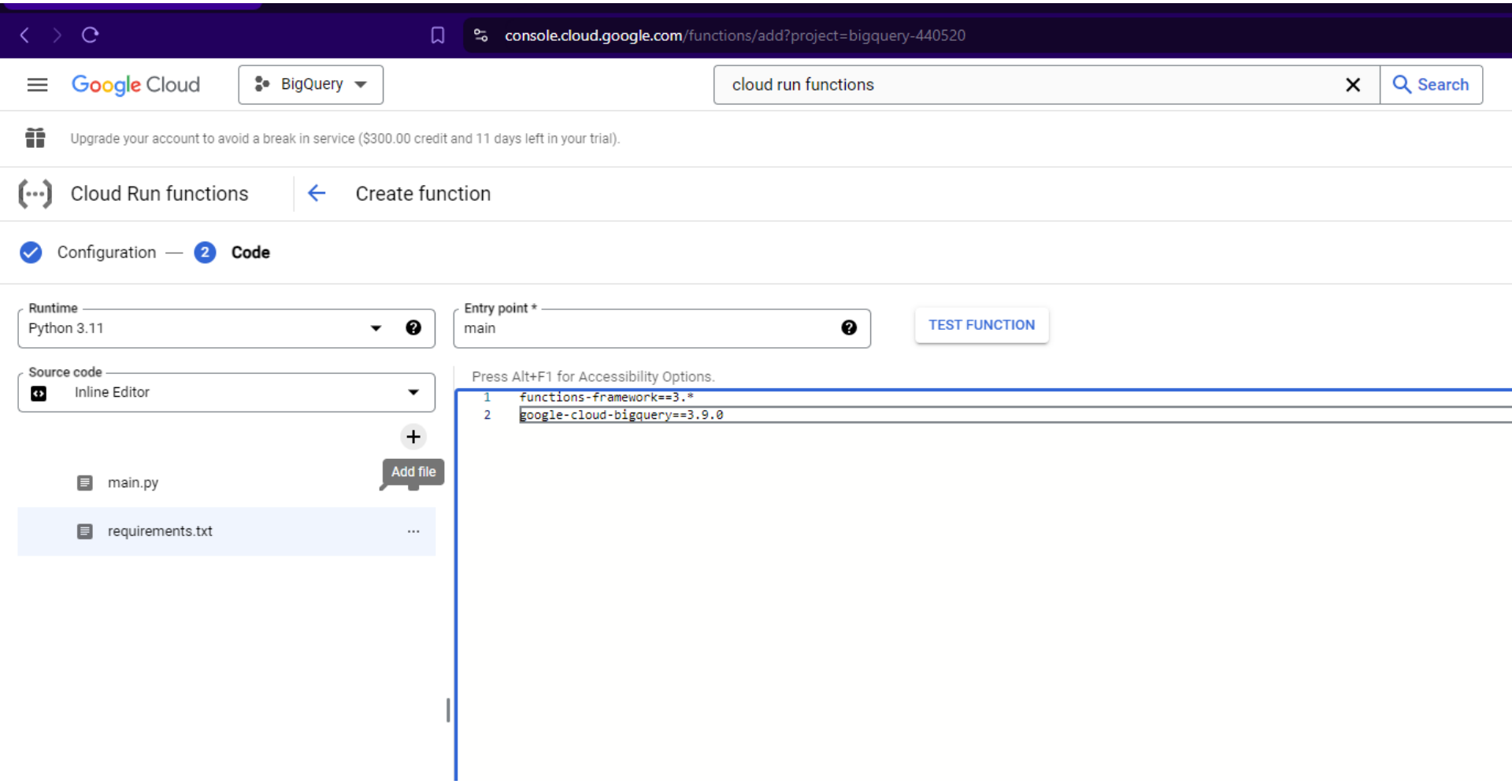
Source code: Inline Editor

main.py

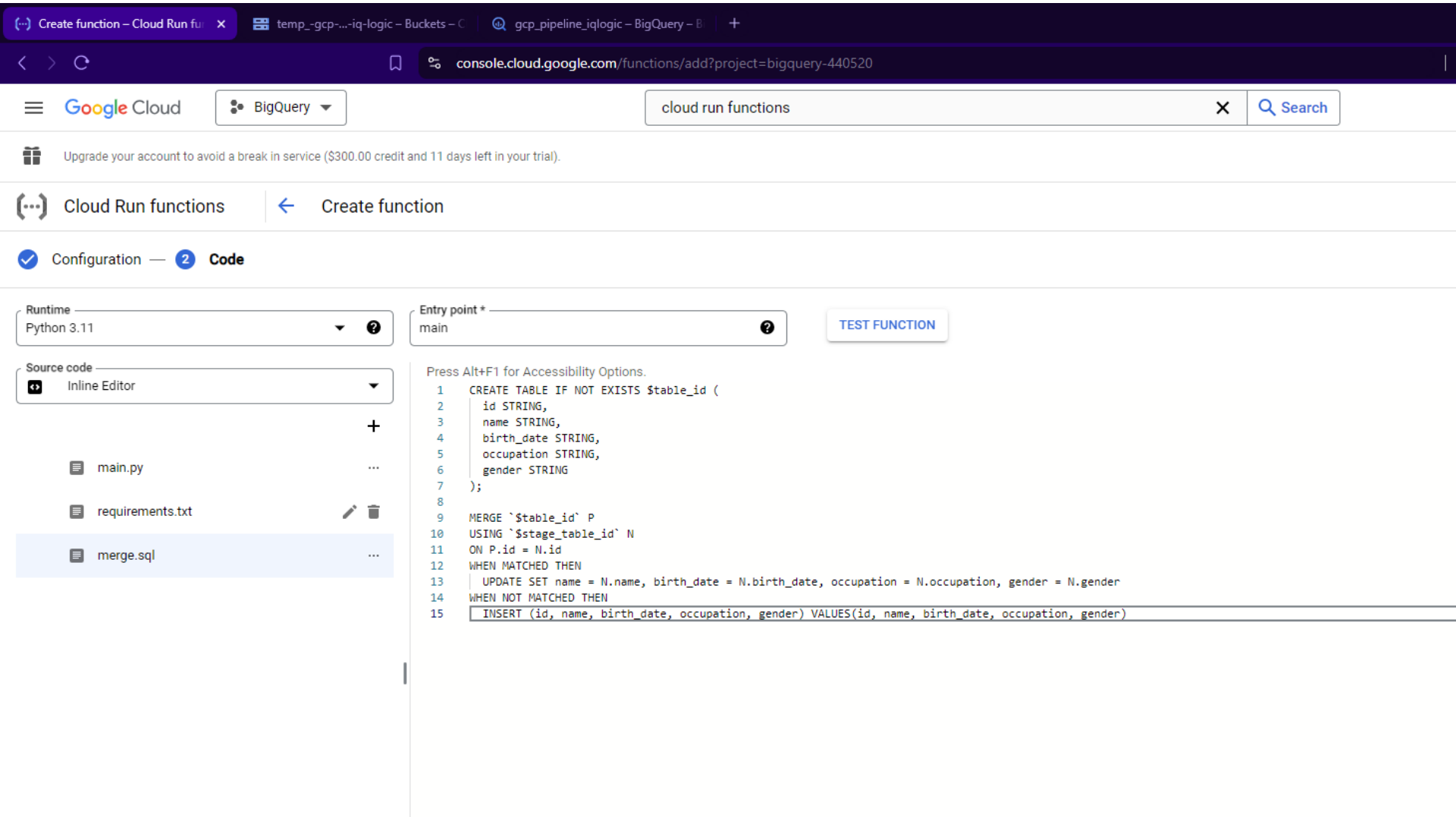
requirements.txt

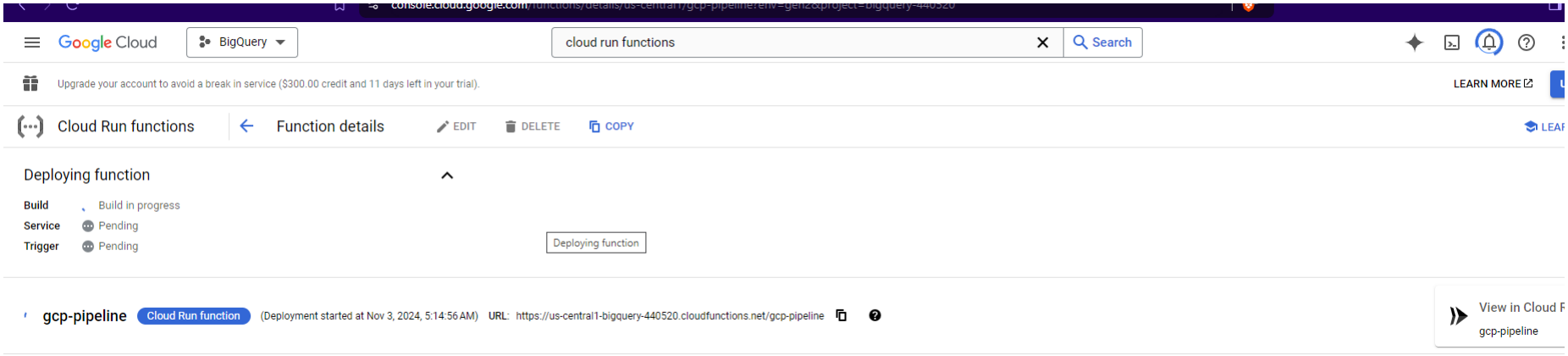
Press Alt+F1 for Accessibility Options.

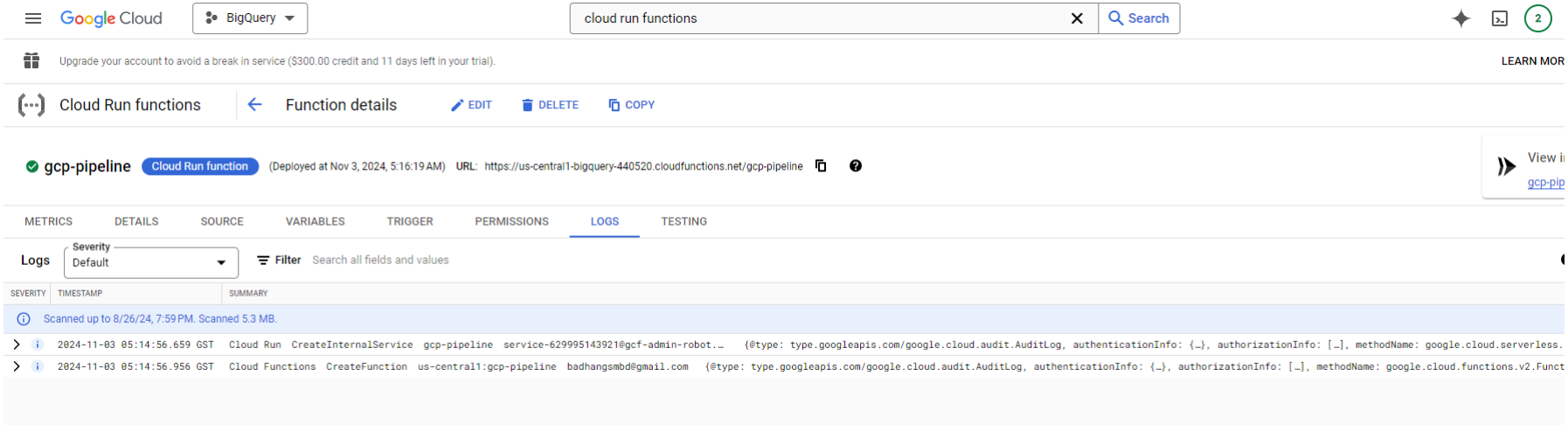
```python
import functions_framework
import re

# Libraries
from google.cloud import bigquery

# Construct a BigQuery client object.
client = bigquery.Client()

stage_table_id = "bigquery-440520.gcp_pipeline_____c.employee_stage"
table_id = "bigquery-440520.sample_pipeline.employee"

file_pattern = r".*sample_[0-9]{8}\.csv$"

table_schema = [
    bigquery.SchemaField("id", "STRING"),
    bigquery.SchemaField("name", "STRING"),
    bigquery.SchemaField("birth_date", "STRING"),
    bigquery.SchemaField("occupation", "STRING"),
    bigquery.SchemaField("gender", "STRING"),
]

# Triggered when file is created in storage bucket
@functions_framework.cloud_event
def main(cloud_event):
    data = cloud_event.data

    bucket = data["bucket"]
    name = data["name"]

    print(f"Bucket: {bucket}")
    print(f"File: {name}")

    if not re.match(file_pattern, name):
        print(f'{name} not match file_pattern')
        return

```

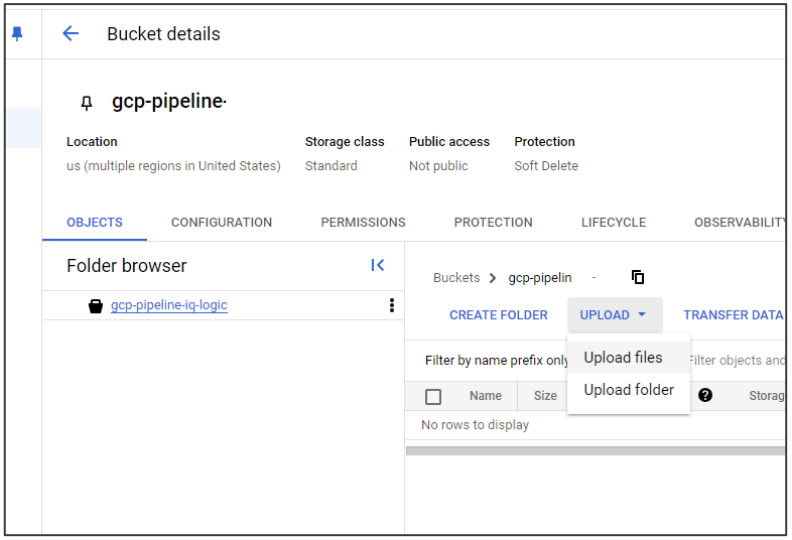

console.cloud.google.com/functions/add?project=bigquery-440520
Google Cloud
BigQuery
cloud run functions
Search
Upgrade your account to avoid a break in service ($300.00 credit and 11 days left in your trial).
Cloud Run functions
Create function
Configuration — 2 Code
Runtime
Python 3.11
Entry point *
main
TEST FUNCTION
Source code
Inline Editor
Press Alt+F1 for Accessibility Options.
1 functions-framework==3.*
2 google-cloud-bigquery==3.9.0
Add file
main.py
requirements.txt

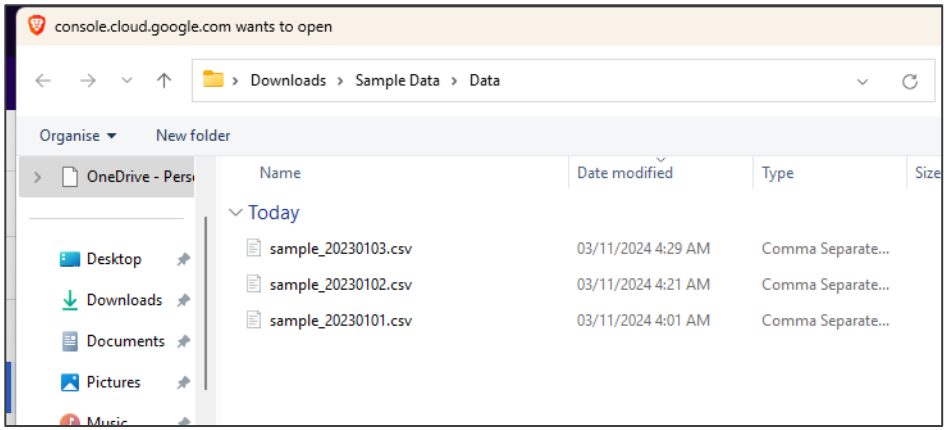
UNIVERSITY OF WOLLONGONG IN DUBAI

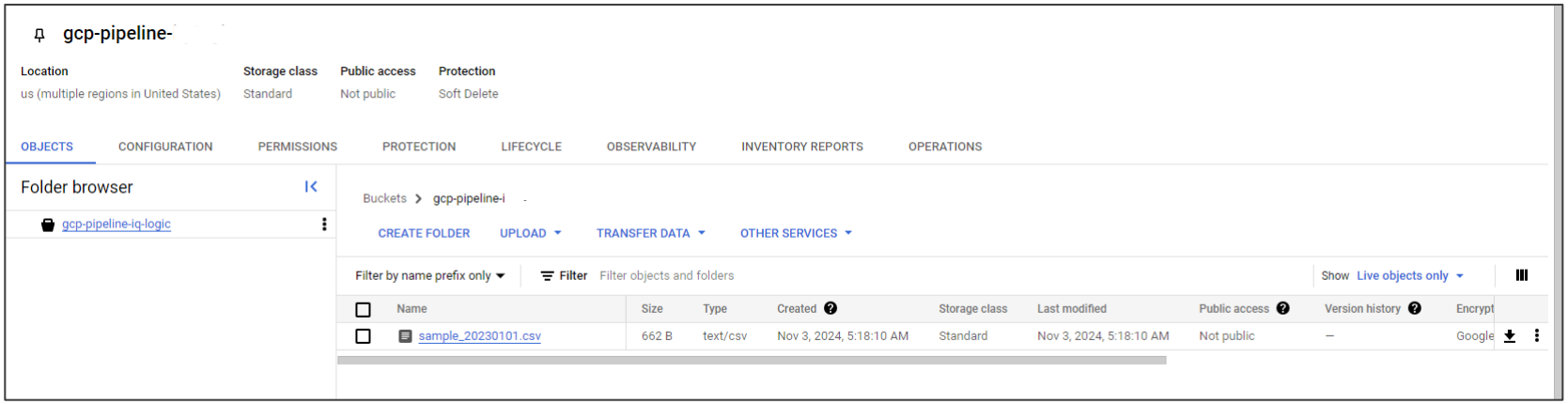



Google Cloud
BigQuery
cloud run functions
Search
Upgrade your account to avoid a break in service ($300.00 credit and 11 days left in your trial).
LEARN MORE
Cloud Run functions
Function details
EDIT
DELETE
COPY
Deploying function
Build
Build in progress
Service
Pending
Trigger
Pending
Deploying function
gcp-pipeline
Cloud Run function
(Deployment started at Nov 3, 2024, 5:14:56 AM)
URL: https://us-central1-bigquery-440520.cloudfunctions.net/gcp-pipeline
View in Cloud F
gcp-pipeline

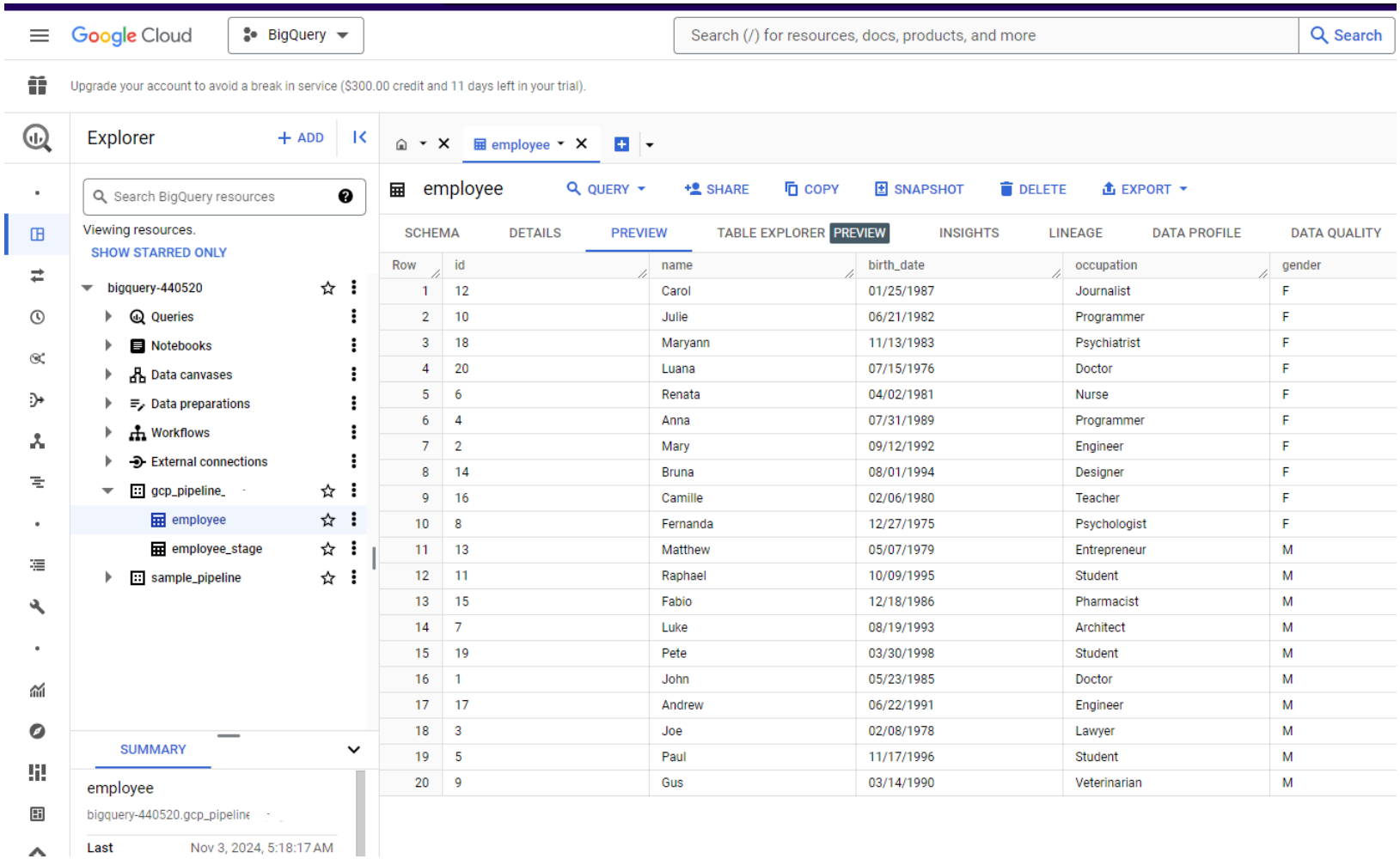

Google Cloud
BigQuery
cloud run functions
Search
2
Upgrade your account to avoid a break in service ($300.00 credit and 11 days left in your trial).
LEARN MOR
Cloud Run functions
Function details
EDIT
DELETE
COPY
gcp-pipeline
Cloud Run function
(Deployed at Nov 3, 2024, 5:16:19 AM)
URL: https://us-central1-bigquery-440520.cloudfunctions.net/gcp-pipeline
View i
gcp-pip
METRICS
DETAILS
SOURCE
VARIABLES
TRIGGER
PERMISSIONS
LOGS
TESTING
Logs
Severity
Default
Filter
Search all fields and values
SEVERITY
TIMESTAMP
SUMMARY
Scanned up to 8/26/24, 7:59 PM. Scanned 5.3 MB.
2024-11-03 05:14:56.659 GST Cloud Run CreateInternalService gcp-pipeline service-629995143921@gcf-admin-robot.… {@type: type.googleapis.com/google.cloud.audit.AuditLog, authenticationInfo: {…}, authorizationInfo: […], methodName: google.cloud.serverless.
2024-11-03 05:14:56.956 GST Cloud Functions CreateFunction us-central1:gcp-pipeline badhangsmbd@gmail.com {@type: type.googleapis.com/google.cloud.audit.AuditLog, authenticationInfo: {…}, authorizationInfo: […], methodName: google.cloud.functions.v2.Funct


Bucket details
gcp-pipeline-
Location
us (multiple regions in United States)
Storage class
Standard
Public access
Not public
Protection
Soft Delete
OBJECTS
CONFIGURATION
PERMISSIONS
PROTECTION
LIFECYCLE
OBSERVABILITY
Folder browser
gcp-pipeline-iq-logic
Buckets
gcp-pipelin
CREATE FOLDER
UPLOAD
TRANSFER DATA
Upload files
Upload folder
Filter by name prefix only
Name
Size
No rows to display

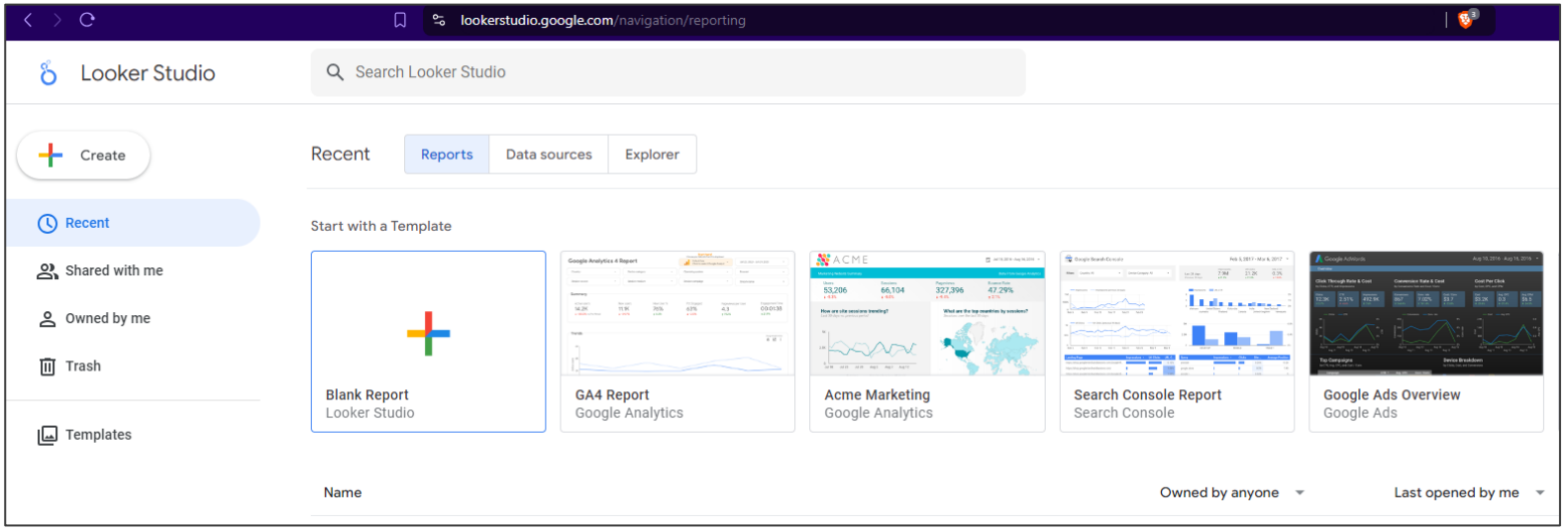

console.cloud.google.com wants to open
Downloads
Sample Data
Data
Organise
New folder
OneDrive - Pers
Desktop
Downloads
Documents
Pictures
Music
Name
Date modified
Type
Today
sample_20230103.csv
03/11/2024 4:29 AM
Comma Separate...
sample_20230102.csv
03/11/2024 4:21 AM
Comma Separate...
sample_20230101.csv
03/11/2024 4:01 AM
Comma Separate...

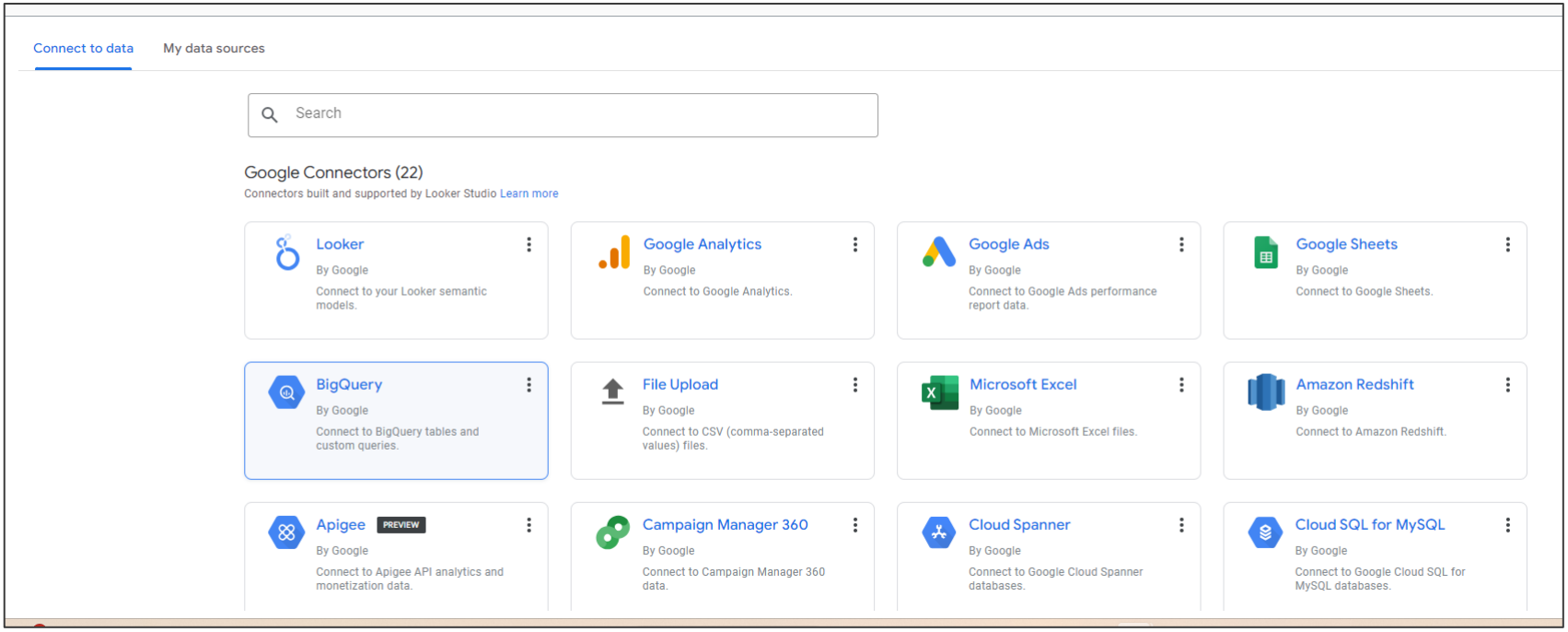

gcp-pipeline-
Location
us (multiple regions in United States)
Storage class
Standard
Public access
Not public
Protection
Soft Delete
OBJECTS
CONFIGURATION
PERMISSIONS
PROTECTION
LIFECYCLE
OBSERVABILITY
INVENTORY REPORTS
OPERATIONS
Folder browser
gcp-pipeline-iq-logic
Buckets
gcp-pipeline-i
CREATE FOLDER
UPLOAD
TRANSFER DATA
OTHER SERVICES
Filter by name prefix only
Filter
Filter objects and folders
Show
Live objects only
Name
Size
Type
Created
Storage class
Last modified
Public access
Version history
Encrypt
sample_20230101.csv
662 B
text/csv
Nov 3, 2024, 5:18:10 AM
Standard
Nov 3, 2024, 5:18:10 AM
Not public
–
Google

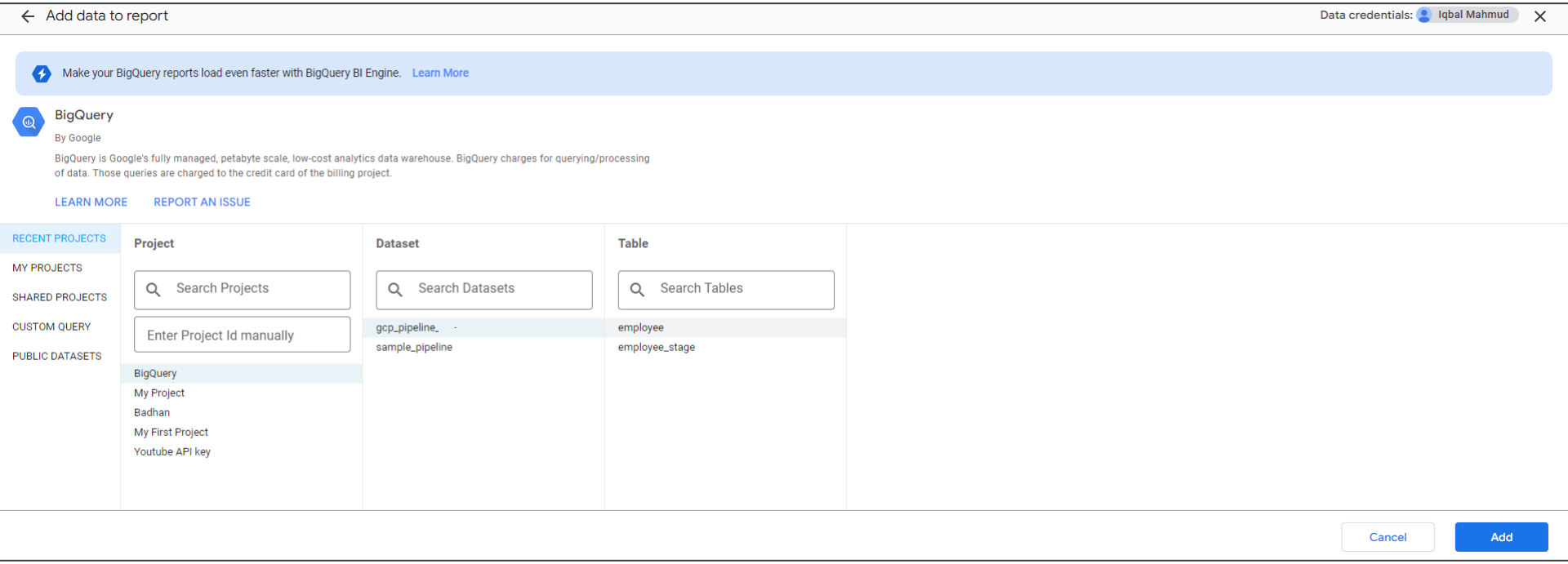
UNIVERSITY OF WOLLONGONG IN DUBAI

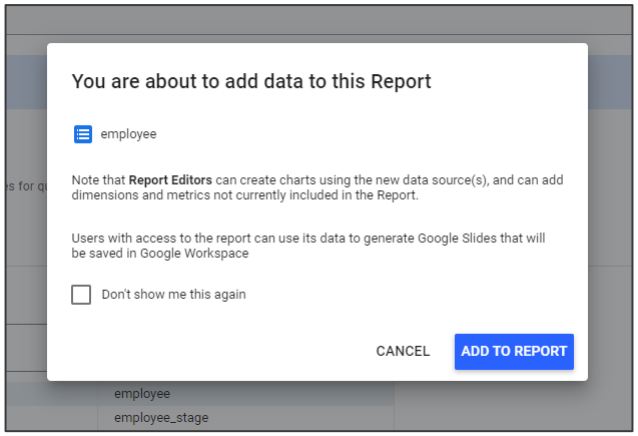


Row	id	name	birth_date	occupation	gender
1	12	Carol	01/25/1987	Journalist	F
2	10	Julie	06/21/1982	Programmer	F
3	18	Maryann	11/13/1983	Psychiatrist	F
4	20	Luana	07/15/1976	Doctor	F
5	6	Renata	04/02/1981	Nurse	F
6	4	Anna	07/31/1989	Programmer	F
7	2	Mary	09/12/1992	Engineer	F
8	14	Bruna	08/01/1994	Designer	F
9	16	Camille	02/06/1980	Teacher	F
10	8	Fernanda	12/27/1975	Psychologist	F
11	13	Matthew	05/07/1979	Entrepreneur	M
12	11	Raphael	10/09/1995	Student	M
13	15	Fabio	12/18/1986	Pharmacist	M
14	7	Luke	08/19/1993	Architect	M
15	19	Pete	03/30/1998	Student	M
16	1	John	05/23/1985	Doctor	M
17	17	Andrew	06/22/1991	Engineer	M
18	3	Joe	02/08/1978	Lawyer	M
19	5	Paul	11/17/1996	Student	M
20	9	Gus	03/14/1990	Veterinarian	M

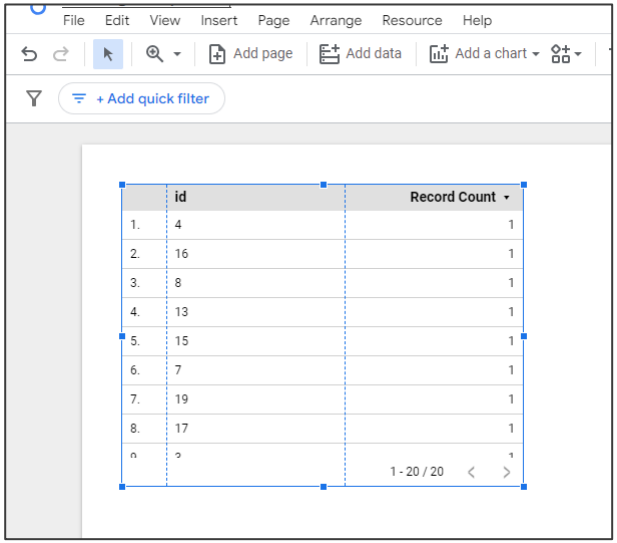

lookerstudio.google.com/navigation/reporting
Looker Studio
Search Looker Studio
Create
Recent
Shared with me
Owned by me
Trash
Templates
Recent
Reports
Data sources
Explorer
Start with a Template
Blank Report
Looker Studio
GA4 Report
Google Analytics
Acme Marketing
Google Analytics
Search Console Report
Search Console
Google Ads Overview
Google Ads
Name
Owned by anyone
Last opened by me

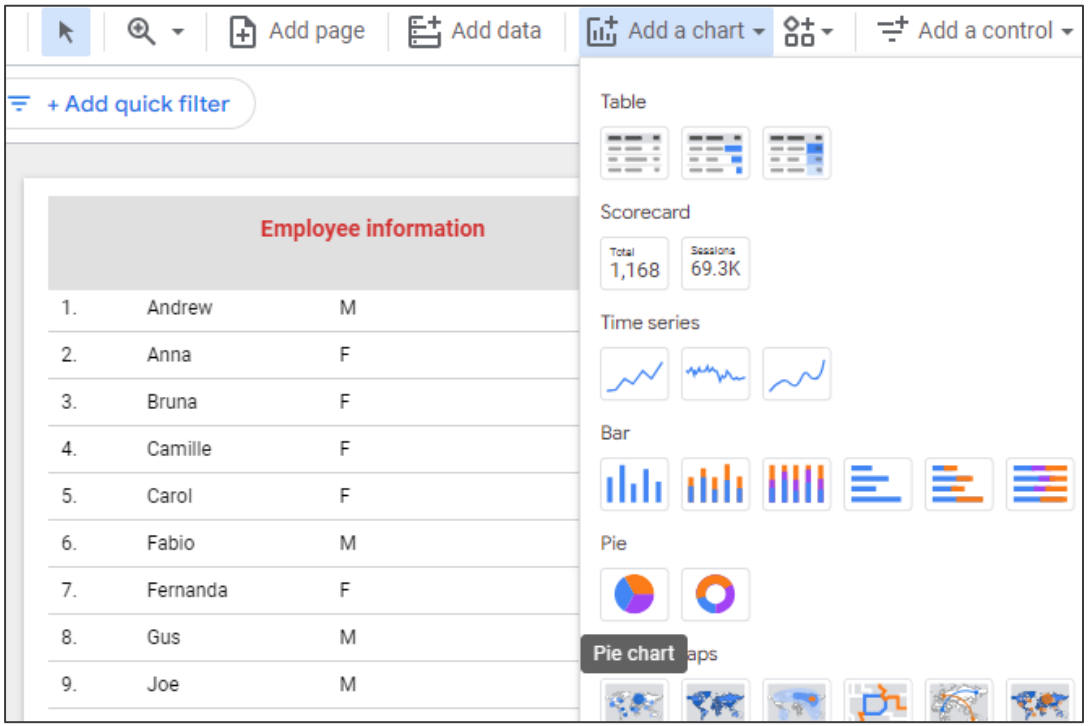

Connect to data
My data sources
Search
Google Connectors (22)
Connectors built and supported by Looker Studio Learn more
Looker
By Google
Connect to your Looker semantic models.
Google Analytics
By Google
Connect to Google Analytics.
Google Ads
By Google
Connect to Google Ads performance report data.
Google Sheets
By Google
Connect to Google Sheets.
BigQuery
By Google
Connect to BigQuery tables and custom queries.
File Upload
By Google
Connect to CSV (comma-separated values) files.
Microsoft Excel
By Google
Connect to Microsoft Excel files.
Amazon Redshift
By Google
Connect to Amazon Redshift.
Apigee
PREVIEW
By Google
Connect to Apigee API analytics and monetization data.
Campaign Manager 360
By Google
Connect to Campaign Manager 360 data.
Cloud Spanner
By Google
Connect to Google Cloud Spanner databases.
Cloud SQL for MySQL
By Google
Connect to Google Cloud SQL for MySQL databases.

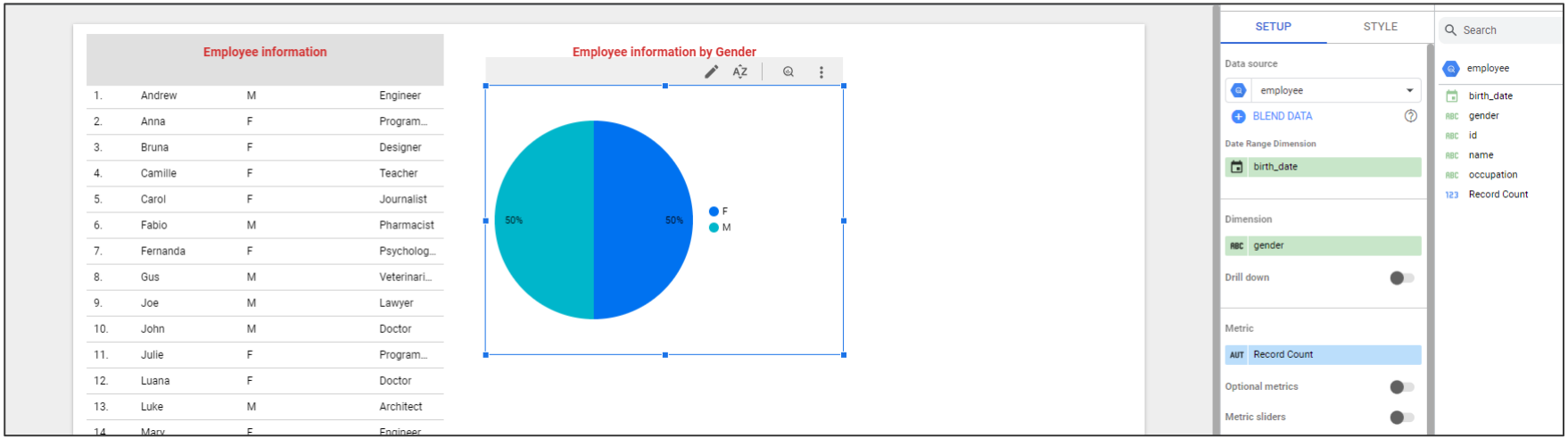

Add data to report
Data credentials: Iqbal Mahmud
Make your BigQuery reports load even faster with BigQuery BI Engine. Learn More
BigQuery
By Google
BigQuery is Google's fully managed, petabyte scale, low-cost analytics data warehouse. BigQuery charges for querying/processing of data. Those queries are charged to the credit card of the billing project.
LEARN MORE
REPORT AN ISSUE
RECENT PROJECTS
MY PROJECTS
SHARED PROJECTS
CUSTOM QUERY
PUBLIC DATASETS
Project
Search Projects
Enter Project Id manually
BigQuery
My Project
Badhan
My First Project
Youtube API key
Dataset
Search Datasets
gcp_pipeline_
sample_pipeline
Table
Search Tables
employee
employee_stage
Cancel
Add

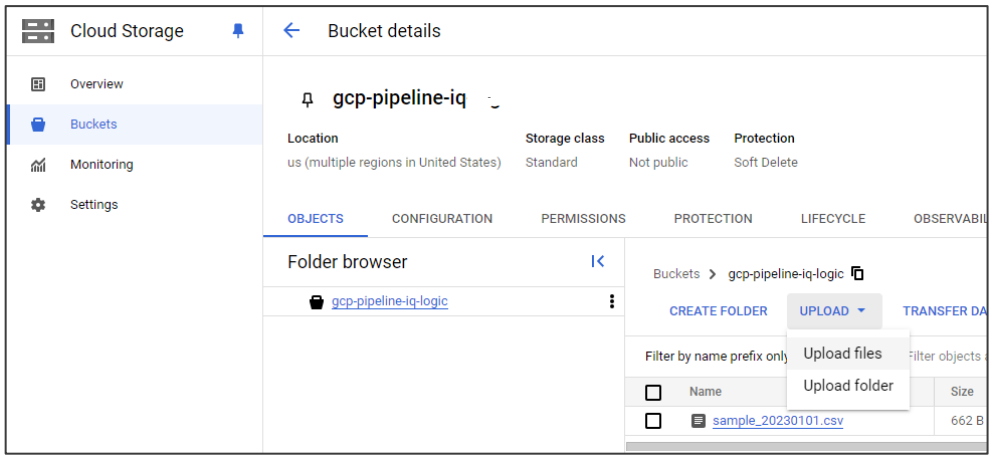

You are about to add data to this Report
employee
Note that Report Editors can create charts using the new data source(s), and can add dimensions and metrics not currently included in the Report.
Users with access to the report can use its data to generate Google Slides that will be saved in Google Workspace
Don't show me this again
CANCEL
ADD TO REPORT
employee
employee_stage

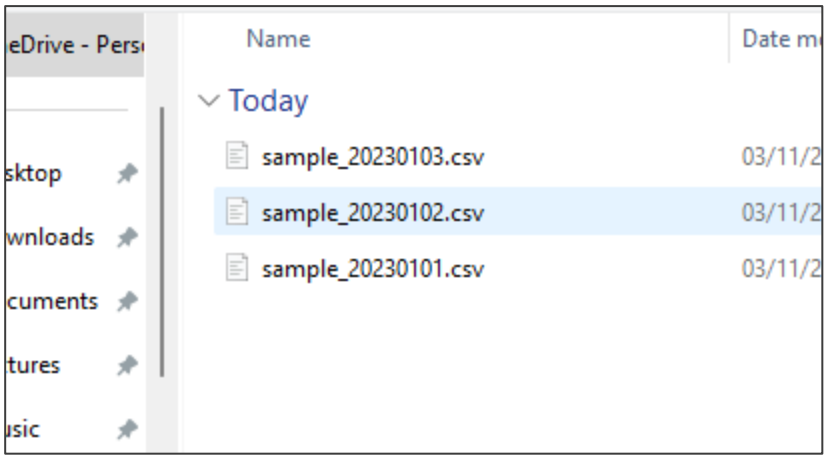
UNIVERSITY OF WOLLONGONG IN DUBAI

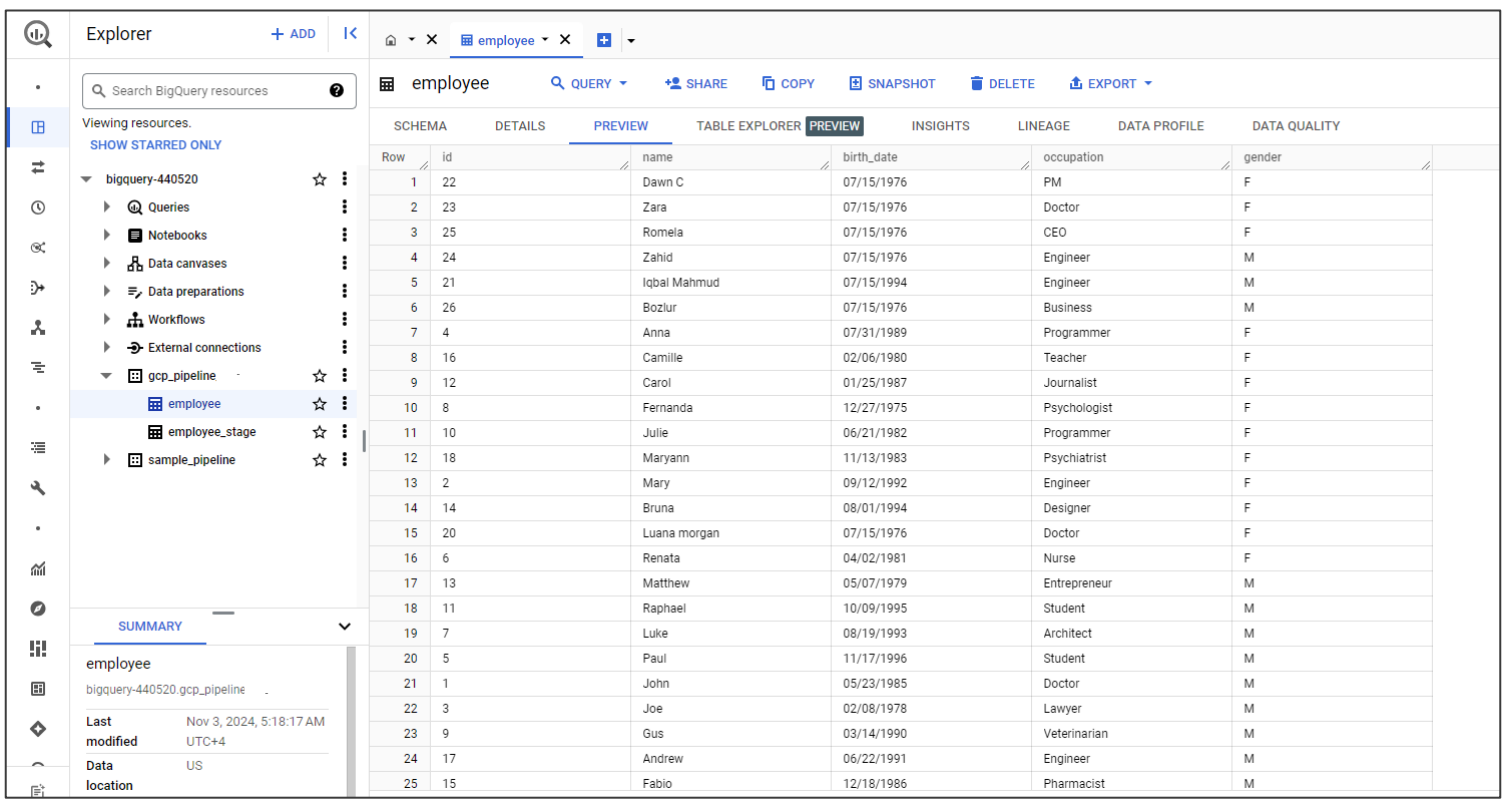












Row	id	name	birth_date	occupation	gender
1	22	Dawn C	07/15/1976	PM	F
2	23	Zara	07/15/1976	Doctor	F
3	25	Romela	07/15/1976	CEO	F
4	24	Zahid	07/15/1976	Engineer	M
5	21	Iqbal Mahmud	07/15/1994	Engineer	M
6	26	Bozlur	07/15/1976	Business	M
7	4	Anna	07/31/1989	Programmer	F
8	16	Camille	02/06/1980	Teacher	F
9	12	Carol	01/25/1987	Journalist	F
10	8	Fernanda	12/27/1975	Psychologist	F
11	10	Julie	06/21/1982	Programmer	F
12	18	Maryann	11/13/1983	Psychiatrist	F
13	2	Mary	09/12/1992	Engineer	F
14	14	Bruna	08/01/1994	Designer	F
15	20	Luana morgan	07/15/1976	Doctor	F
16	6	Renata	04/02/1981	Nurse	F
17	13	Matthew	05/07/1979	Entrepreneur	M
18	11	Raphael	10/09/1995	Student	M
19	7	Luke	08/19/1993	Architect	M
20	5	Paul	11/17/1996	Student	M
21	1	John	05/23/1985	Doctor	M
22	3	Joe	02/08/1978	Lawyer	M
23	9	Gus	03/14/1990	Veterinarian	M
24	17	Andrew	06/22/1991	Engineer	M
25	15	Fabio	12/18/1986	Pharmacist	M

File Edit View Insert Page Arrange Resource Help

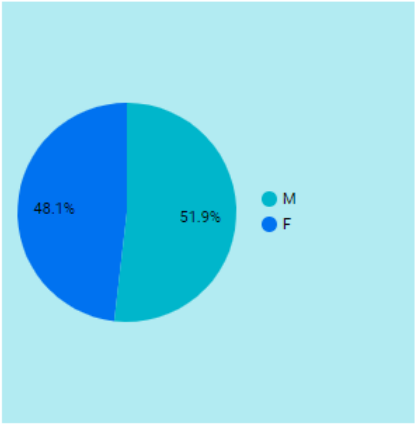
Add page | Add data | Add a chart | Add a control | Theme and layout

+ Add quick filter | Reset

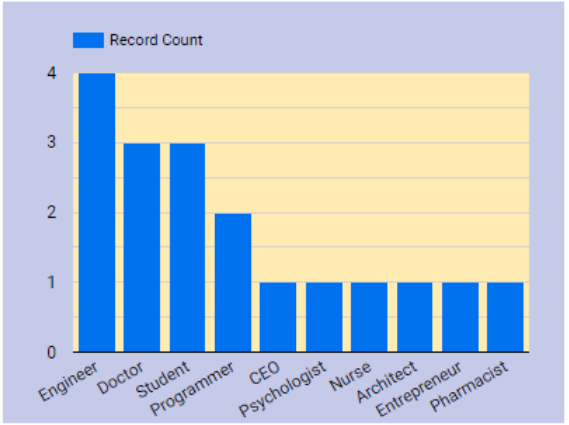
Employee information

1.	Andrew	M	Engineer
2.	Anna	F	Program...
3.	Bozlur	M	Business
4.	Bruna	F	Designer
5.	Camille	F	Teacher
6.	Carol	F	Journalist
7.	Dawn C	F	PM
8.	Fabio	M	Pharmacist
9.	Fernanda	F	Psycholog...
10.	Gus	M	Veterinari...
11.	Iqbal Mahmud	M	Engineer
12.	Joe	M	Lawyer
13.	John	M	Doctor
14.	Julie	F	Program...
15.	Luana morgan	F	Doctor
16.	Luke	M	Architect
17.	Mary	F	Engineer
18.	Maryann	F	Psychiatrist
19.	Matthew	M	Entrepren...
20.	Nazmul H	M	Private
21.	Paul	M	Student

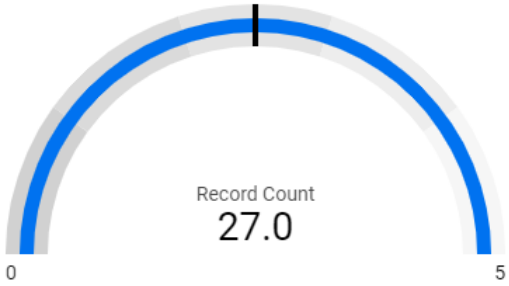
Employee information by Gender



Employee information by Gender



Total Employee



UNIVERSITY OF WOLLONGONG IN DUBAI

Thank You

PRACTICE:

1. SETUP YOUE ENVIRONMENT
2. DOWNLOAD CONFIGURATION FILES FROM

https://github.com/iqbalmbadhan/gcp_bigquery_looker

3. PICK A FILE FROM ANY PUBLIC DATASET
4. BUILD YOUR OWN DASHBOARD