FINAL PROJECT

GOOGLE CLOUD PLATFORM

PROJECT TOPICS:

- Create two buckets.
- Load the data one bucket to another bucket using cloud function.
- Create bigquery.
- Load the data from bucket into the bigquery using dataflow.

A Final Project Report Submitted in the fulfilment of the requirements for Final-Project Evaluation.

Submitted by

Anji L (2320471)
Ayyasamy M (2320474)
Bharath Kumar B (2320821)
Mushi Sai B (2320746)
Madhusri (2320460)
Venkata Swamy P(2320477)

COHORT ID: CSDAIA24GP003

cognizant

ABSTRACTION:

This project focuses on automating the process of ingesting, processing, and loading data from Google Cloud Storage buckets into BigQuery using Dataflow and Cloud Functions. The workflow involves triggering Cloud Functions in response to new files being uploaded to a bucket, which then initiates a Dataflow job to transform and load the data into BigQuery. The project aims to demonstrate efficient data orchestration techniques in a cloud environment, showcasing the integration of various Google Cloud Platform services.

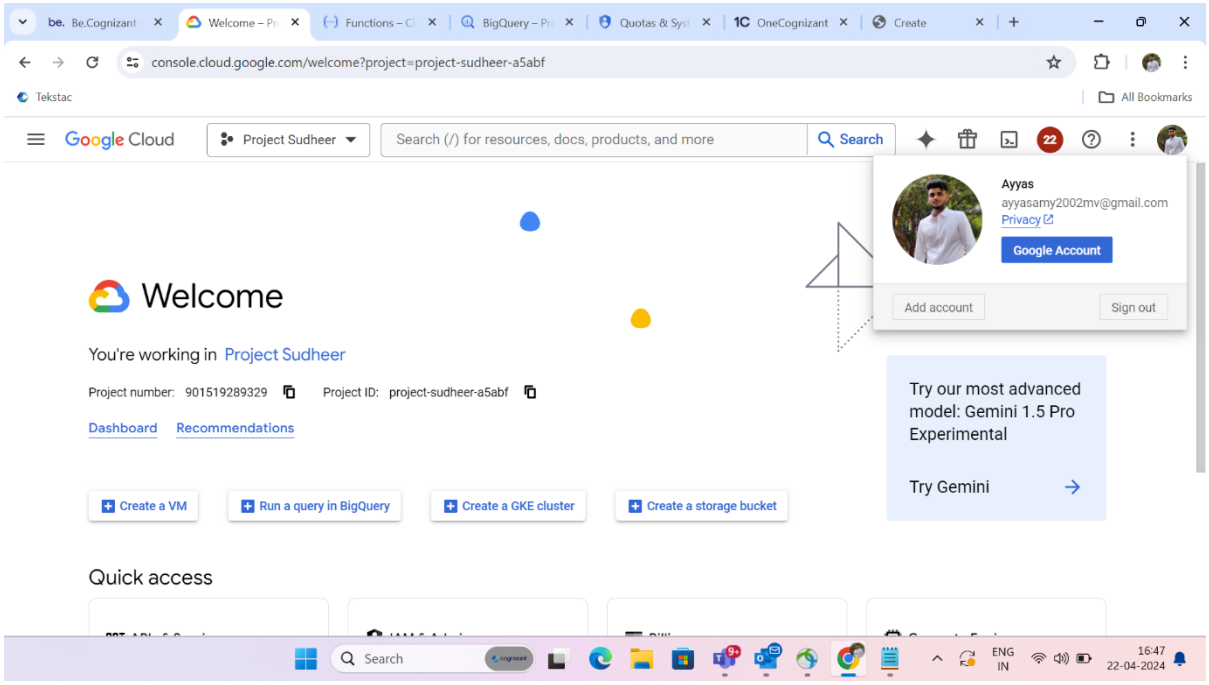
Key Components:

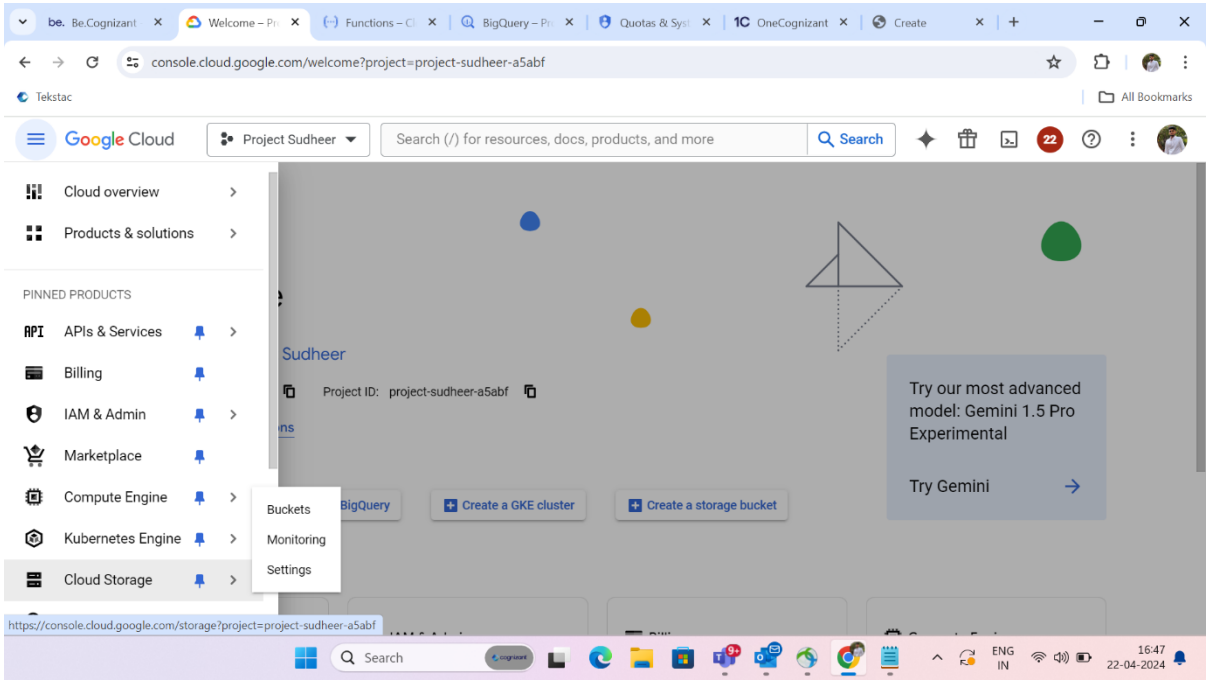
1. **Google Cloud Storage Buckets**: Used for storing raw data files.
2. **Cloud Functions**: Triggered by new file uploads, initiating the data processing pipeline.
3. **Dataflow**: Processes and transforms the data from Cloud Storage and loads it into BigQuery.
4. **BigQuery**: Serves as the data warehouse for the processed data.

Objectives:

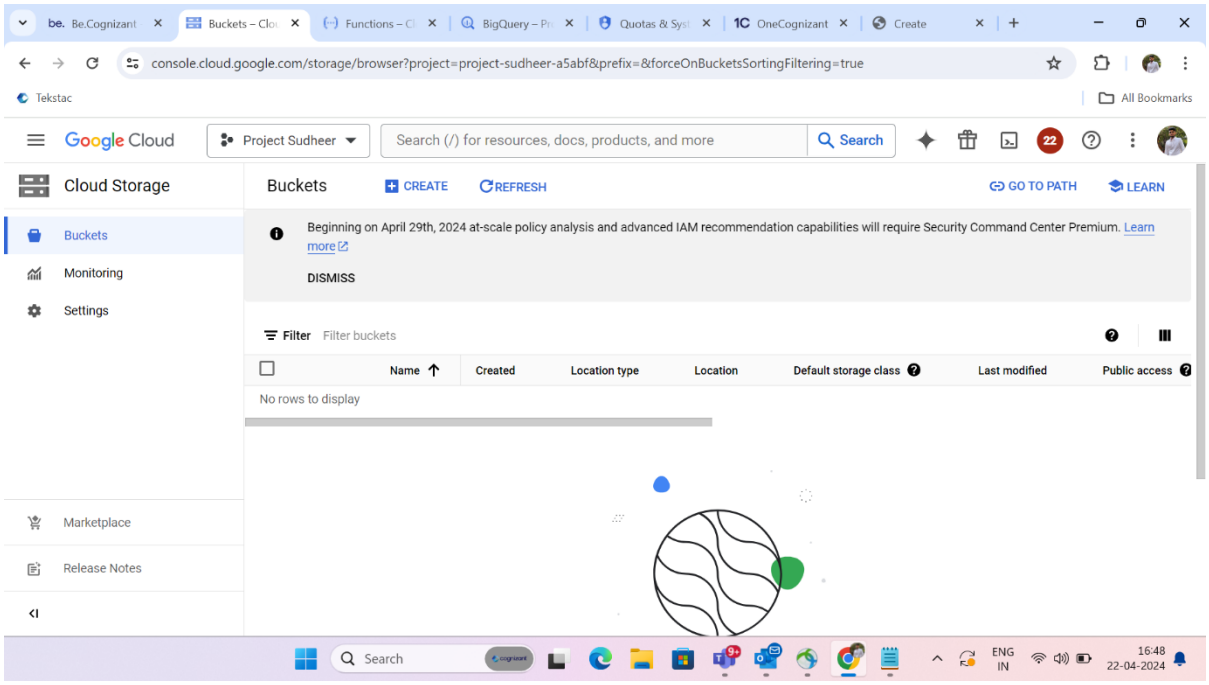
- Automatically trigger data processing tasks in response to new data uploads.
- Perform ETL (Extract, Transform, Load) operations on the data using Dataflow.
- Load transformed data into BigQuery for analysis and reporting.
- Demonstrate the scalability and efficiency of cloud-based data orchestration using Google Cloud Platform services.

GOOGLE CLOUD PLATFORM

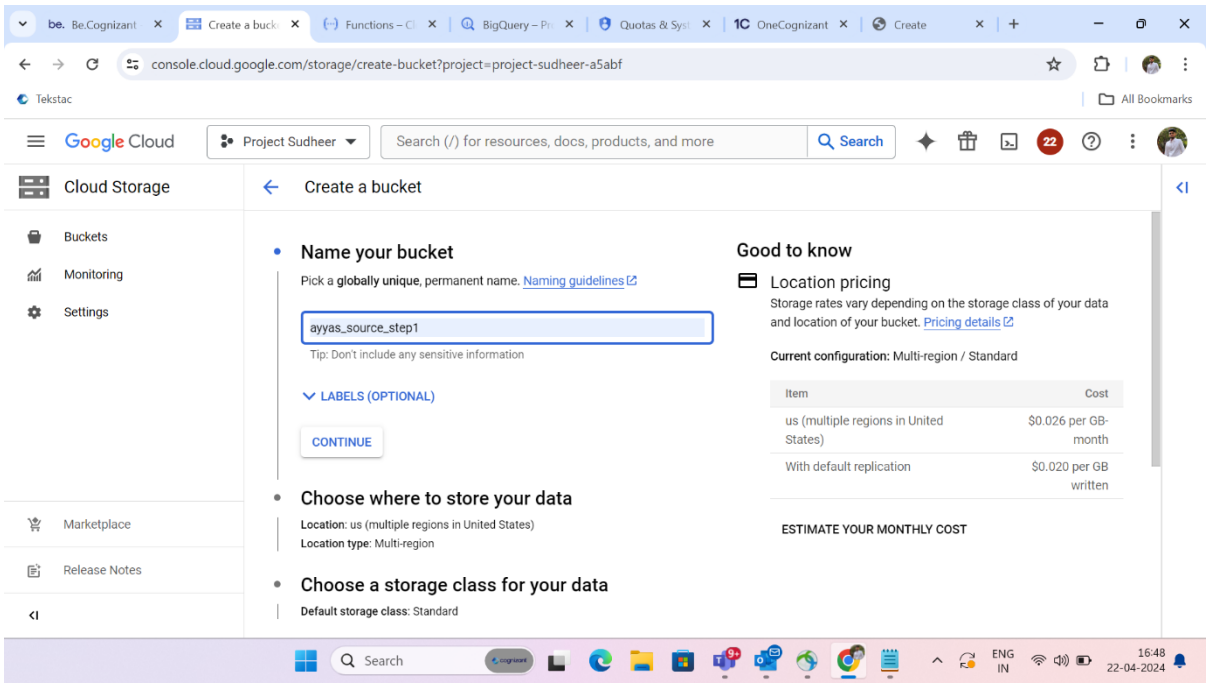


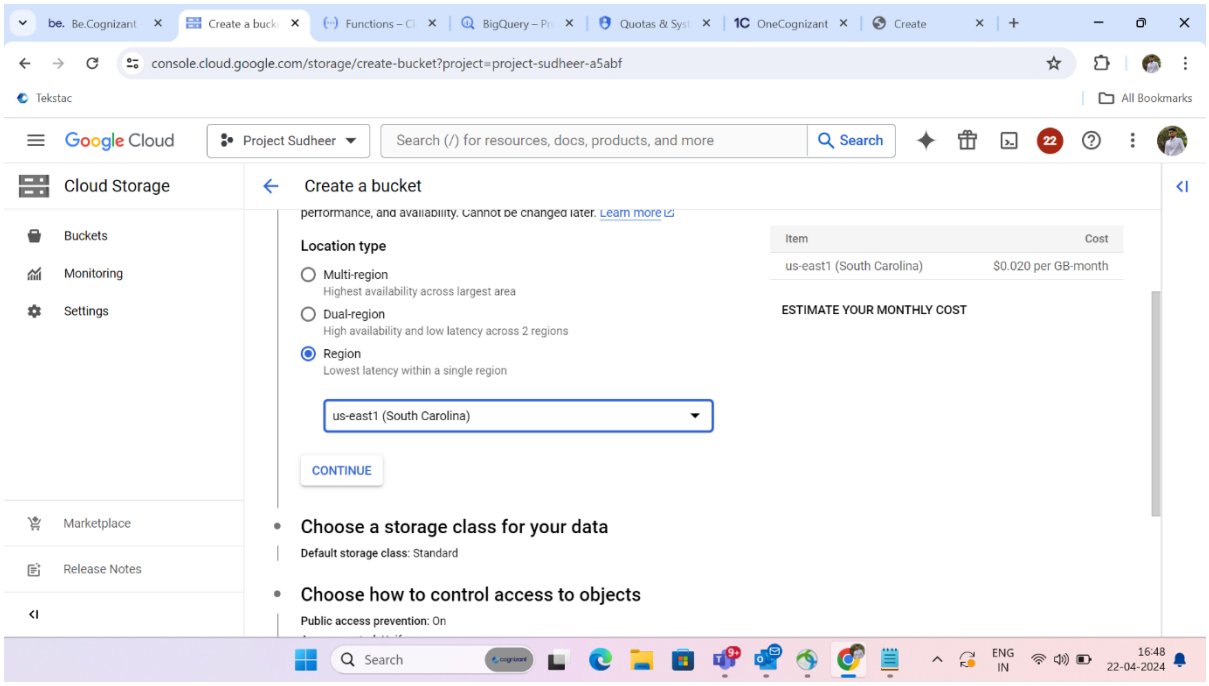


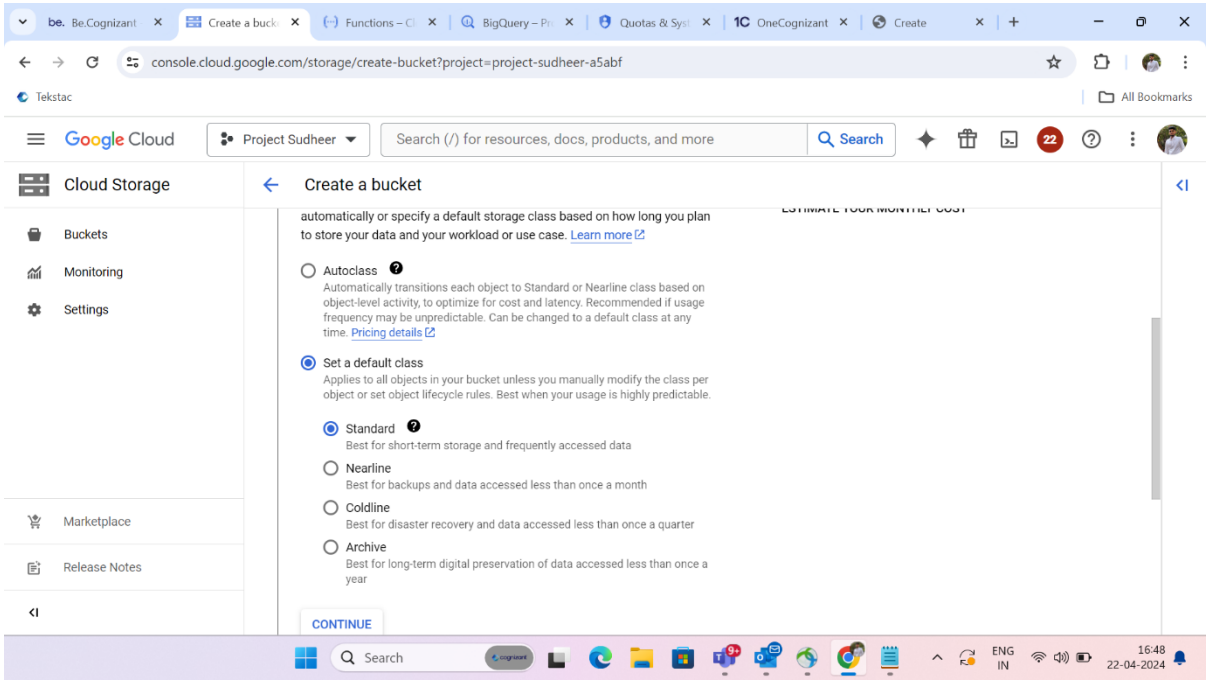
BUCKET CREATION :

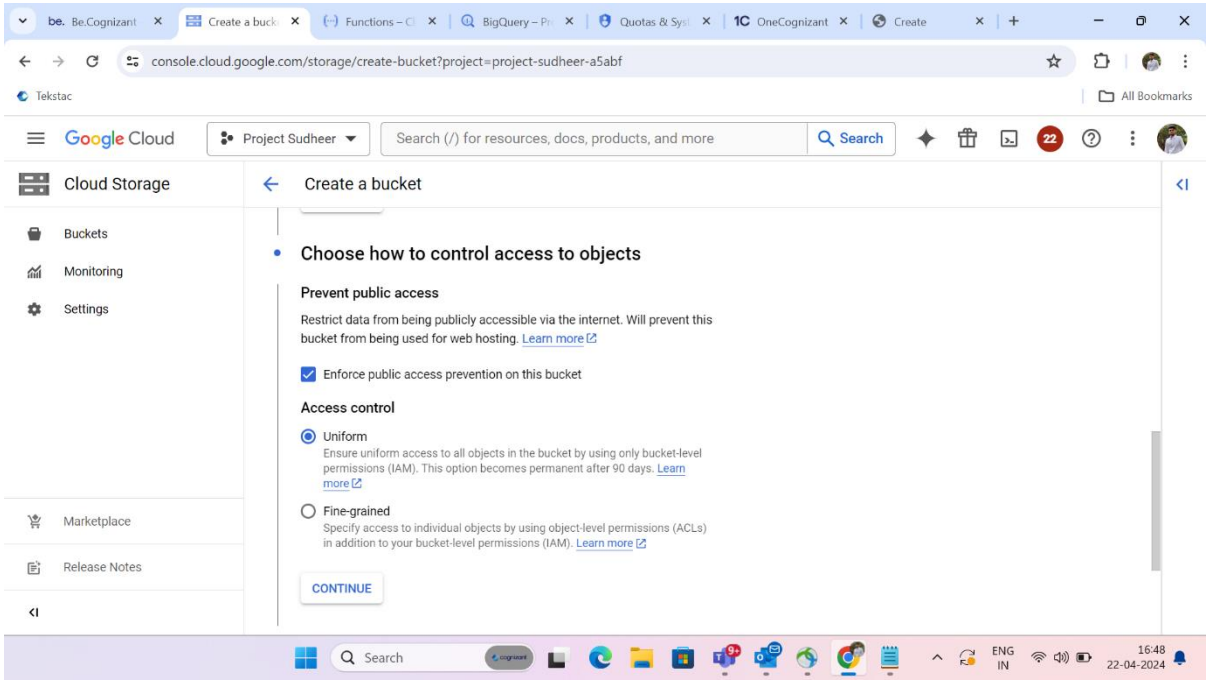


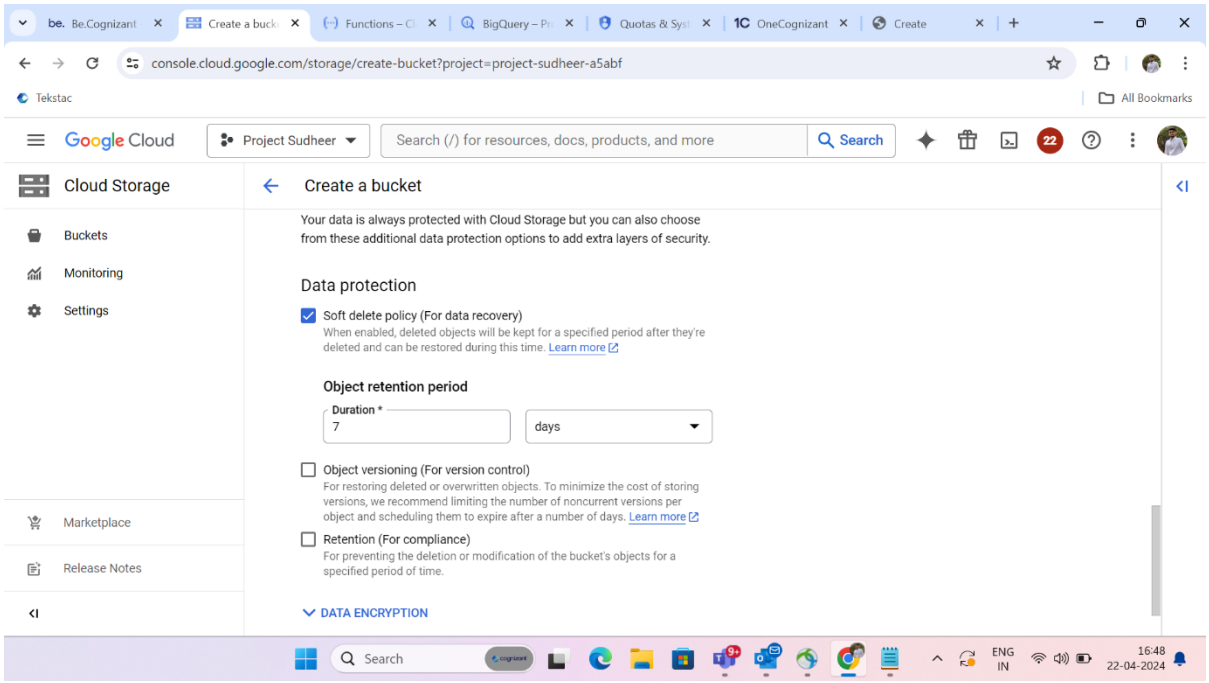
SOURCE BUCKET:

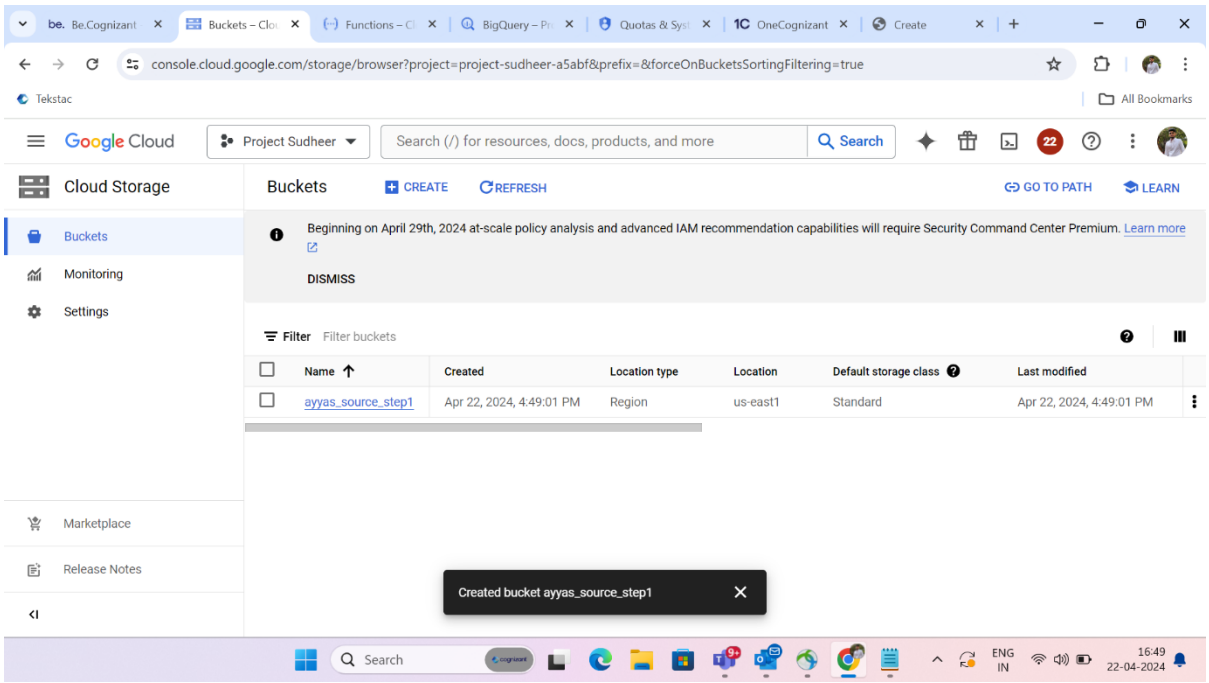



console.cloud.google.com/storage/create-bucket?project=project-sudheer-a5abf
Google Cloud
Project Sudheer
Search (/) for resources, docs, products, and more
Search
Cloud Storage
Create a bucket
Buckets
Monitoring
Settings
Marketplace
Release Notes
performance, and availability. Cannot be changed later. Learn more
Location type
Multi-region
Highest availability across largest area
Dual-region
High availability and low latency across 2 regions
Region
Lowest latency within a single region
us-east1 (South Carolina)
CONTINUE
Item
Cost
us-east1 (South Carolina)
$0.020 per GB-month
ESTIMATE YOUR MONTHLY COST
Choose a storage class for your data
Default storage class: Standard
Choose how to control access to objects
Public access prevention: On
ENG
IN
16:48
22-04-2024

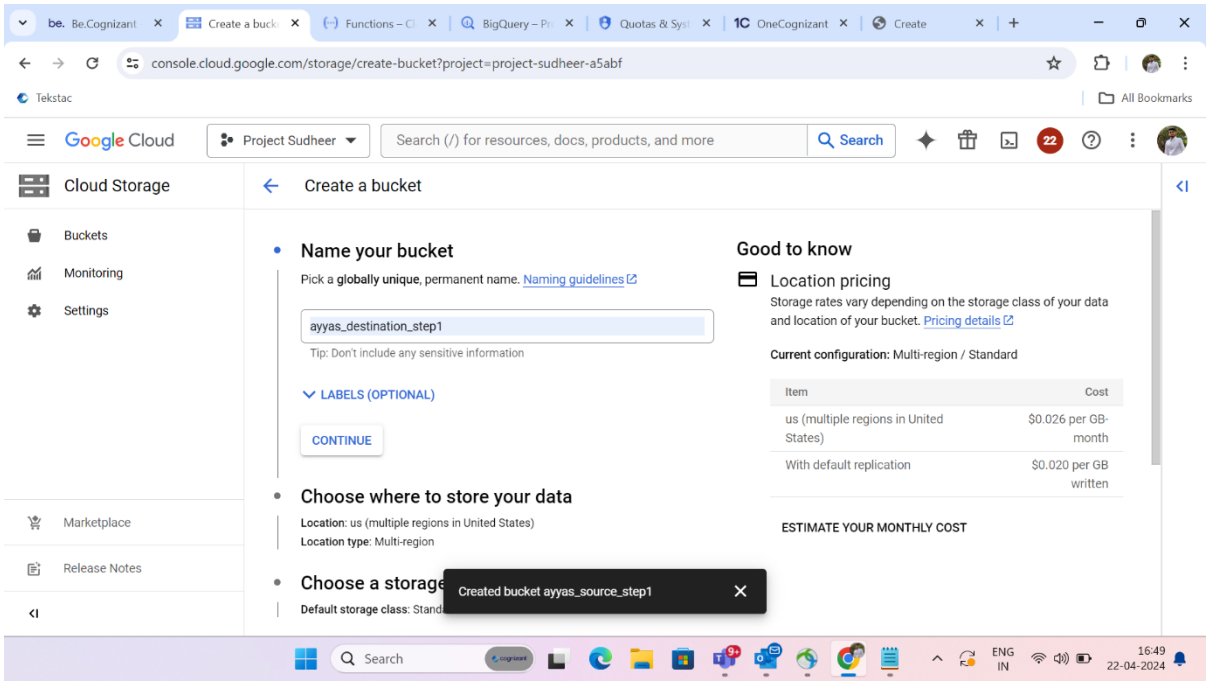

console.cloud.google.com/storage/create-bucket?project=project-sudheer-a5abf
Google Cloud
Project Sudheer
Search (/) for resources, docs, products, and more
Search
Cloud Storage
Create a bucket
Buckets
Monitoring
Settings
Marketplace
Release Notes
automatically or specify a default storage class based 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
ENG
IN
16:48
22-04-2024

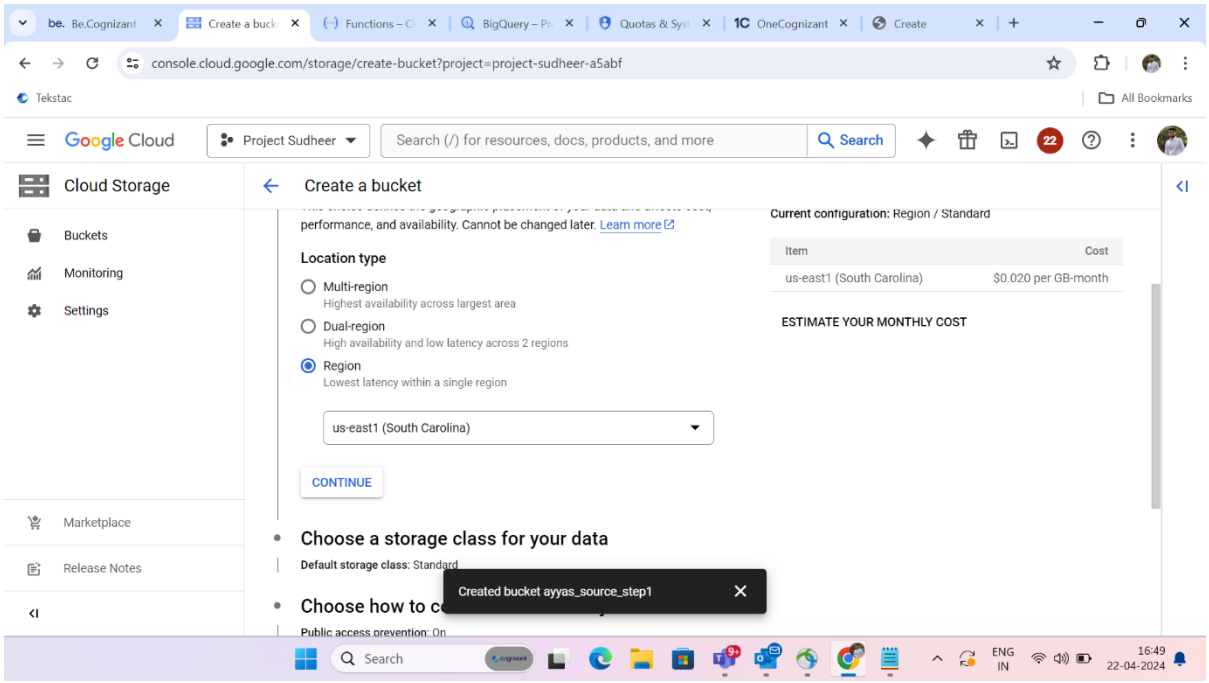

Be.Cognizant
Create a buck
Functions – Cl
BigQuery – Pr
Quotas & Sys
OneCognizant
Create
console.cloud.google.com/storage/create-bucket?project=project-sudheer-a5abf
Tekstac
All Bookmarks
Google Cloud
Project Sudheer
Search (/) for resources, docs, products, and more
Search
Cloud Storage
Create a bucket
Buckets
Monitoring
Settings
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
Marketplace
Release Notes
Search
ENG
IN
16:48
22-04-2024

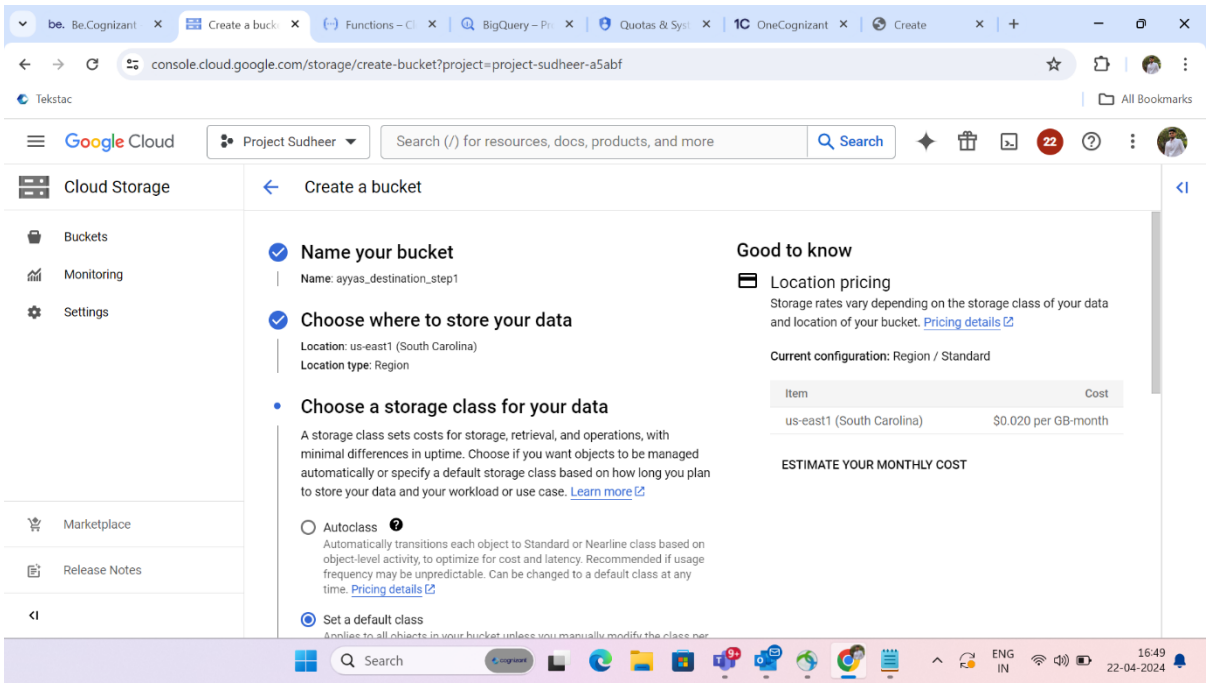


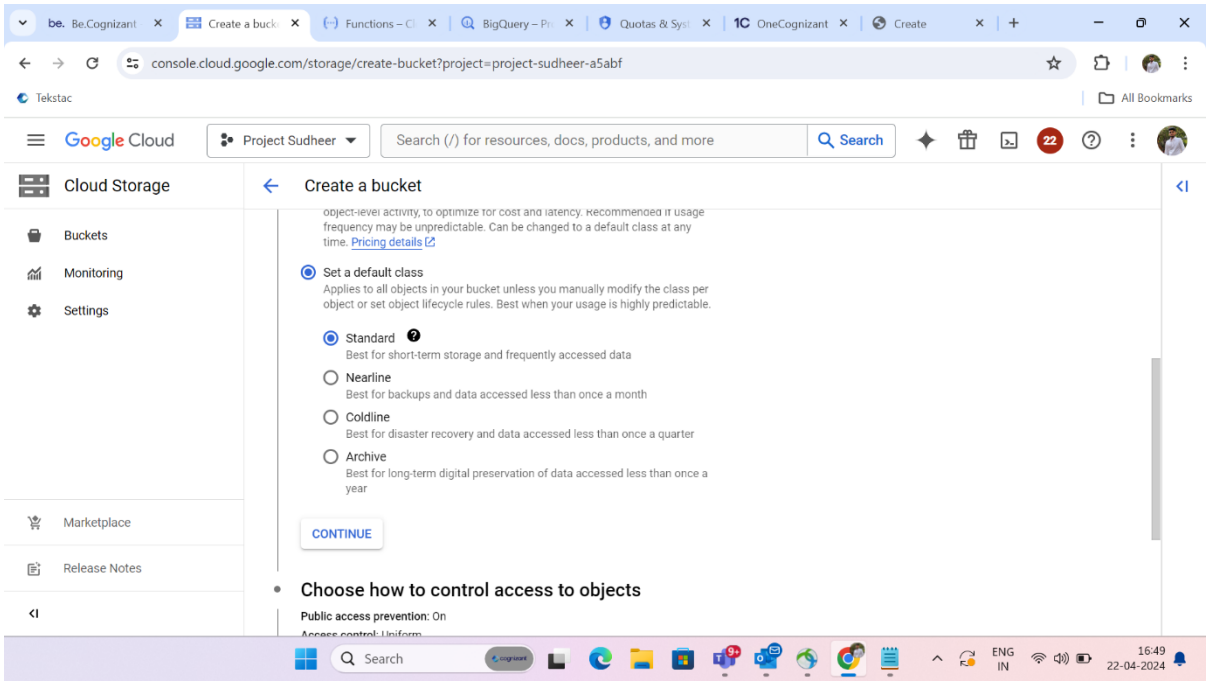


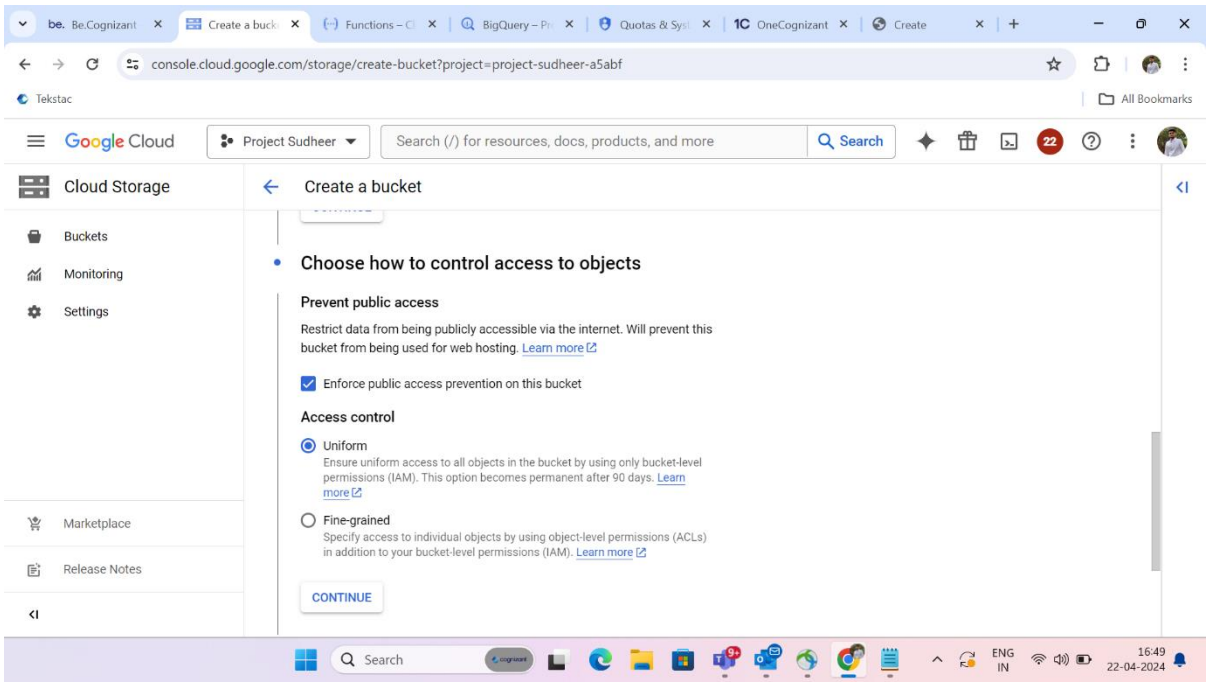
DESTINATION BUCKET:

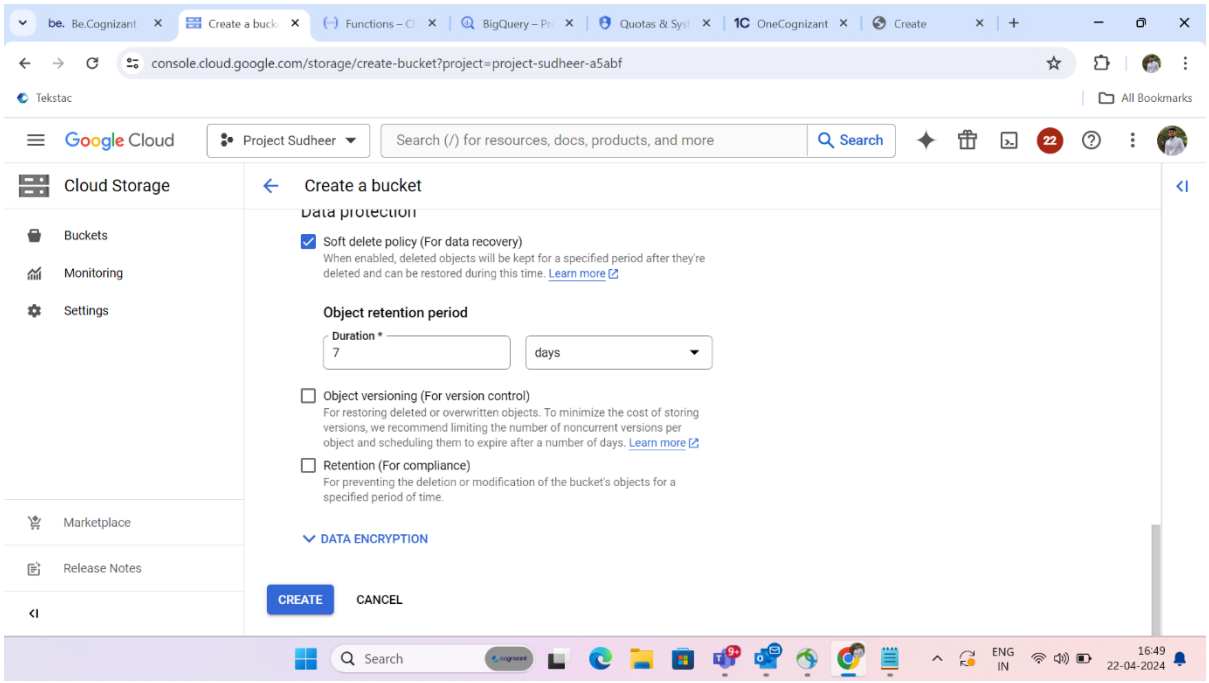

console.cloud.google.com/storage/create-bucket?project=project-sudheer-a5abf
Google Cloud
Project Sudheer
Search (/) for resources, docs, products, and more
Search
Cloud Storage
Buckets
Monitoring
Settings
Marketplace
Release Notes
Create a bucket
Name your bucket
Pick a globally unique, permanent name. Naming guidelines
ayyas_destination_step1
Tip: Don't include any sensitive information
LABELS (OPTIONAL)
CONTINUE
Choose where to store your data
Location: us (multiple regions in United States)
Location type: Multi-region
Choose a storage
Default storage class: Stand
Created bucket ayyas_source_step1
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: Multi-region / Standard
Item
Cost
us (multiple regions in United States)
$0.026 per GB-month
With default replication
$0.020 per GB written
ESTIMATE YOUR MONTHLY COST
ENG
IN
16:49
22-04-2024


console.cloud.google.com/storage/create-bucket?project=project-sudheer-a5abf
Google Cloud
Project Sudheer
Search (/) for resources, docs, products, and more
Search
Cloud Storage
Create a bucket
Buckets
Monitoring
Settings
Marketplace
Release Notes
performance, and availability. Cannot be changed later. Learn more
Location type
Multi-region
Highest availability across largest area
Dual-region
High availability and low latency across 2 regions
Region
Lowest latency within a single region
us-east1 (South Carolina)
CONTINUE
Choose a storage class for your data
Default storage class: Standard
Choose how to c
Public access prevention: On
Created bucket ayyas_source_step1
Current configuration: Region / Standard
Item
Cost
us-east1 (South Carolina)
$0.020 per GB-month
ESTIMATE YOUR MONTHLY COST

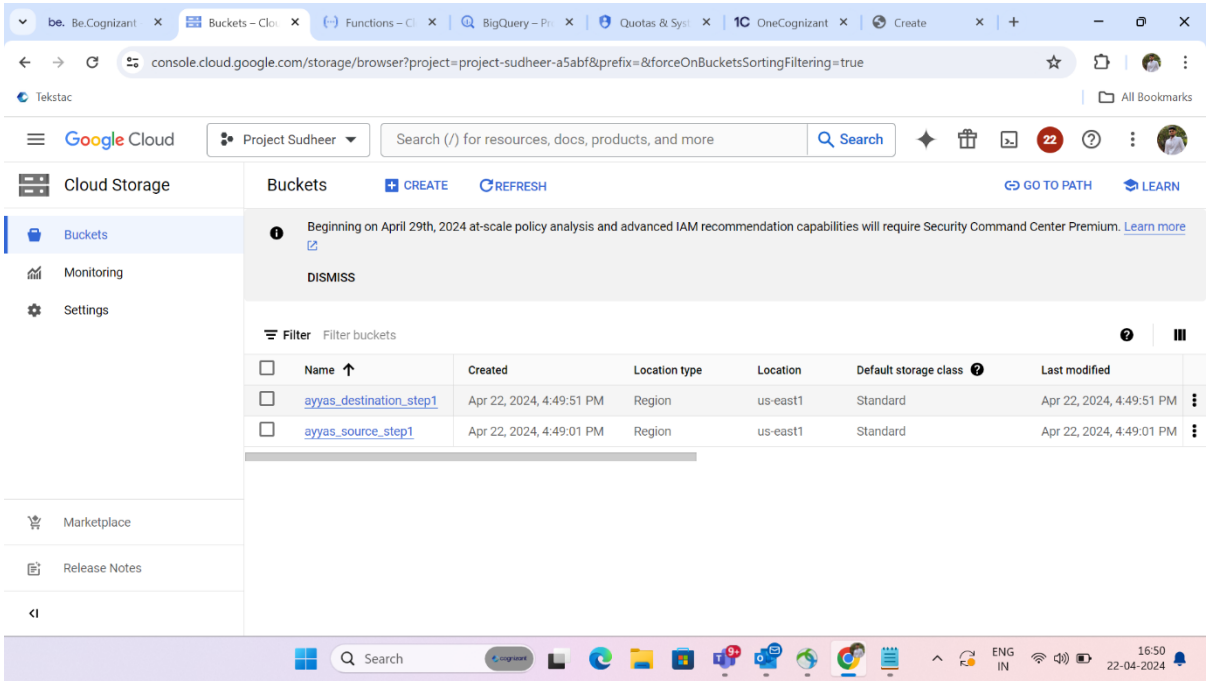

console.cloud.google.com/storage/create-bucket?project=project-sudheer-a5abf
Google Cloud
Project Sudheer
Search (/) for resources, docs, products, and more
Search
Cloud Storage
Create a bucket
Buckets
Monitoring
Settings
Marketplace
Release Notes
Name your bucket
Name: ayyas_destination_step1
Choose where to store your data
Location: us-east1 (South Carolina)
Location type: Region
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
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
us-east1 (South Carolina)
$0.020 per GB-month
ESTIMATE YOUR MONTHLY COST


console.cloud.google.com/storage/create-bucket?project=project-sudheer-a5abf
Google Cloud
Project Sudheer
Search (/) for resources, docs, products, and more
Search
Cloud Storage
Create a bucket
Buckets
Monitoring
Settings
Marketplace
Release Notes
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

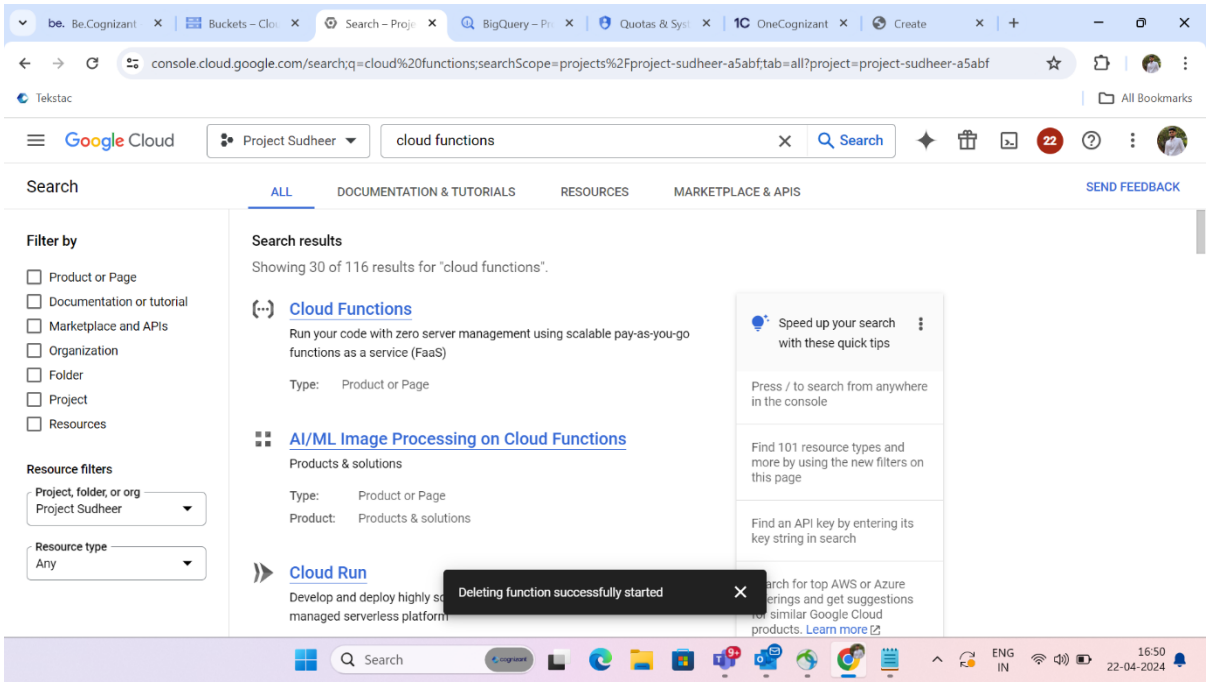

console.cloud.google.com/storage/create-bucket?project=project-sudheer-a5abf
Google Cloud
Project Sudheer
Search (/) for resources, docs, products, and more
Search
Cloud Storage
Create a bucket
Buckets
Monitoring
Settings
Marketplace
Release Notes
Choose how to control access to objects
Prevent public 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

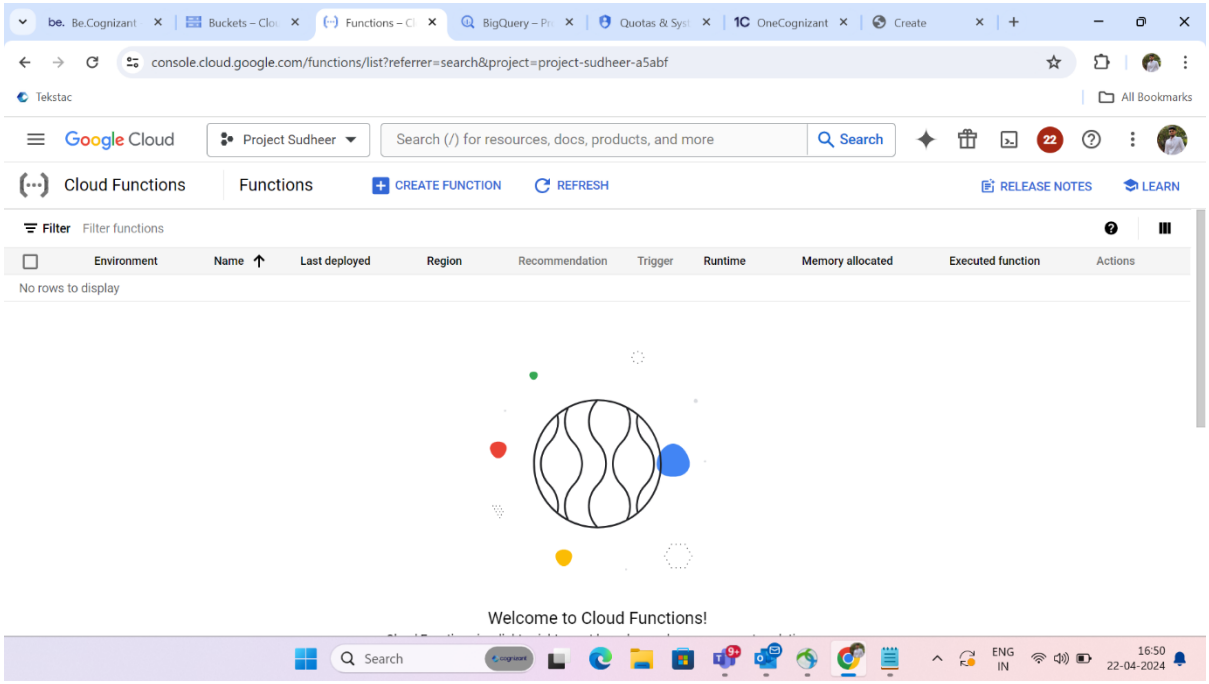


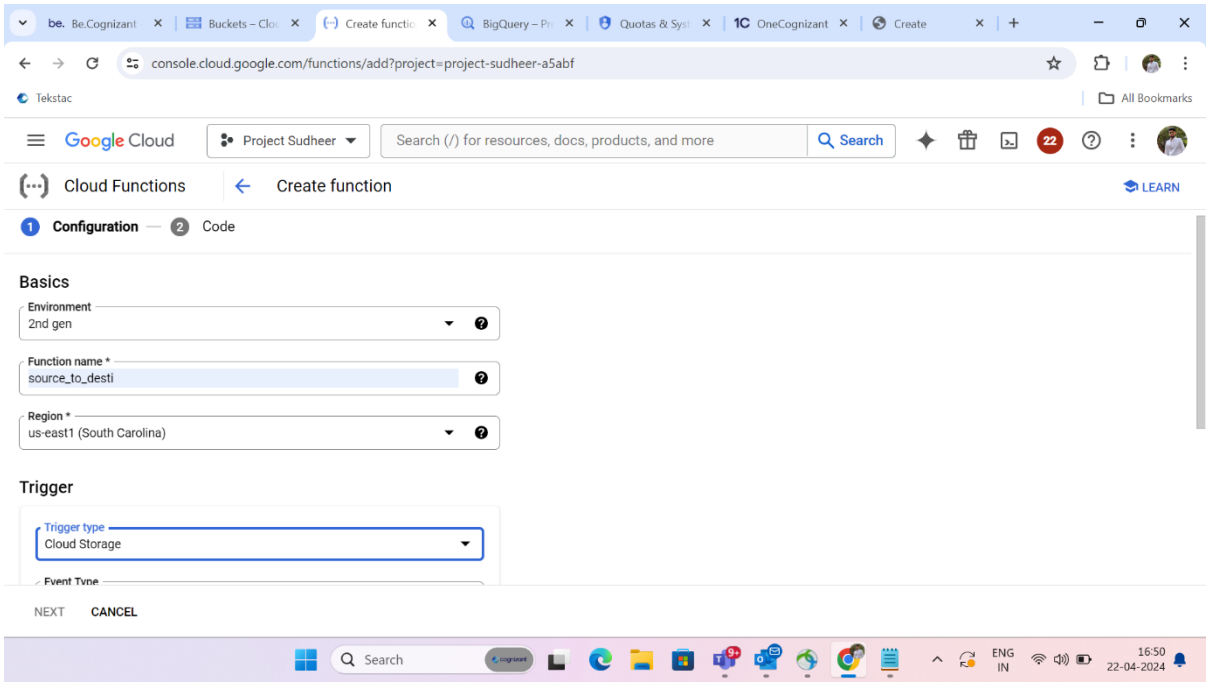
CHECK BUCKET:

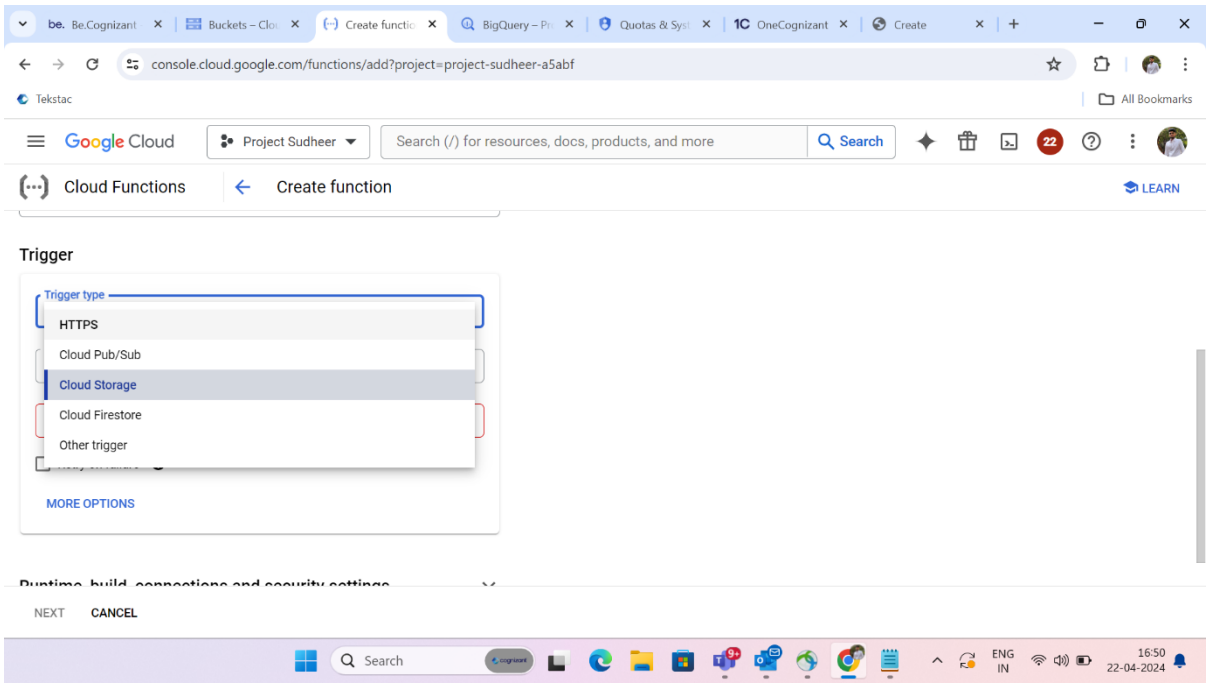


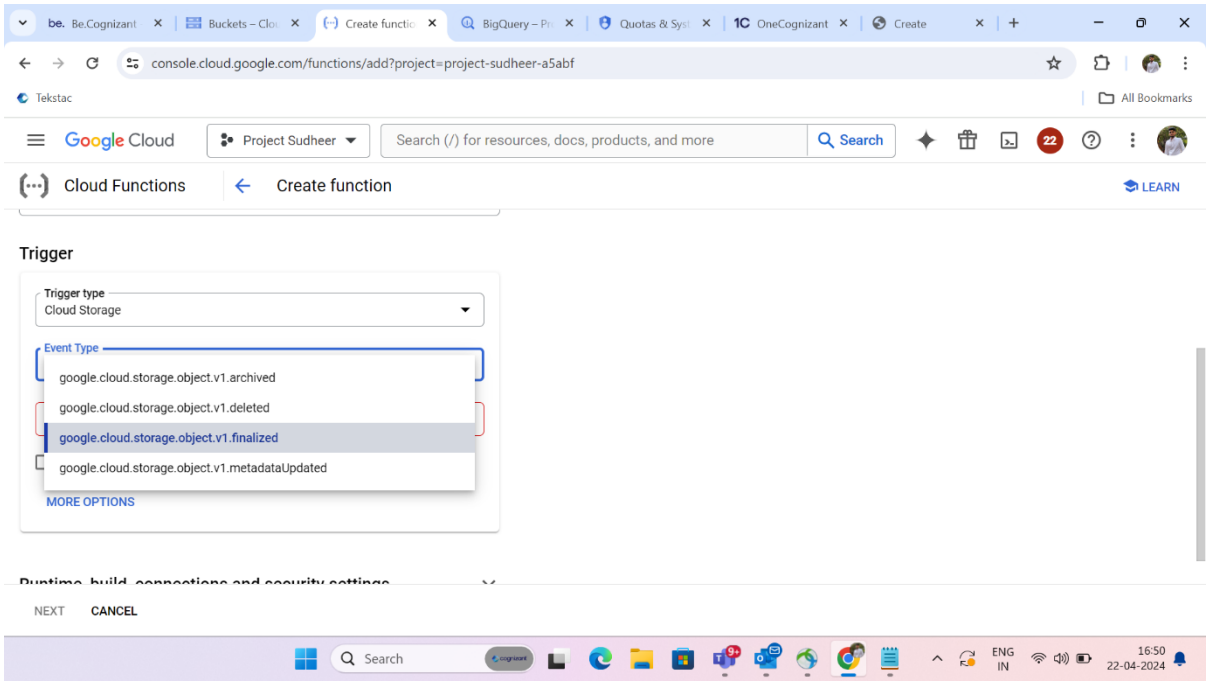
CLOUD FUNCTION:

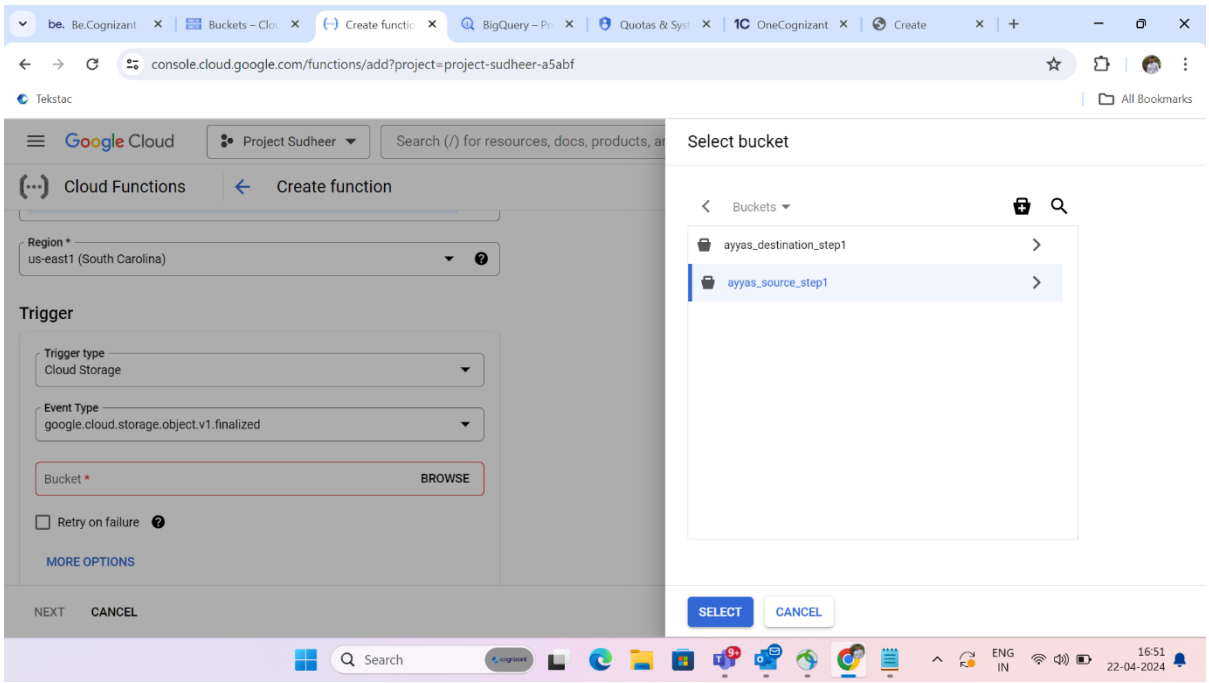


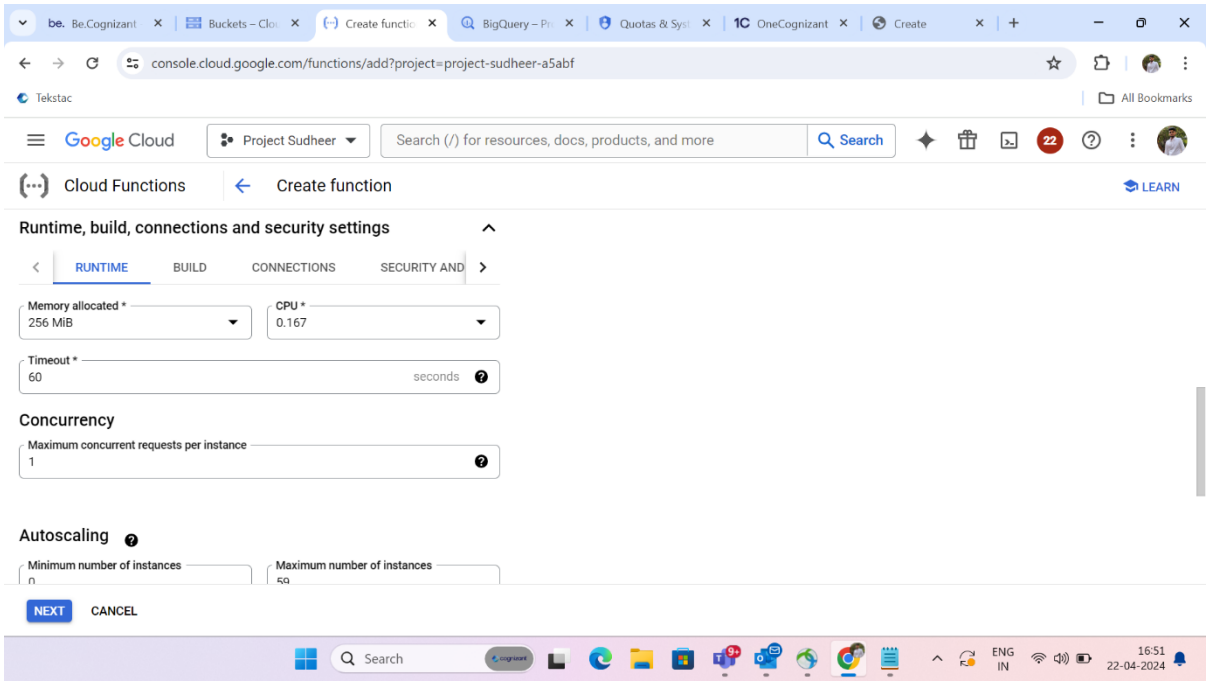


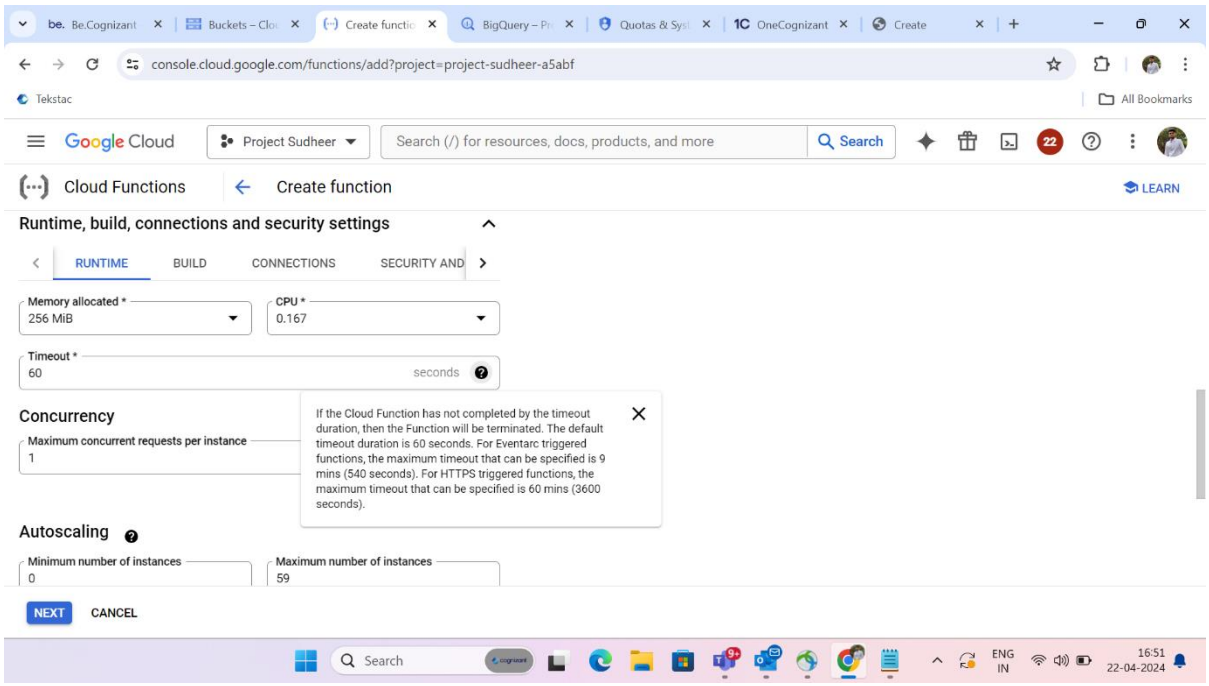

console.cloud.google.com/functions/add?project=project-sudheer-a5abf
Google Cloud
Project Sudheer
Search (/) for resources, docs, products, and more
Search
Cloud Functions
Create function
LEARN
1 Configuration
2 Code
Basics
Environment
2nd gen
Function name *
source_to_desti
Region *
us-east1 (South Carolina)
Trigger
Trigger type
Cloud Storage
NEXT
CANCEL

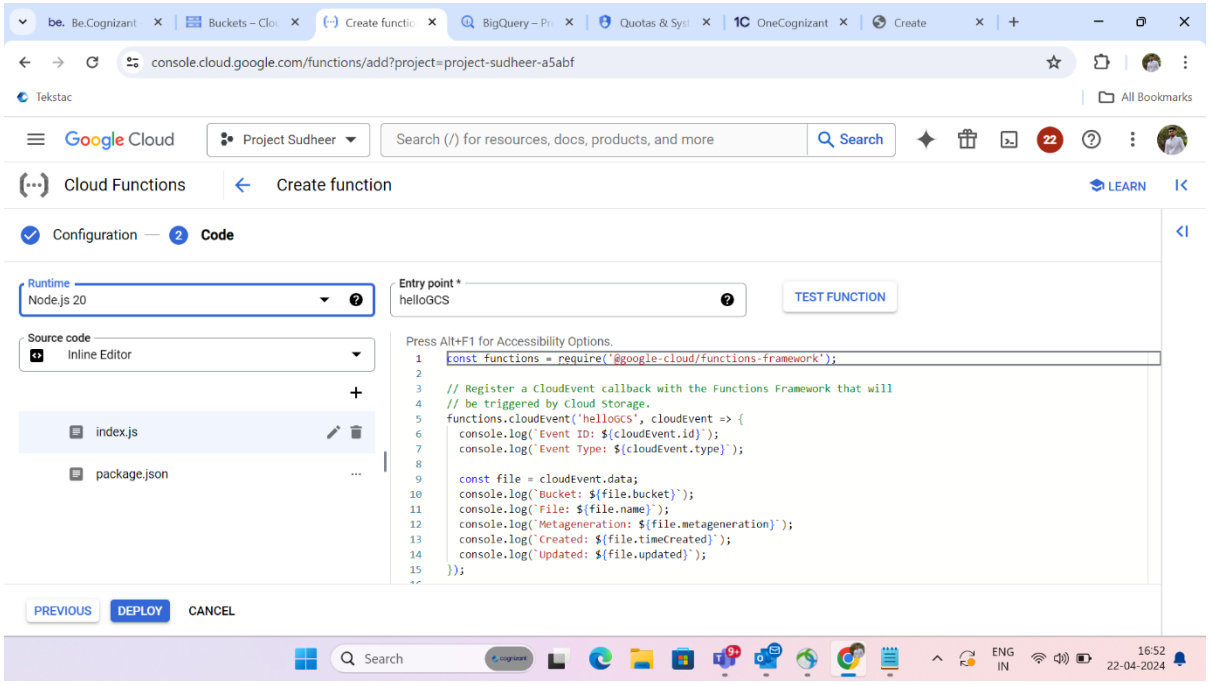

console.cloud.google.com/functions/add?project=project-sudheer-a5abf
Google Cloud
Project Sudheer
Search (/) for resources, docs, products, and more
Search
Cloud Functions
Create function
LEARN
Trigger
Trigger type
HTTPS
Cloud Pub/Sub
Cloud Storage
Cloud Firestore
Other trigger
MORE OPTIONS
NEXT
CANCEL

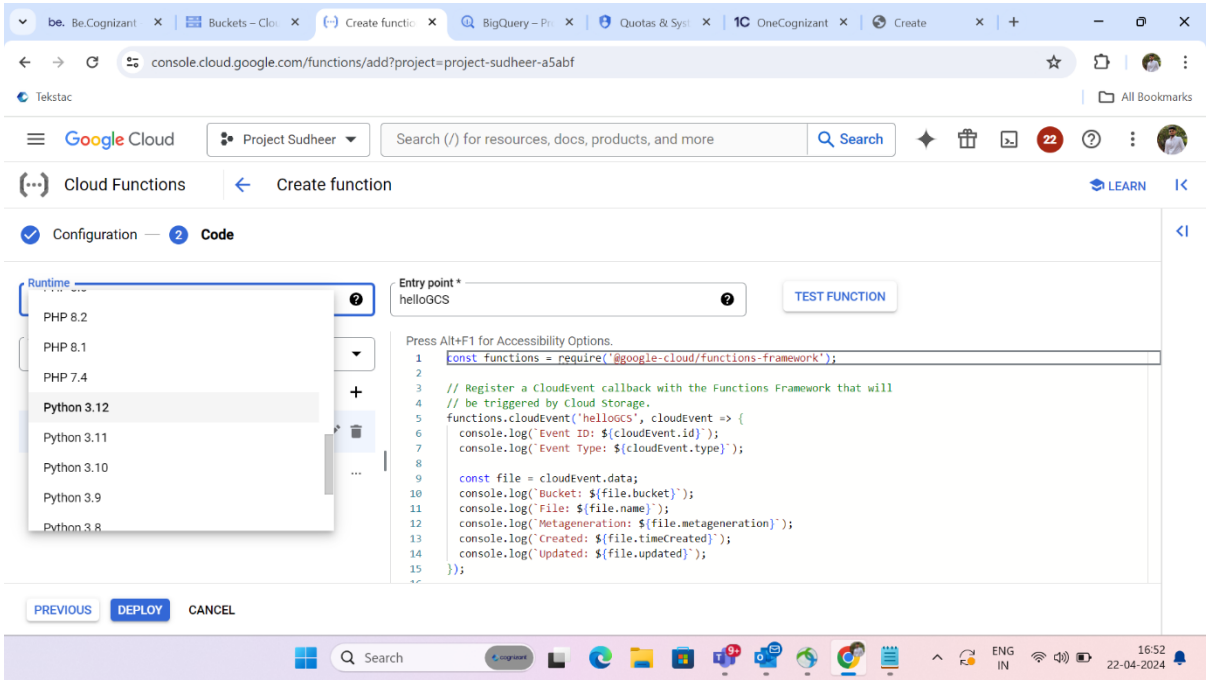

console.cloud.google.com/functions/add?project=project-sudheer-a5abf
Google Cloud
Project Sudheer
Search (/) for resources, docs, products, and more
Search
Cloud Functions
Create function
LEARN
Trigger
Trigger type
Cloud Storage
Event Type
google.cloud.storage.object.v1.archived
google.cloud.storage.object.v1.deleted
google.cloud.storage.object.v1.finalized
google.cloud.storage.object.v1.metadataUpdated
MORE OPTIONS
NEXT
CANCEL

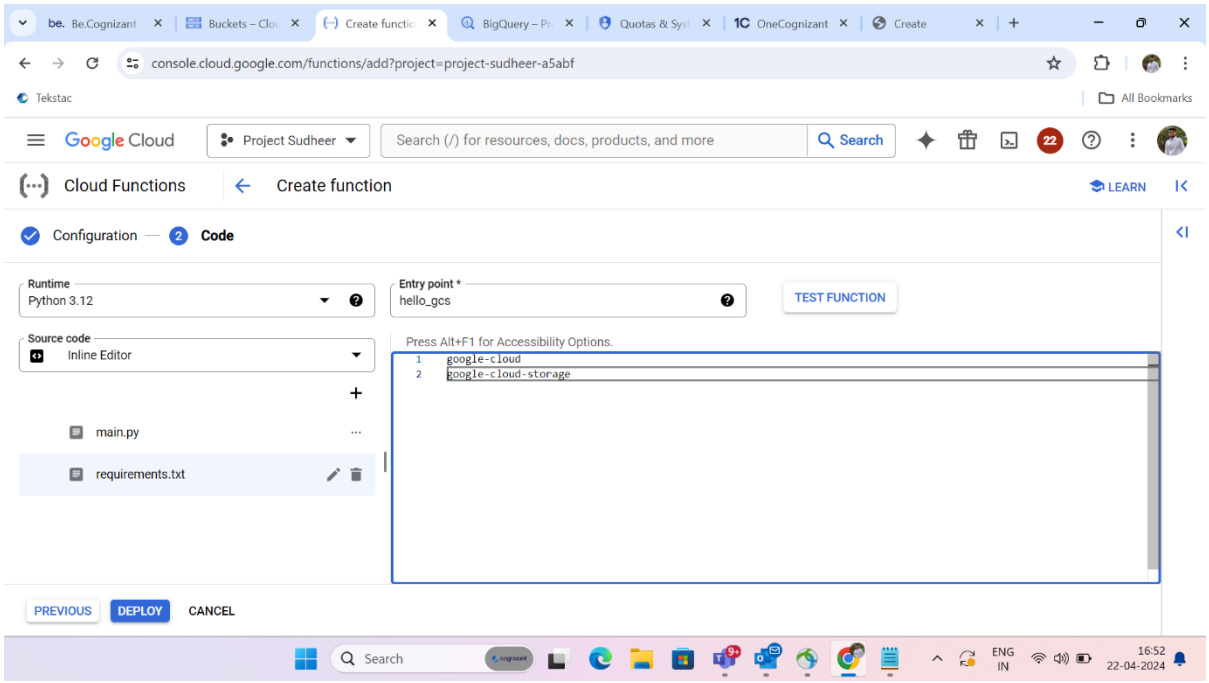

console.cloud.google.com/functions/add?project=project-sudheer-a5abf
Google Cloud
Project Sudheer
Cloud Functions
Create function
Region *
us-east1 (South Carolina)
Trigger
Trigger type
Cloud Storage
Event Type
google.cloud.storage.object.v1.finalized
Bucket *
BROWSE
Retry on failure
MORE OPTIONS
NEXT
CANCEL
Select bucket
Buckets
ayyas_destination_step1
ayyas_source_step1
SELECT
CANCEL

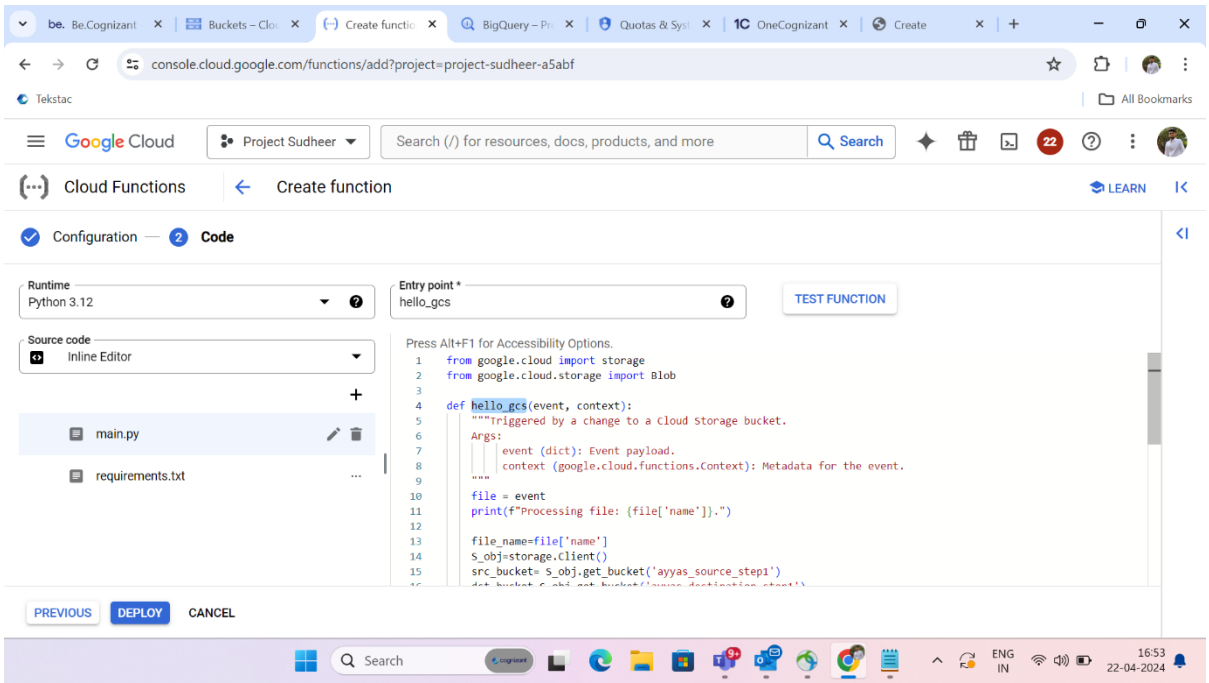

console.cloud.google.com/functions/add?project=project-sudheer-a5abf
Google Cloud
Project Sudheer
Search (/) for resources, docs, products, and more
Search
Cloud Functions
Create function
LEARN
Runtime, build, connections and security settings
RUNTIME
BUILD
CONNECTIONS
SECURITY AND
Memory allocated *
256 MiB
CPU *
0.167
Timeout *
60
seconds
Concurrency
Maximum concurrent requests per instance
1
Autoscaling
Minimum number of instances
Maximum number of instances
59
NEXT
CANCEL

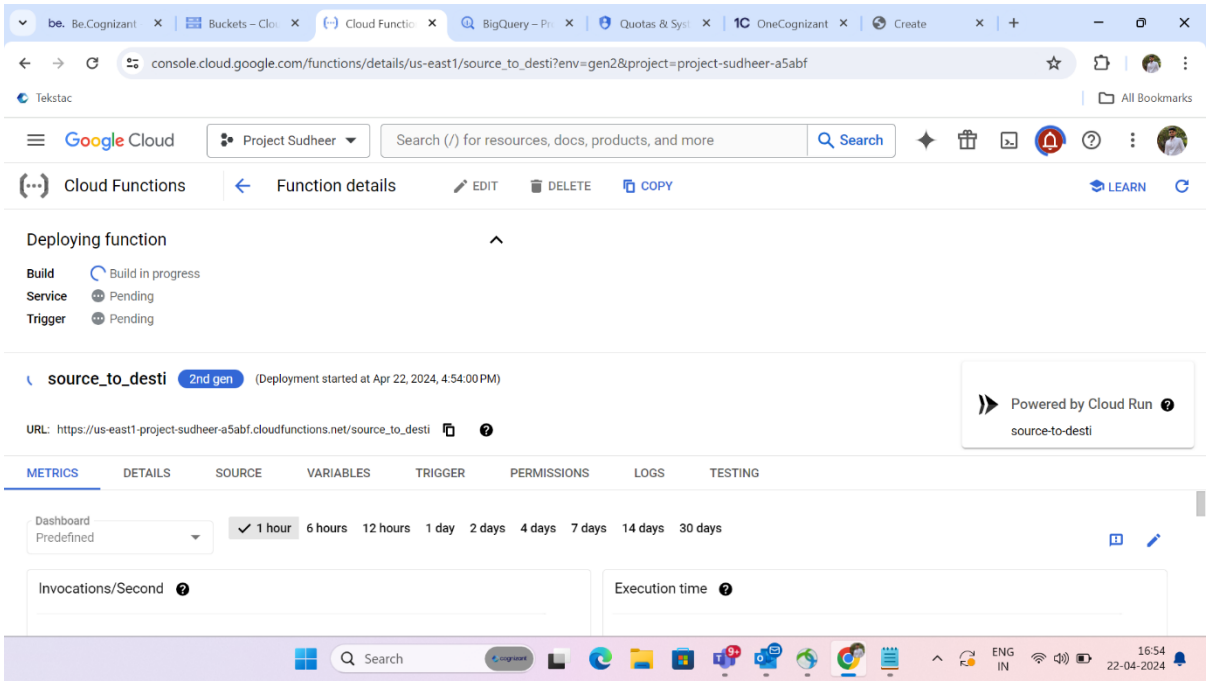

console.cloud.google.com/functions/add?project=project-sudheer-a5abf
Google Cloud
Project Sudheer
Search (/) for resources, docs, products, and more
Search
Cloud Functions
Create function
LEARN
Runtime, build, connections and security settings
RUNTIME
BUILD
CONNECTIONS
SECURITY AND
Memory allocated *
256 MiB
CPU *
0.167
Timeout *
60
seconds
Concurrency
Maximum concurrent requests per instance
1
If the Cloud Function has not completed by the timeout duration, then the Function will be terminated. The default timeout duration is 60 seconds. For Eventarc triggered functions, the maximum timeout that can be specified is 9 mins (540 seconds). For HTTPS triggered functions, the maximum timeout that can be specified is 60 mins (3600 seconds).
Autoscaling
Minimum number of instances
0
Maximum number of instances
59
NEXT
CANCEL

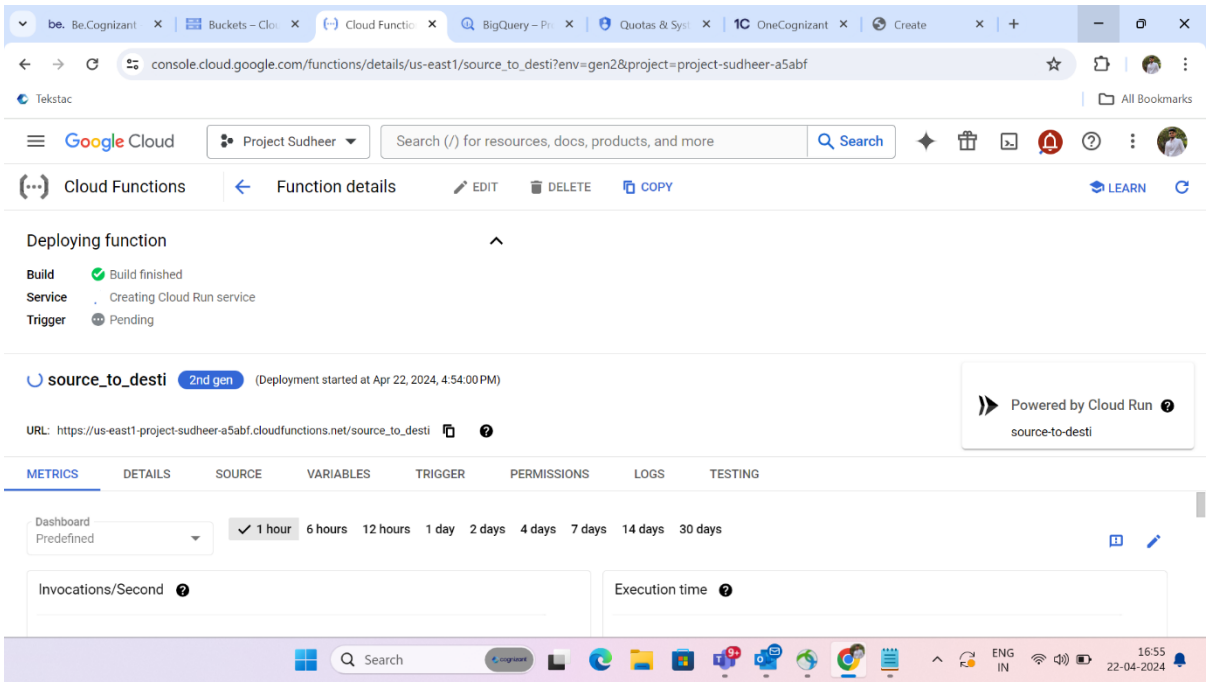

console.cloud.google.com/functions/add?project=project-sudheer-a5abf
Google Cloud
Project Sudheer
Search (/) for resources, docs, products, and more
Search
Cloud Functions
Create function
LEARN
Configuration
2 Code
Runtime
Node.js 20
Entry point *
helloGCS
TEST FUNCTION
Source code
Inline Editor
index.js
package.json
Press Alt+F1 for Accessibility Options.
const functions = require('@google-cloud/functions-framework');
// Register a CloudEvent callback with the Functions Framework that will
// be triggered by Cloud Storage.
functions.cloudEvent('helloGCS', cloudEvent => {
console.log(`Event ID: ${cloudEvent.id}`);
console.log(`Event Type: ${cloudEvent.type}`);
const file = cloudEvent.data;
console.log(`Bucket: ${file.bucket}`);
console.log(`File: ${file.name}`);
console.log(`Metageneration: ${file.metageneration}`);
console.log(`Created: ${file.timeCreated}`);
console.log(`Updated: ${file.updated}`);
});
PREVIOUS
DEPLOY
CANCEL
Search
ENG
IN
16:52
22-04-2024

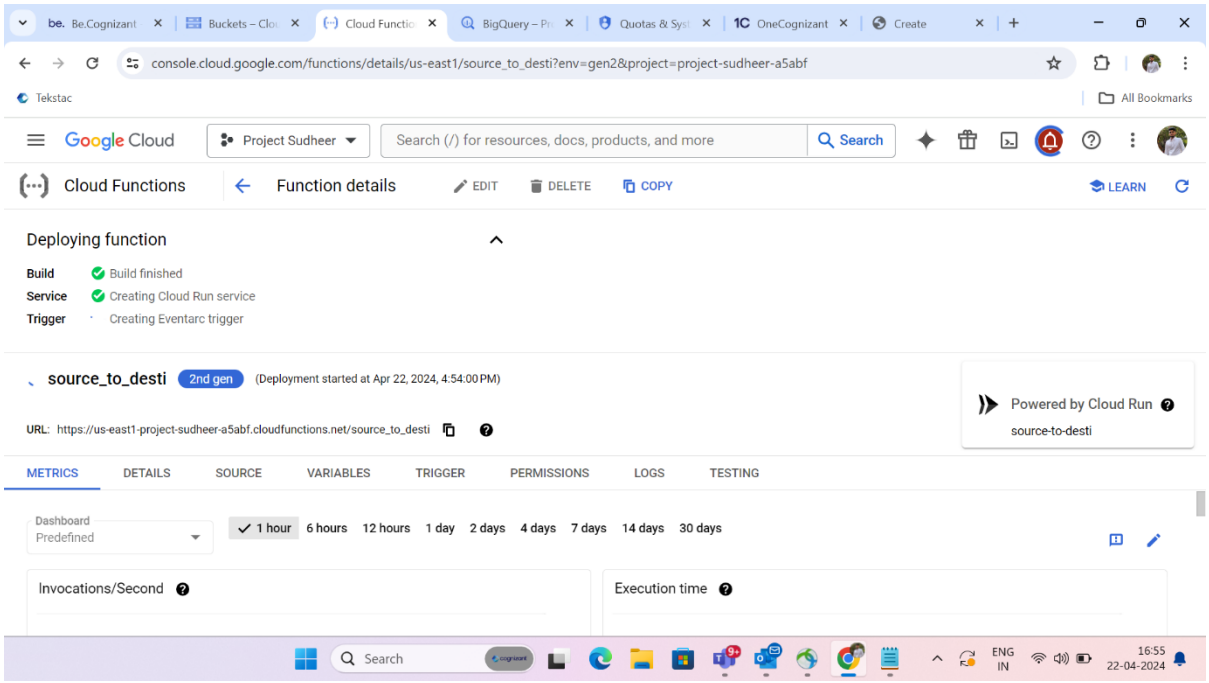

console.cloud.google.com/functions/add?project=project-sudheer-a5abf
Google Cloud
Project Sudheer
Search (/) for resources, docs, products, and more
Search
Cloud Functions
Create function
LEARN
Configuration
2 Code
Runtime
PHP 8.2
PHP 8.1
PHP 7.4
Python 3.12
Python 3.11
Python 3.10
Python 3.9
Entry point *
helloGCS
TEST FUNCTION
Press Alt+F1 for Accessibility Options.
const functions = require('@google-cloud/functions-framework');
// Register a CloudEvent callback with the Functions Framework that will
// be triggered by Cloud Storage.
functions.cloudEvent('helloGCS', cloudEvent => {
console.log(`Event ID: ${cloudEvent.id}`);
console.log(`Event Type: ${cloudEvent.type}`);
const file = cloudEvent.data;
console.log(`Bucket: ${file.bucket}`);
console.log(`File: ${file.name}`);
console.log(`Metageneration: ${file.metageneration}`);
console.log(`Created: ${file.timeCreated}`);
console.log(`Updated: ${file.updated}`);
});
PREVIOUS
DEPLOY
CANCEL
Search
ENG
IN
16:52
22-04-2024

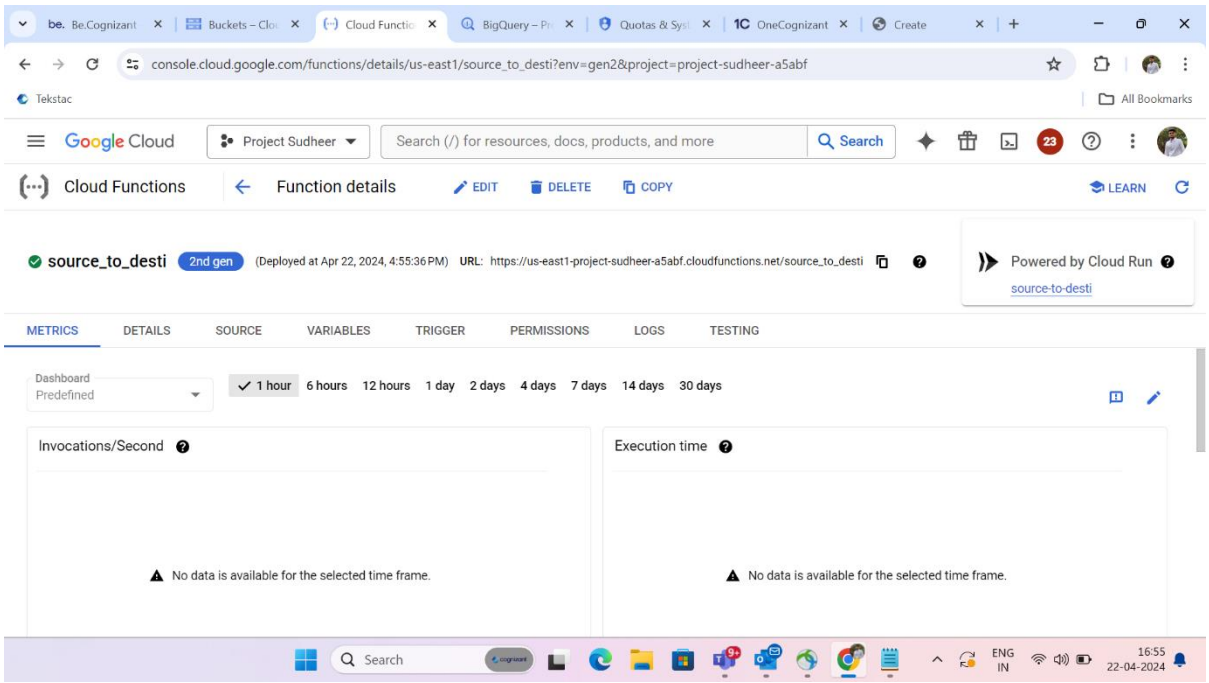

console.cloud.google.com/functions/add?project=project-sudheer-a5abf
Google Cloud
Project Sudheer
Search (/) for resources, docs, products, and more
Search
Cloud Functions
Create function
LEARN
Configuration
2
Code
Runtime
Python 3.12
Entry point *
hello_gcs
TEST FUNCTION
Source code
Inline Editor
Press Alt+F1 for Accessibility Options.
main.py
requirements.txt
google-cloud
google-cloud-storage
PREVIOUS
DEPLOY
CANCEL
Search
ENG
IN
16:52
22-04-2024


console.cloud.google.com/functions/add?project=project-sudheer-a5abf
Google Cloud
Project Sudheer
Search (/) for resources, docs, products, and more
Search
Cloud Functions
Create function
LEARN
Configuration
2
Code
Runtime
Python 3.12
Entry point *
hello_gcs
TEST FUNCTION
Source code
Inline Editor
Press Alt+F1 for Accessibility Options.
main.py
requirements.txt
from google.cloud import storage
from google.cloud.storage import Blob
def hello_gcs(event, context):
"""Triggered by a change to a Cloud Storage bucket.
Args:
event (dict): Event payload.
context (google.cloud.functions.Context): Metadata for the event.
"""
file = event
print(f"Processing file: {file['name']}.")
file_name=file['name']
S_obj=storage.Client()
src_bucket= S_obj.get_bucket('ayyas_source_step1')
PREVIOUS
DEPLOY
CANCEL
Search
ENG
IN
16:53
22-04-2024

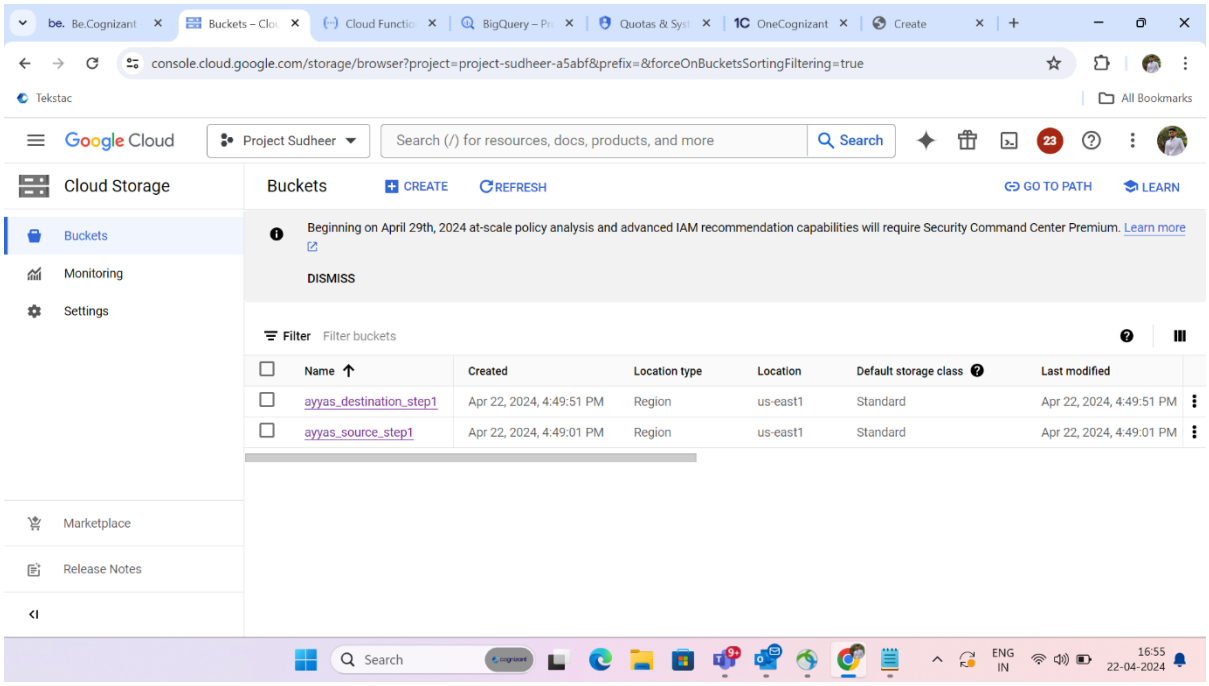

console.cloud.google.com/functions/details/us-east1/source_to_desti?env=gen2&project=project-sudheer-a5abf
Google Cloud
Project Sudheer
Search (/) for resources, docs, products, and more
Search
Cloud Functions
Function details
EDIT
DELETE
COPY
LEARN
Deploying function
Build
Build in progress
Service
Pending
Trigger
Pending
source_to_desti
2nd gen
(Deployment started at Apr 22, 2024, 4:54:00 PM)
URL: https://us-east1-project-sudheer-a5abf.cloudfunctions.net/source_to_desti
Powered by Cloud Run
source-to-desti
METRICS
DETAILS
SOURCE
VARIABLES
TRIGGER
PERMISSIONS
LOGS
TESTING
Dashboard
Predefined
1 hour
6 hours
12 hours
1 day
2 days
4 days
7 days
14 days
30 days
Invocations/Second
Execution time
Search
ENG
IN
16:54
22-04-2024

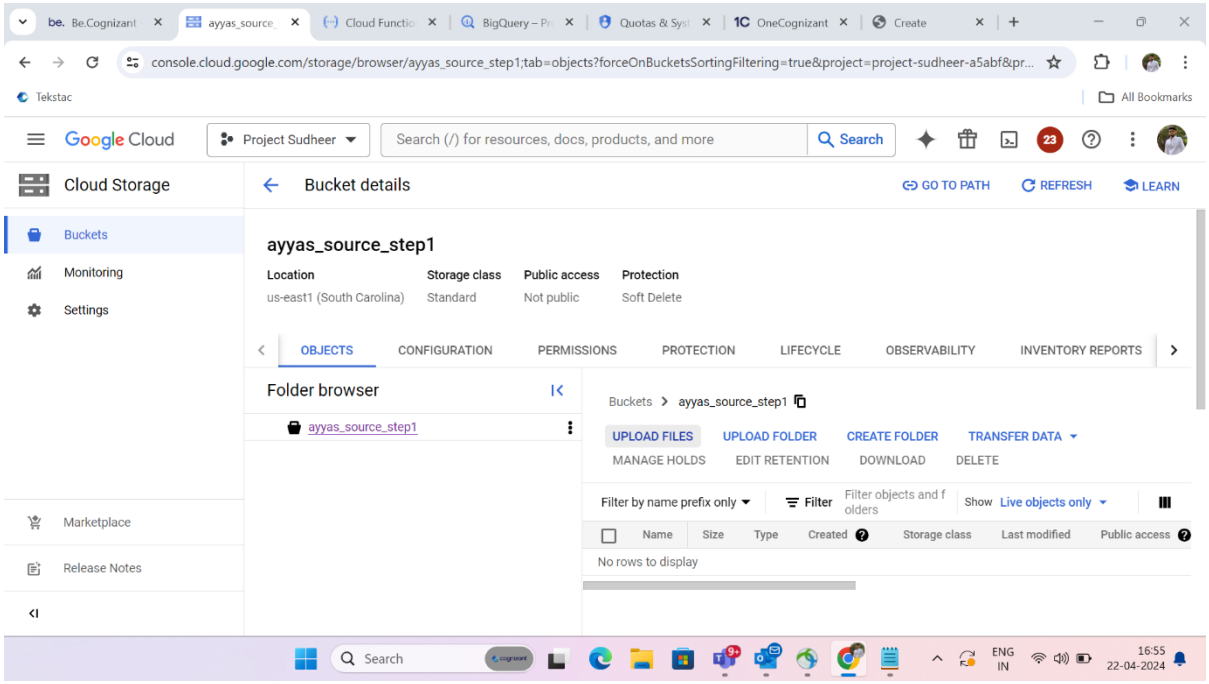

console.cloud.google.com/functions/details/us-east1/source_to_desti?env=gen2&project=project-sudheer-a5abf
Google Cloud
Project Sudheer
Search (/) for resources, docs, products, and more
Search
Cloud Functions
Function details
EDIT
DELETE
COPY
LEARN
Deploying function
Build
Build finished
Service
Creating Cloud Run service
Trigger
Pending
source_to_desti
2nd gen
(Deployment started at Apr 22, 2024, 4:54:00 PM)
URL: https://us-east1-project-sudheer-a5abf.cloudfunctions.net/source_to_desti
Powered by Cloud Run
source-to-desti
METRICS
DETAILS
SOURCE
VARIABLES
TRIGGER
PERMISSIONS
LOGS
TESTING
Dashboard
Predefined
1 hour
6 hours
12 hours
1 day
2 days
4 days
7 days
14 days
30 days
Invocations/Second
Execution time
Search
ENG
IN
16:55
22-04-2024

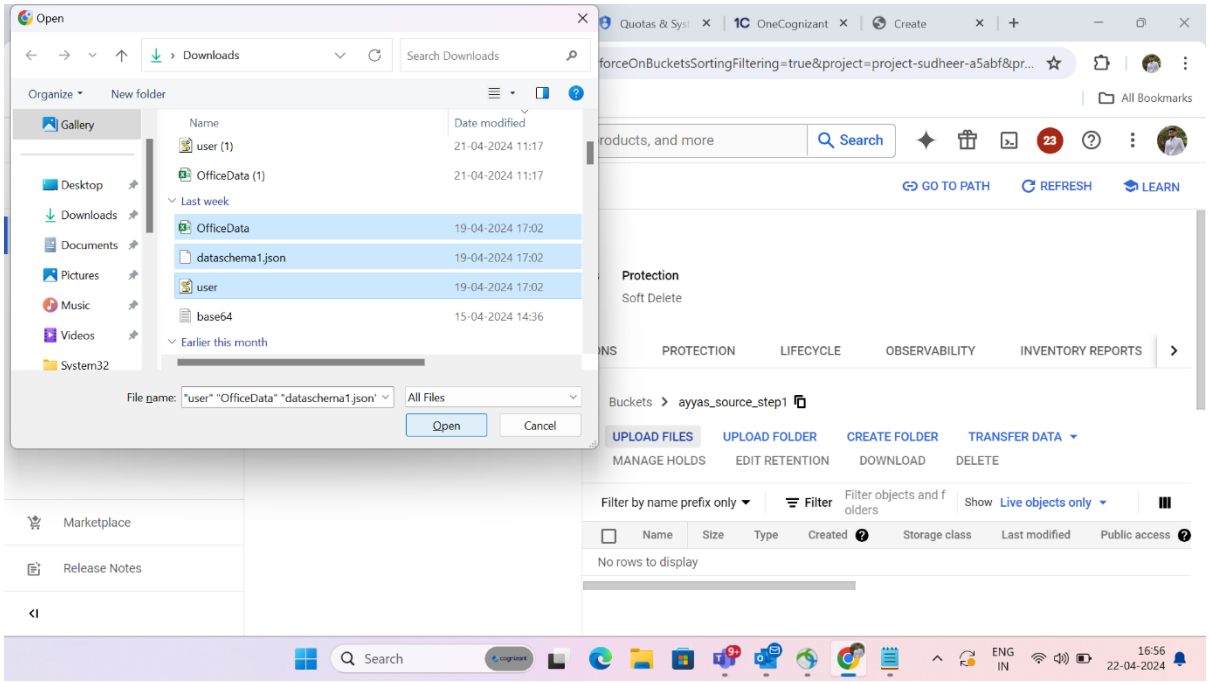

console.cloud.google.com/functions/details/us-east1/source_to_desti?env=gen2&project=project-sudheer-a5abf
Google Cloud
Project Sudheer
Search (/) for resources, docs, products, and more
Search
Cloud Functions
Function details
EDIT
DELETE
COPY
LEARN
Deploying function
Build
Build finished
Service
Creating Cloud Run service
Trigger
Creating Eventarc trigger
source_to_desti
2nd gen
(Deployment started at Apr 22, 2024, 4:54:00 PM)
URL: https://us-east1-project-sudheer-a5abf.cloudfunctions.net/source_to_desti
Powered by Cloud Run
source-to-desti
METRICS
DETAILS
SOURCE
VARIABLES
TRIGGER
PERMISSIONS
LOGS
TESTING
Dashboard
Predefined
1 hour
6 hours
12 hours
1 day
2 days
4 days
7 days
14 days
30 days
Invocations/Second
Execution time

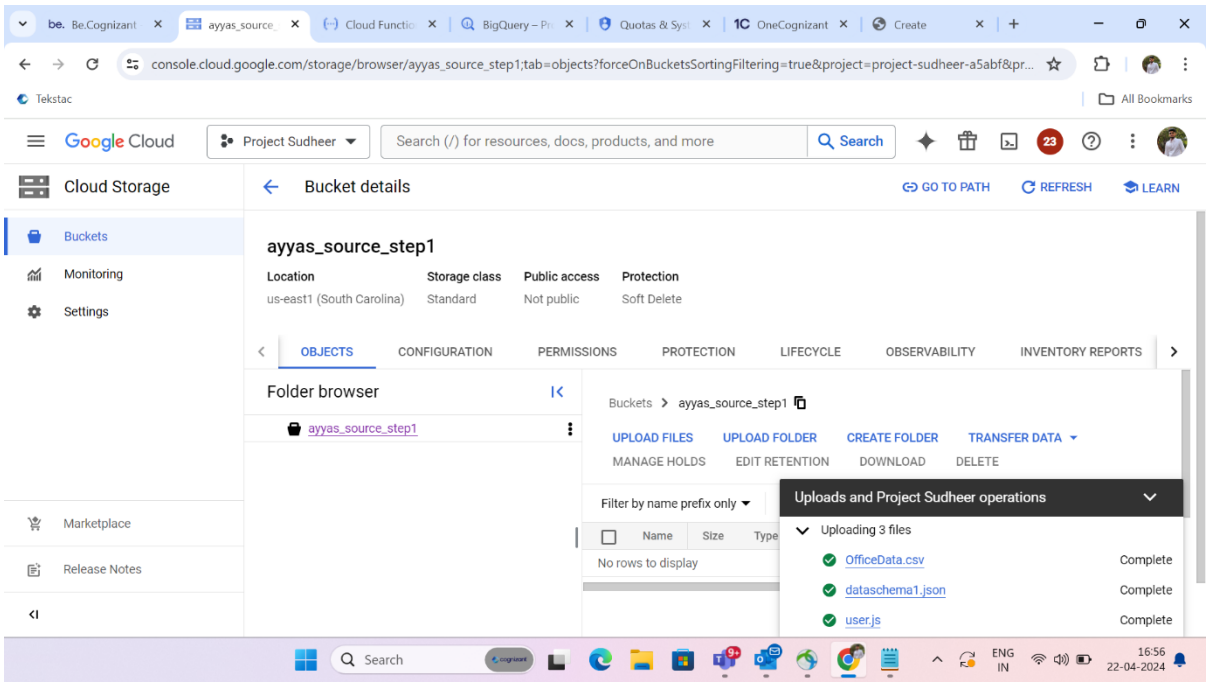

console.cloud.google.com/functions/details/us-east1/source_to_desti?env=gen2&project=project-sudheer-a5abf
Google Cloud
Project Sudheer
Search (/) for resources, docs, products, and more
Search
Cloud Functions
Function details
EDIT
DELETE
COPY
LEARN
source_to_desti
2nd gen
(Deployed at Apr 22, 2024, 4:55:36 PM)
URL: https://us-east1-project-sudheer-a5abf.cloudfunctions.net/source_to_desti
Powered by Cloud Run
source-to-desti
METRICS
DETAILS
SOURCE
VARIABLES
TRIGGER
PERMISSIONS
LOGS
TESTING
Dashboard
Predefined
1 hour
6 hours
12 hours
1 day
2 days
4 days
7 days
14 days
30 days
Invocations/Second
No data is available for the selected time frame.
Execution time
No data is available for the selected time frame.

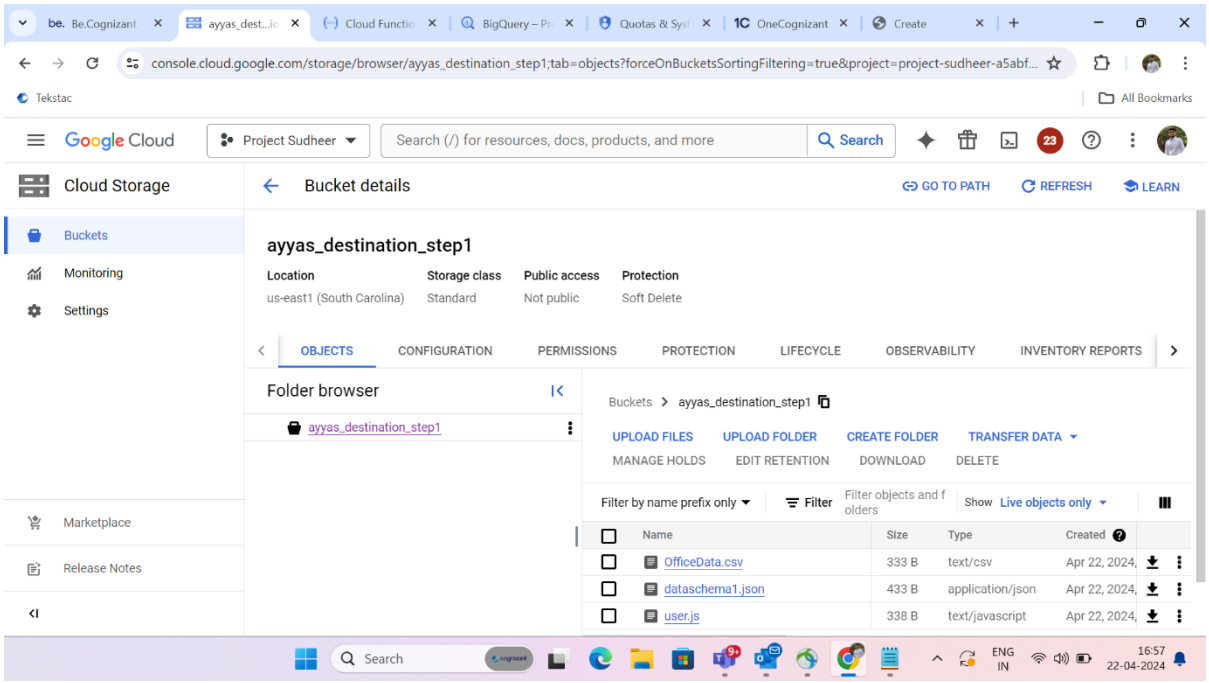
Upload files to Buckets:



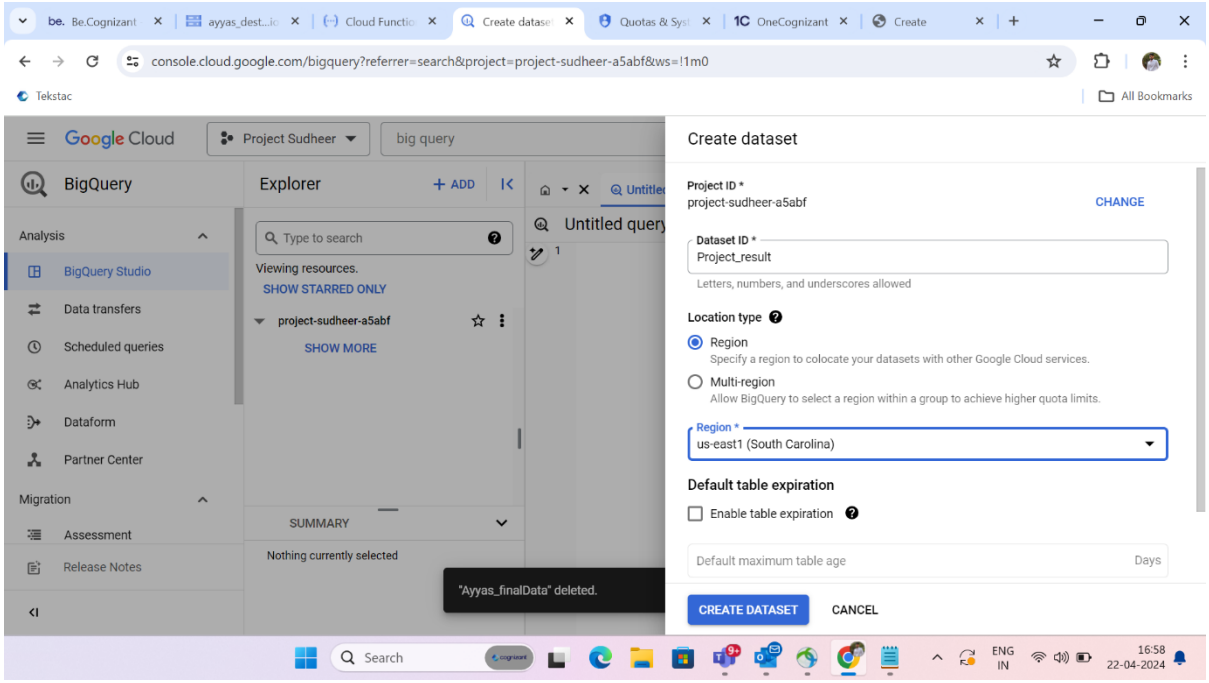


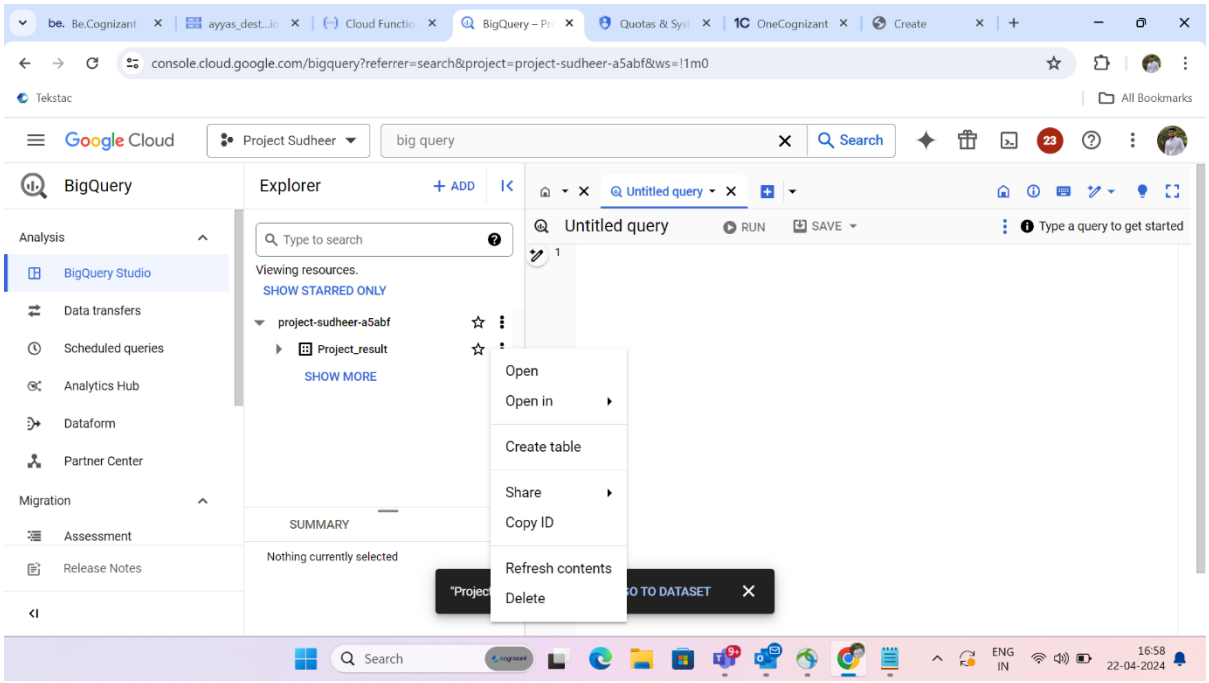

Open
Downloads
Search Downloads
Organize
New folder
Gallery
Desktop
Downloads
Documents
Pictures
Music
Videos
System32
Name
Date modified
user (1)
21-04-2024 11:17
OfficeData (1)
21-04-2024 11:17
Last week
OfficeData
19-04-2024 17:02
dataschema1.json
19-04-2024 17:02
user
19-04-2024 17:02
base64
15-04-2024 14:36
Earlier this month
File name: "user" "OfficeData" "dataschema1.json"
All Files
Open
Cancel
Search
GO TO PATH
REFRESH
LEARN
Protection
Soft Delete
PROTECTION
LIFECYCLE
OBSERVABILITY
INVENTORY REPORTS
Buckets
ayyas_source_step1
UPLOAD FILES
UPLOAD FOLDER
CREATE FOLDER
TRANSFER DATA
MANAGE HOLDS
EDIT RETENTION
DOWNLOAD
DELETE
Filter by name prefix only
Filter
Show
Live objects only
Name
Size
Type
Created
Storage class
Last modified
Public access
No rows to display
Marketplace
Release Notes

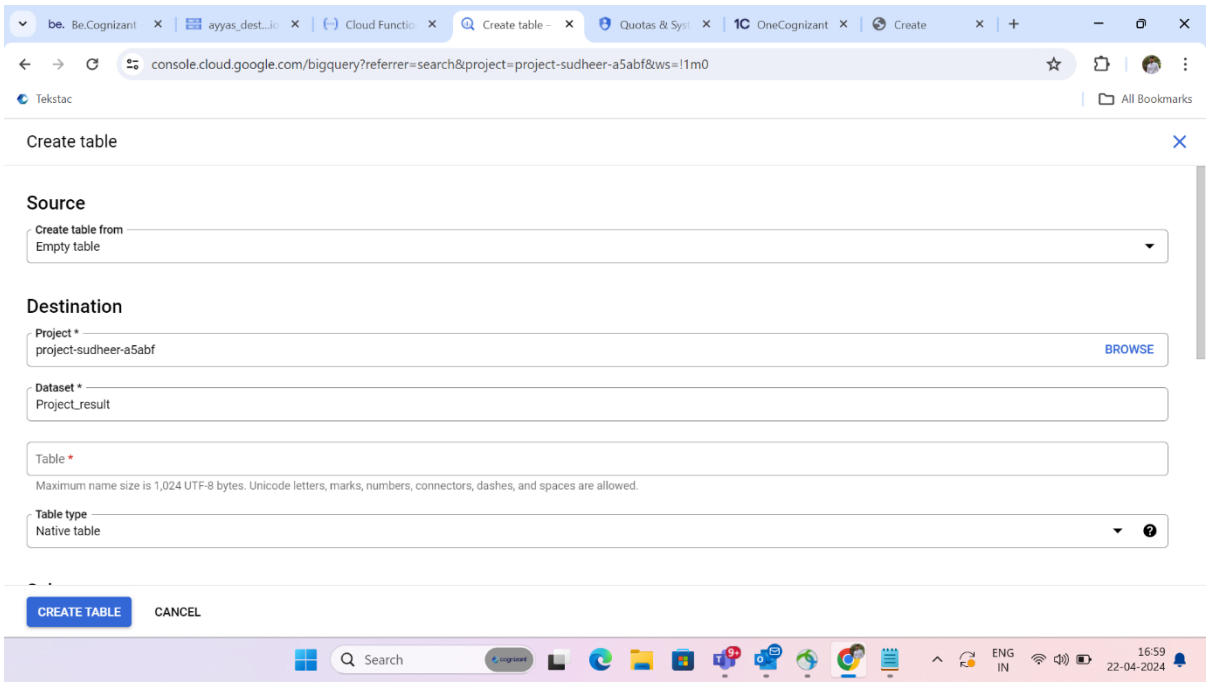

console.cloud.google.com/storage/browser/ayyas_source_step1;tab=objects?forceOnBucketsSortingFiltering=true&project=project-sudheer-a5abf&pr...
Google Cloud
Project Sudheer
Search (/) for resources, docs, products, and more
Search
Cloud Storage
Bucket details
GO TO PATH
REFRESH
LEARN
Buckets
Monitoring
Settings
ayyas_source_step1
Location
us-east1 (South Carolina)
Storage class
Standard
Public access
Not public
Protection
Soft Delete
OBJECTS
CONFIGURATION
PERMISSIONS
PROTECTION
LIFECYCLE
OBSERVABILITY
INVENTORY REPORTS
Folder browser
ayyas_source_step1
Buckets
ayyas_source_step1
UPLOAD FILES
UPLOAD FOLDER
CREATE FOLDER
TRANSFER DATA
MANAGE HOLDS
EDIT RETENTION
DOWNLOAD
DELETE
Filter by name prefix only
Uploads and Project Sudheer operations
Uploading 3 files
OfficeData.csv
Complete
dataschema1.json
Complete
user.js
Complete
Name
Size
No rows to display
Marketplace
Release Notes



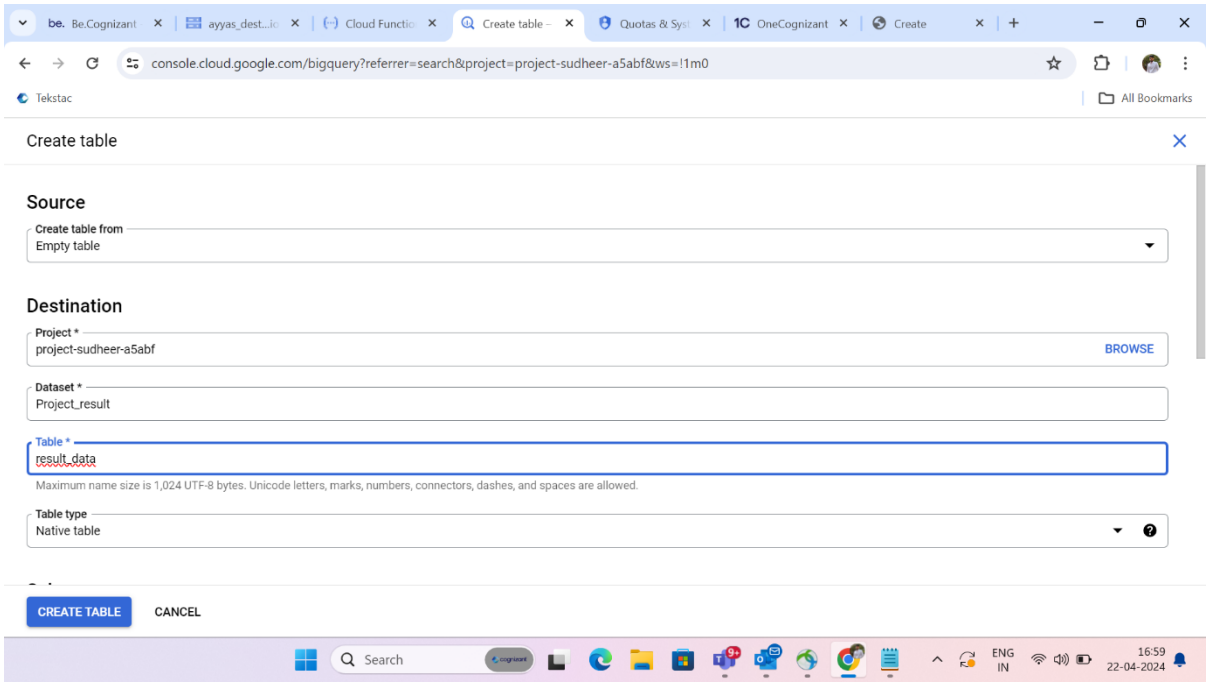
Create Dataset in Big query:

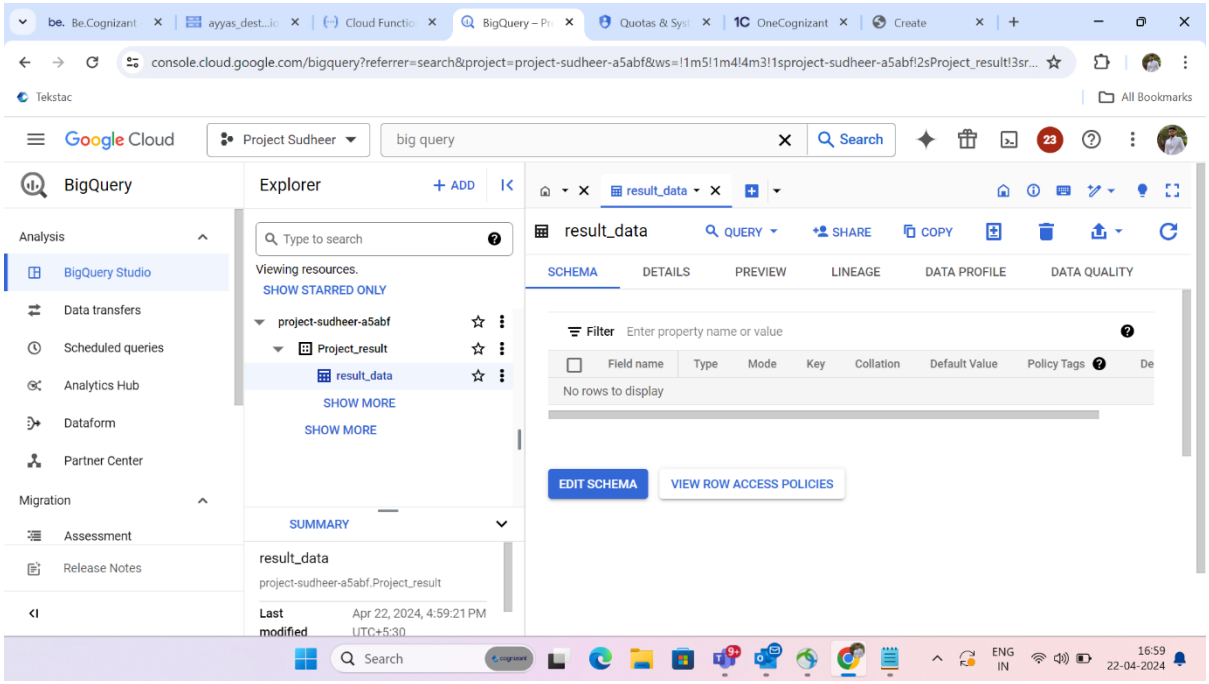


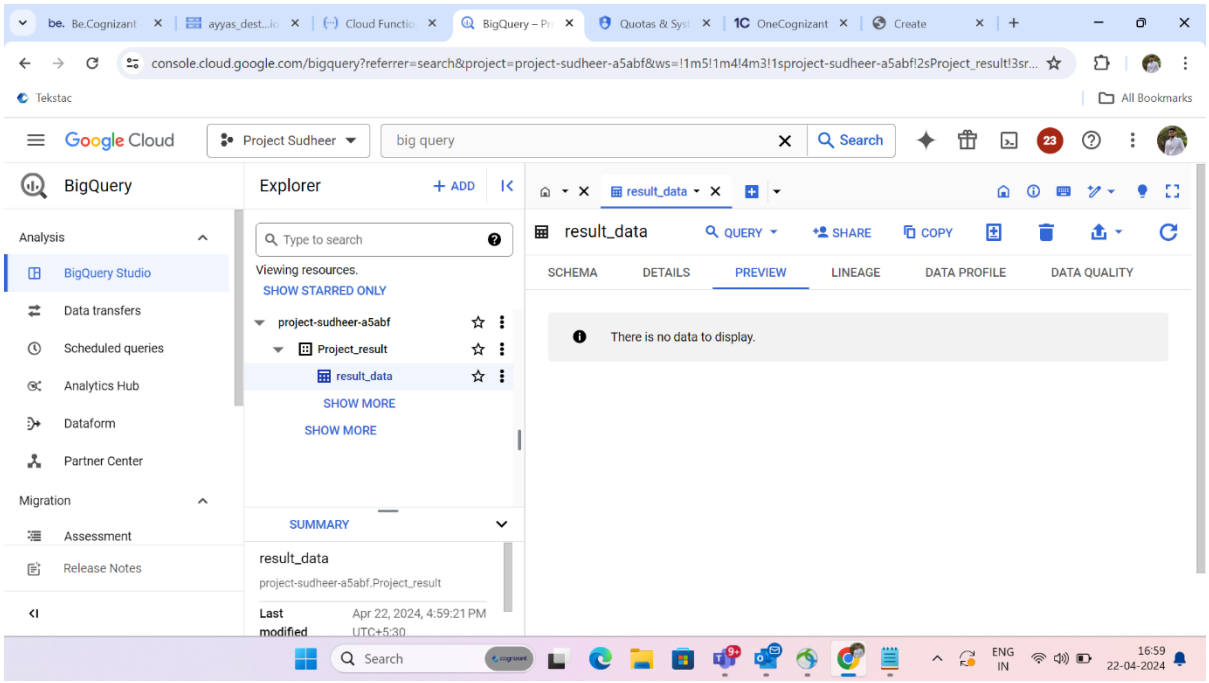

console.cloud.google.com/bigquery?referrer=search&project=project-sudheer-a5abf&ws=!1m0
Google Cloud
Project Sudheer
big query
Search
BigQuery
Explorer
+ ADD
Analysis
BigQuery Studio
Data transfers
Scheduled queries
Analytics Hub
Dataform
Partner Center
Migration
Assessment
Release Notes
Type to search
Viewing resources.
SHOW STARRED ONLY
project-sudheer-a5abf
Project_result
SHOW MORE
SUMMARY
Nothing currently selected
Untitled query
RUN
SAVE
Type a query to get started
Open
Open in
Create table
Share
Copy ID
Refresh contents
Delete
GO TO DATASET


console.cloud.google.com/bigquery?referrer=search&project=project-sudheer-a5abf&ws=!1m0
Create table
Source
Create table from
Empty table
Destination
Project *
project-sudheer-a5abf
BROWSE
Dataset *
Project_result
Table *
Maximum name size is 1,024 UTF-8 bytes. Unicode letters, marks, numbers, connectors, dashes, and spaces are allowed.
Table type
Native table
CREATE TABLE
CANCEL

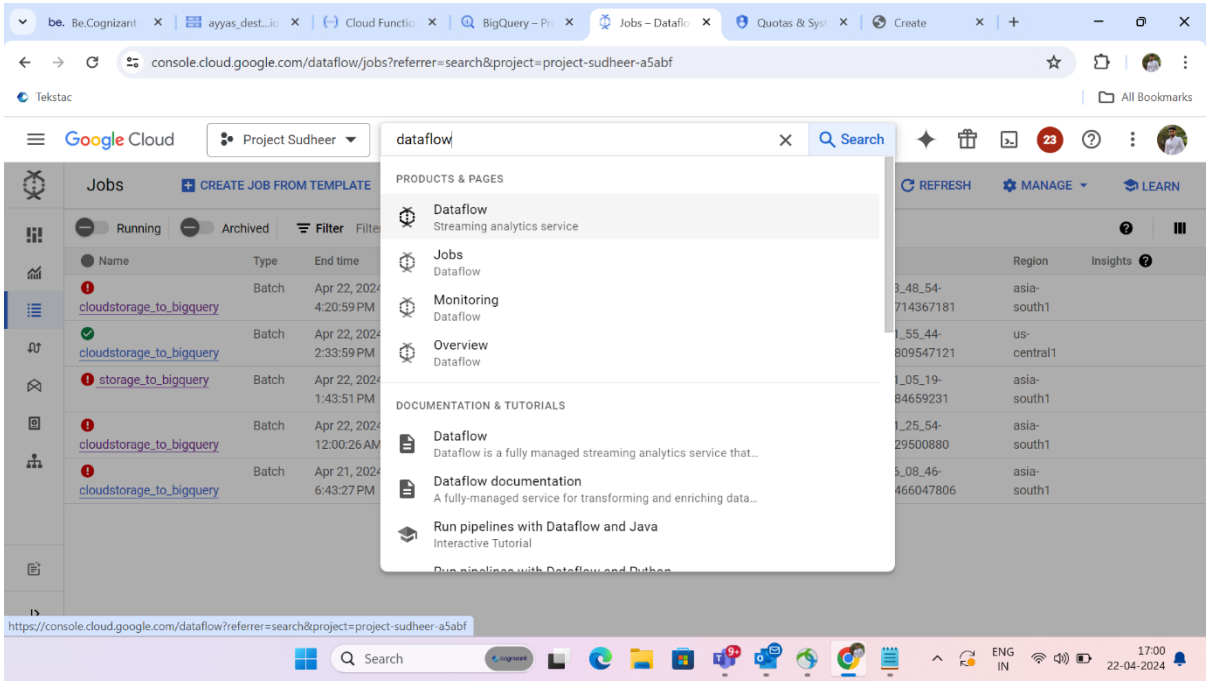
Create Tables:

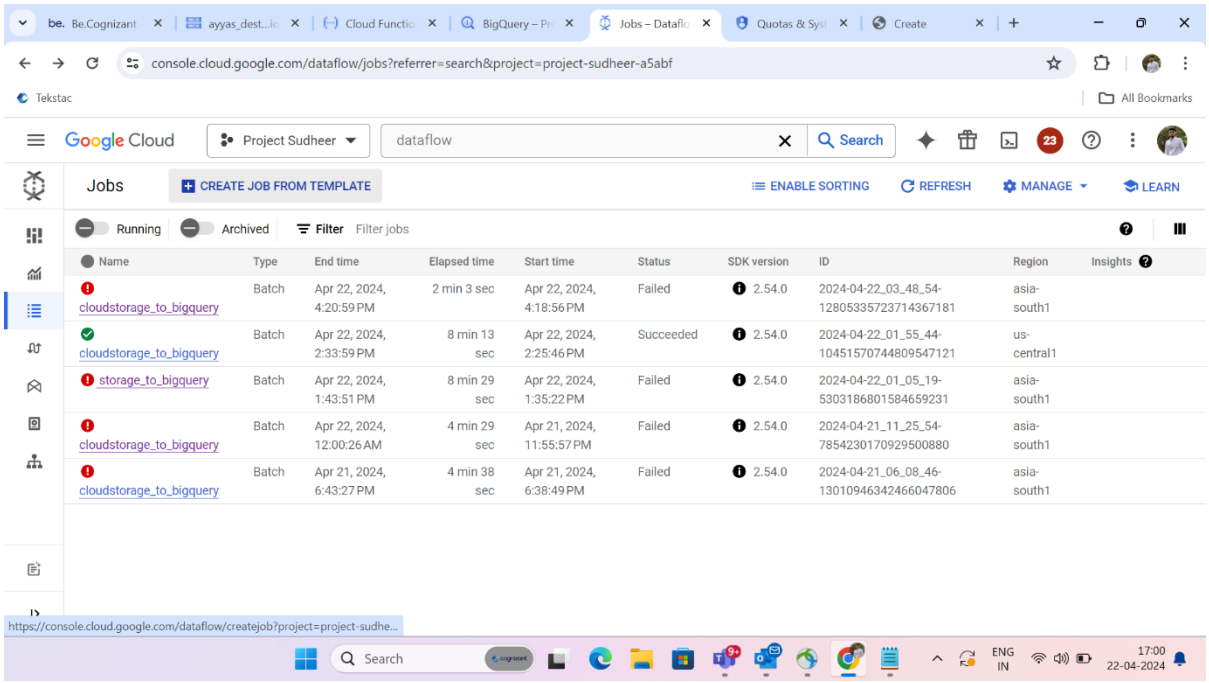


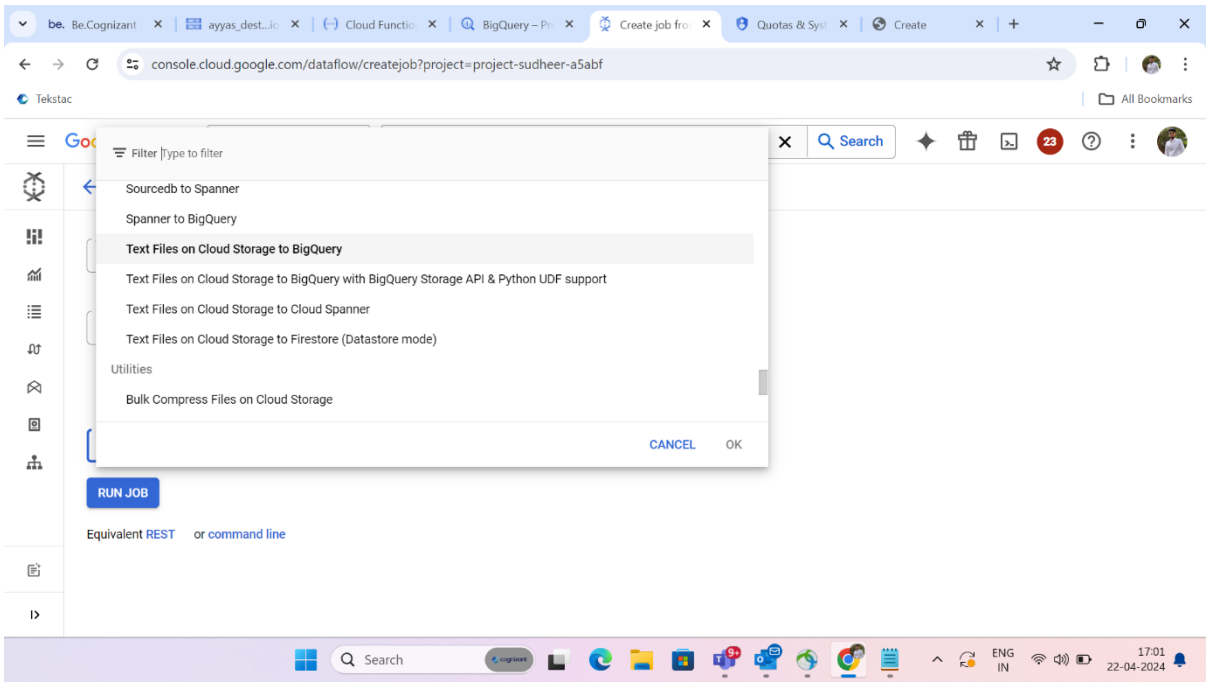


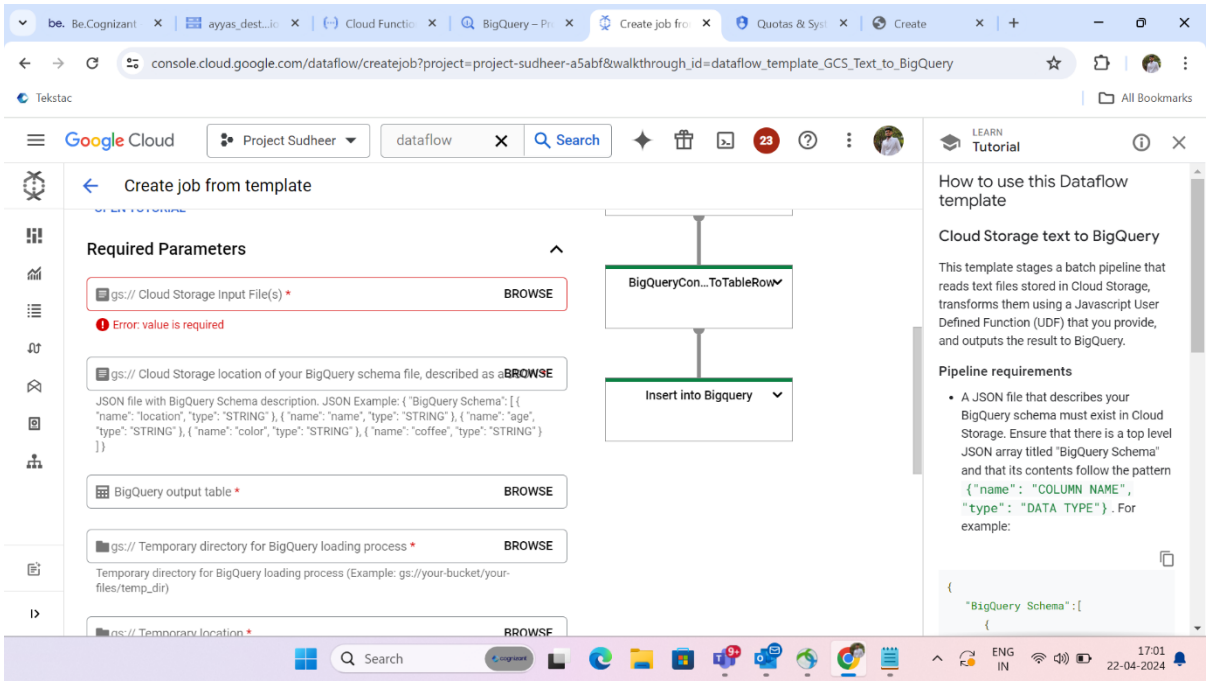


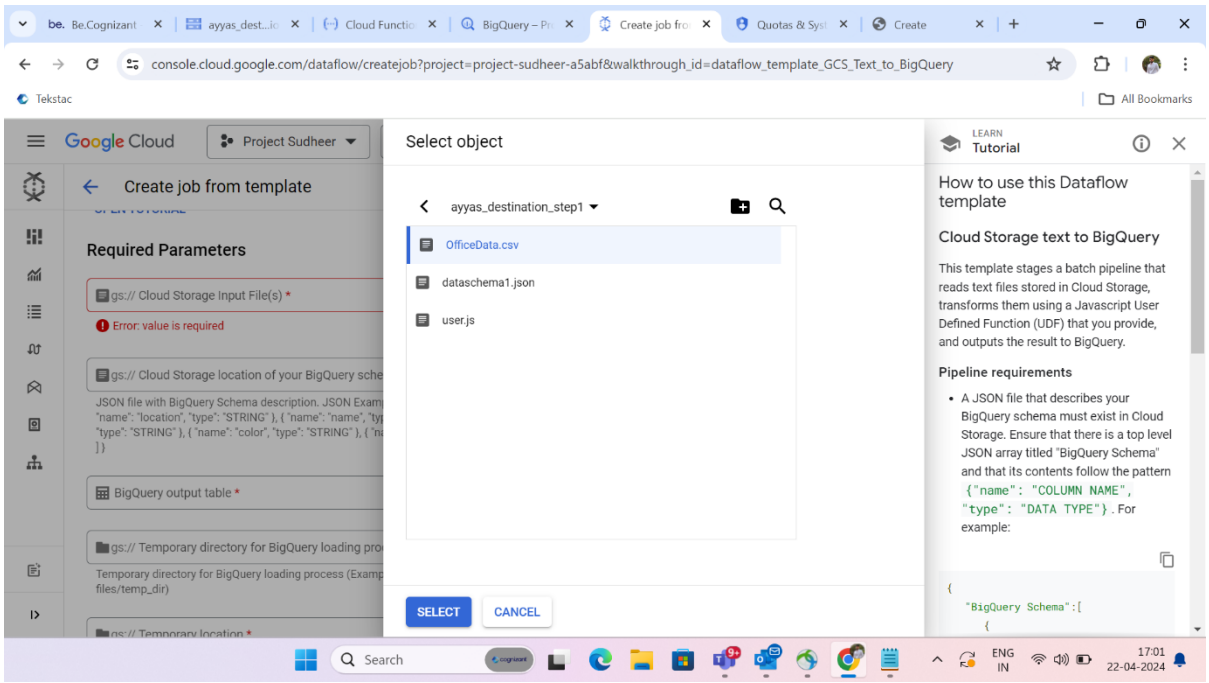
Create Dataflow:

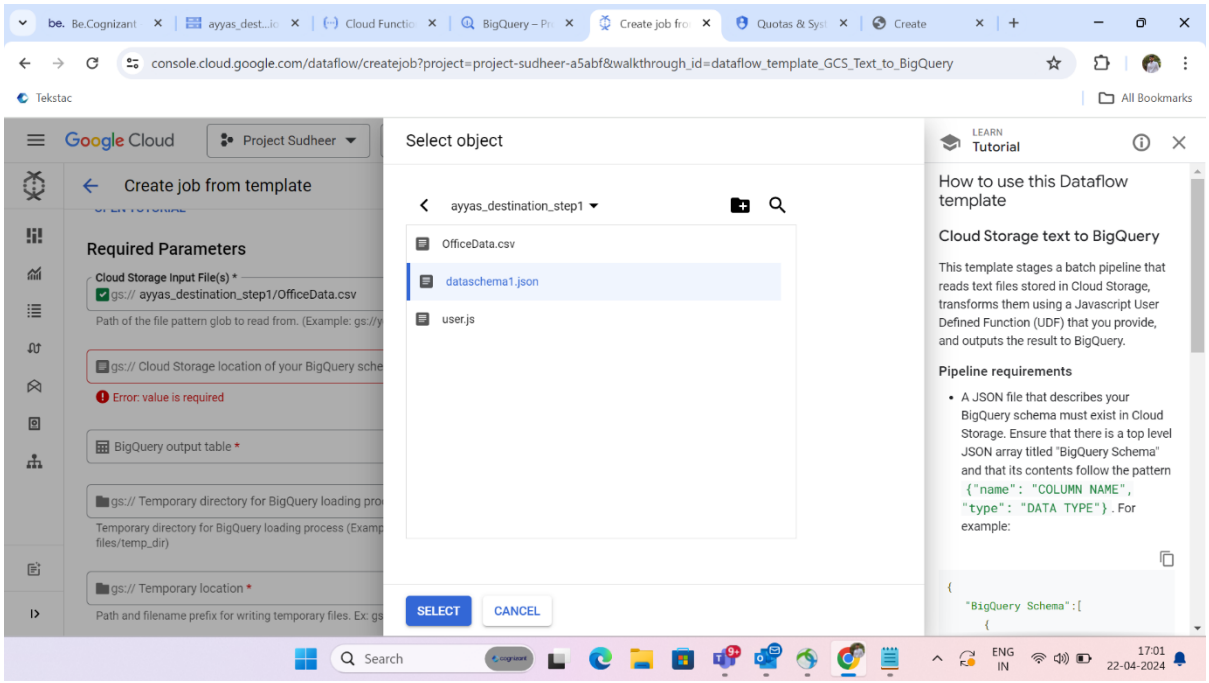


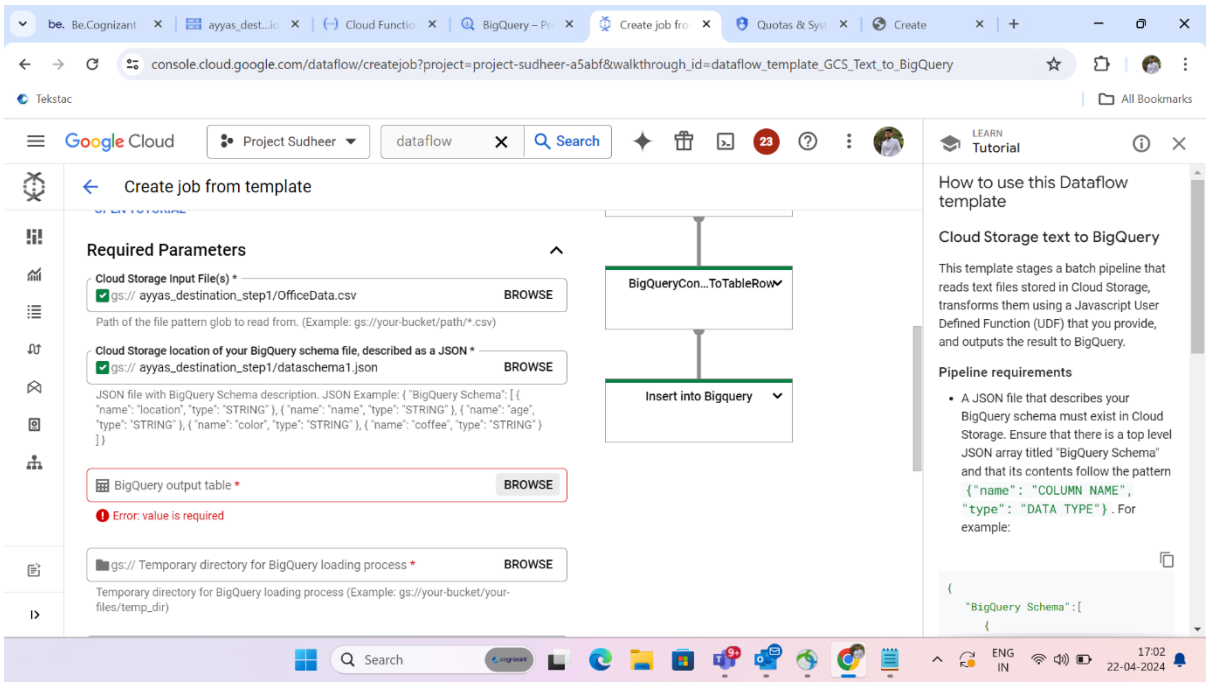

console.cloud.google.com/dataflow/jobs?referrer=search&project=project-sudheer-a5abf
Google Cloud
Project Sudheer
dataflow
Search
Jobs
CREATE JOB FROM TEMPLATE
ENABLE SORTING
REFRESH
MANAGE
LEARN
Running
Archived
Filter
Filter jobs
Name
Type
End time
Elapsed time
Start time
Status
SDK version
ID
Region
Insights
cloudstorage_to_bigquery
Batch
Apr 22, 2024, 4:20:59 PM
2 min 3 sec
Apr 22, 2024, 4:18:56 PM
Failed
2.54.0
2024-04-22_03_48_54-12805335723714367181
asia-south1
cloudstorage_to_bigquery
Batch
Apr 22, 2024, 2:33:59 PM
8 min 13 sec
Apr 22, 2024, 2:25:46 PM
Succeeded
2.54.0
2024-04-22_01_55_44-10451570744809547121
us-central1
storage_to_bigquery
Batch
Apr 22, 2024, 1:43:51 PM
8 min 29 sec
Apr 22, 2024, 1:35:22 PM
Failed
2.54.0
2024-04-22_01_05_19-5303186801584659231
asia-south1
cloudstorage_to_bigquery
Batch
Apr 22, 2024, 12:00:26 AM
4 min 29 sec
Apr 21, 2024, 11:55:57 PM
Failed
2.54.0
2024-04-21_11_25_54-7854230170929500880
asia-south1
cloudstorage_to_bigquery
Batch
Apr 21, 2024, 6:43:27 PM
4 min 38 sec
Apr 21, 2024, 6:38:49 PM
Failed
2.54.0
2024-04-21_06_08_46-13010946342466047806
asia-south1
https://console.cloud.google.com/dataflow/createjob?project=project-sudhe...
17:00
22-04-2024

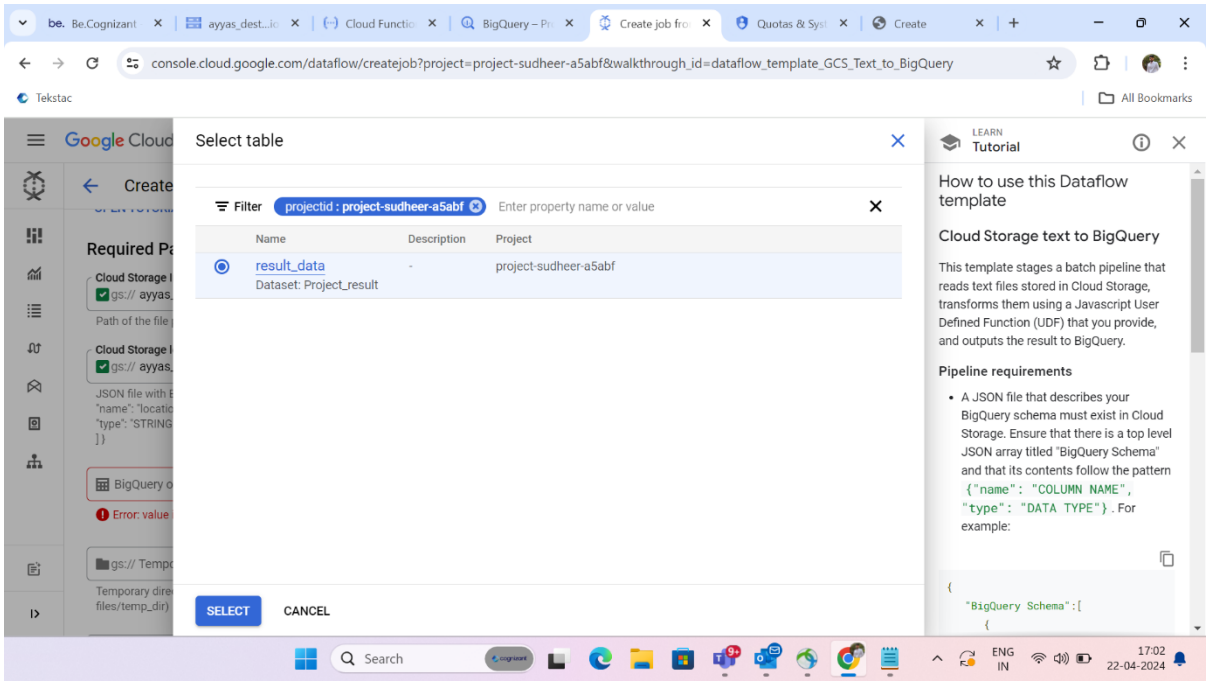

console.cloud.google.com/dataflow/createjob?project=project-sudheer-a5abf
Filter Type to filter
Sourcedb to Spanner
Spanner to BigQuery
Text Files on Cloud Storage to BigQuery
Text Files on Cloud Storage to BigQuery with BigQuery Storage API & Python UDF support
Text Files on Cloud Storage to Cloud Spanner
Text Files on Cloud Storage to Firestore (Datastore mode)
Utilities
Bulk Compress Files on Cloud Storage
CANCEL
OK
RUN JOB
Equivalent REST or command line
17:01
22-04-2024

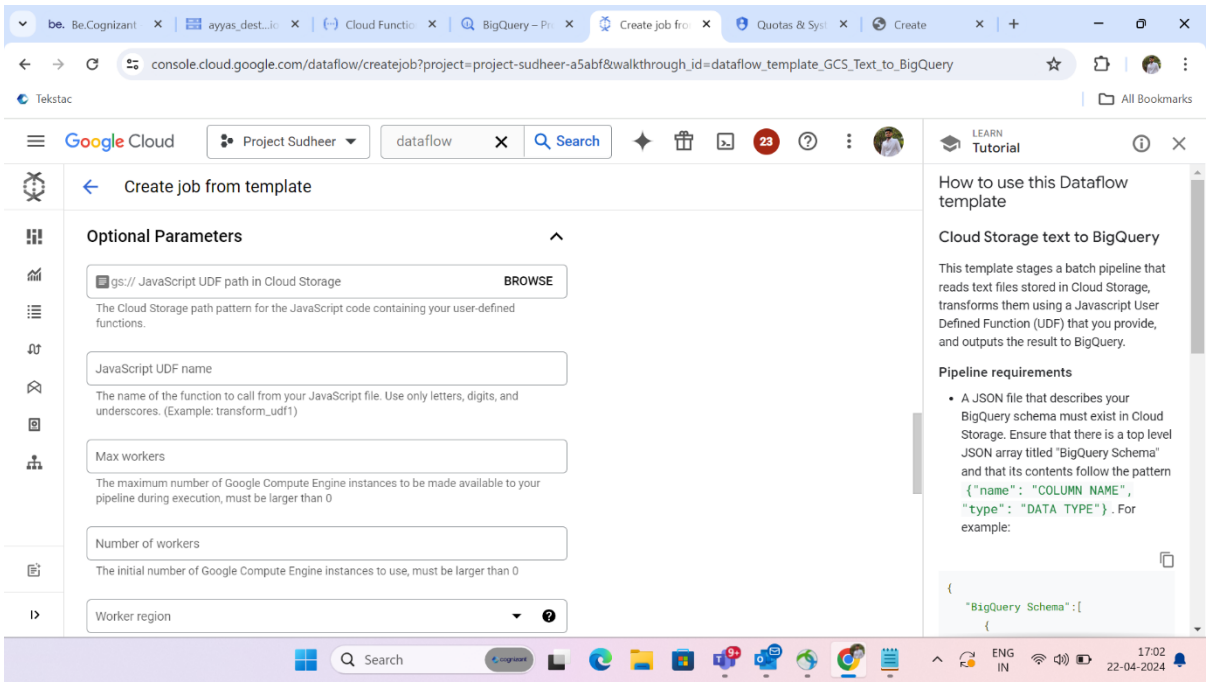

console.cloud.google.com/dataflow/createjob?project=project-sudheer-a5abf&walkthrough_id=dataflow_template_GCS_Text_to_BigQuery
Google Cloud
Project Sudheer
dataflow
Search
Create job from template
Required Parameters
gs:// Cloud Storage Input File(s) *
BROWSE
Error: value is required
gs:// Cloud Storage location of your BigQuery schema file, described as a
JSON file with BigQuery Schema description. JSON Example: { "BigQuery Schema": [{ "name": "location", "type": "STRING" }, { "name": "name", "type": "STRING" }, { "name": "age", "type": "STRING" }, { "name": "color", "type": "STRING" }, { "name": "coffee", "type": "STRING" }] }
BigQuery output table *
gs:// Temporary directory for BigQuery loading process *
Temporary directory for BigQuery loading process (Example: gs://your-bucket/your-files/temp_dir)
BigQueryCon...ToTableRow
Insert into Bigquery
LEARN
Tutorial
How to use this Dataflow template
Cloud Storage text to BigQuery
This template stages a batch pipeline that reads text files stored in Cloud Storage, transforms them using a Javascript User Defined Function (UDF) that you provide, and outputs the result to BigQuery.
Pipeline requirements
A JSON file that describes your BigQuery schema must exist in Cloud Storage. Ensure that there is a top level JSON array titled "BigQuery Schema" and that its contents follow the pattern {"name": "COLUMN NAME", "type": "DATA TYPE"}. For example:
{
"BigQuery Schema":[
{
Search
ENG
IN
17:01
22-04-2024

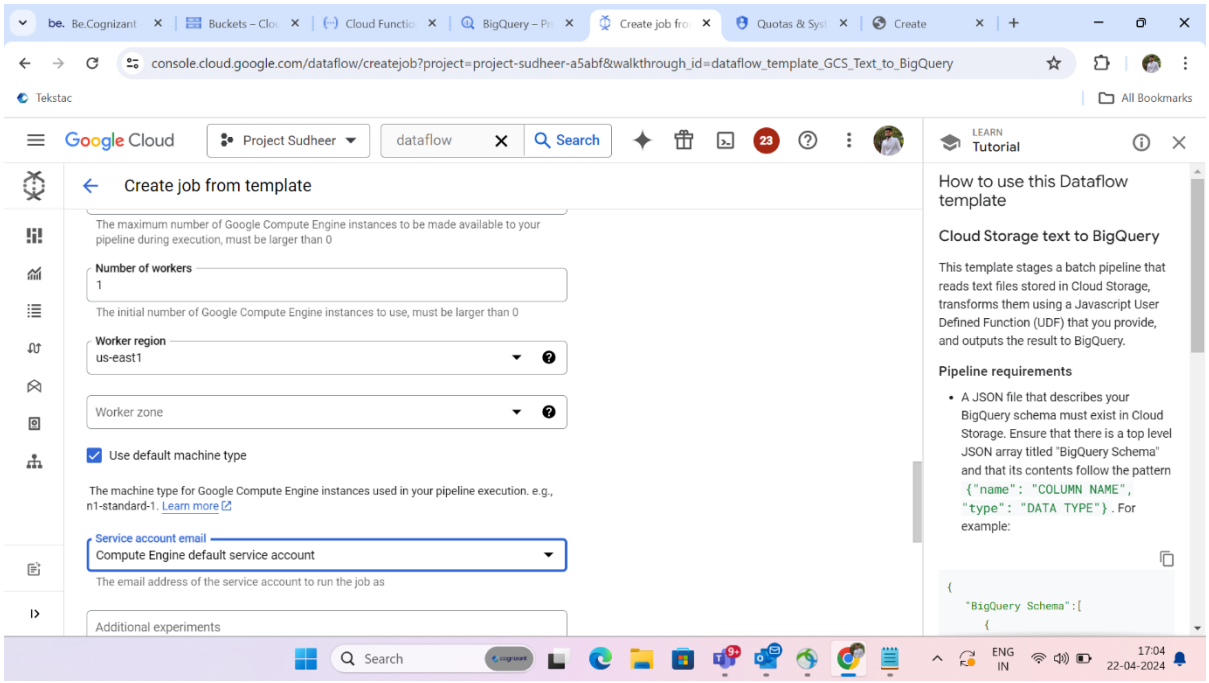

console.cloud.google.com/dataflow/createjob?project=project-sudheer-a5abf&walkthrough_id=dataflow_template_GCS_Text_to_BigQuery
Google Cloud
Project Sudheer
Create job from template
Required Parameters
gs:// Cloud Storage Input File(s) *
Error: value is required
BigQuery output table *
Select object
ayyas_destination_step1
OfficeData.csv
dataschema1.json
user.js
SELECT
CANCEL
LEARN
Tutorial
How to use this Dataflow template
Cloud Storage text to BigQuery
This template stages a batch pipeline that reads text files stored in Cloud Storage, transforms them using a Javascript User Defined Function (UDF) that you provide, and outputs the result to BigQuery.
Pipeline requirements
A JSON file that describes your BigQuery schema must exist in Cloud Storage. Ensure that there is a top level JSON array titled "BigQuery Schema" and that its contents follow the pattern {"name": "COLUMN NAME", "type": "DATA TYPE"}. For example:
{
"BigQuery Schema":[
{
Search
ENG
IN
17:01
22-04-2024

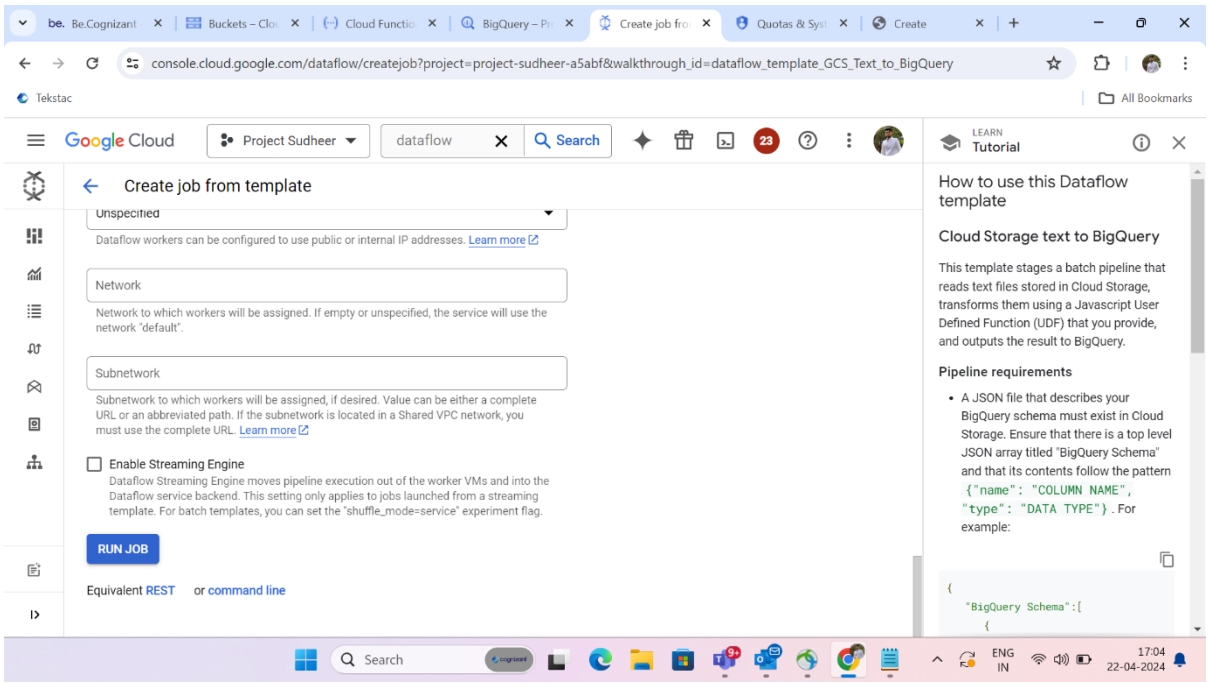

console.cloud.google.com/dataflow/createjob?project=project-sudheer-a5abf&walkthrough_id=dataflow_template_GCS_Text_to_BigQuery
Google Cloud
Project Sudheer
Create job from template
Required Parameters
Cloud Storage Input File(s) *
gs:// ayyas_destination_step1/OfficeData.csv
Path of the file pattern glob to read from. (Example: gs://y
gs:// Cloud Storage location of your BigQuery sche
Error: value is required
BigQuery output table *
gs:// Temporary directory for BigQuery loading pro
Temporary directory for BigQuery loading process (Examp
files/temp_dir)
gs:// Temporary location *
Path and filename prefix for writing temporary files. Ex: gs
Select object
ayyas_destination_step1
OfficeData.csv
dataschema1.json
user.js
SELECT
CANCEL
LEARN
Tutorial
How to use this Dataflow template
Cloud Storage text to BigQuery
This template stages a batch pipeline that reads text files stored in Cloud Storage, transforms them using a Javascript User Defined Function (UDF) that you provide, and outputs the result to BigQuery.
Pipeline requirements
A JSON file that describes your BigQuery schema must exist in Cloud Storage. Ensure that there is a top level JSON array titled "BigQuery Schema" and that its contents follow the pattern {"name": "COLUMN NAME", "type": "DATA TYPE"}. For example:
{
"BigQuery Schema":[
{
Search
ENG
IN
17:01
22-04-2024

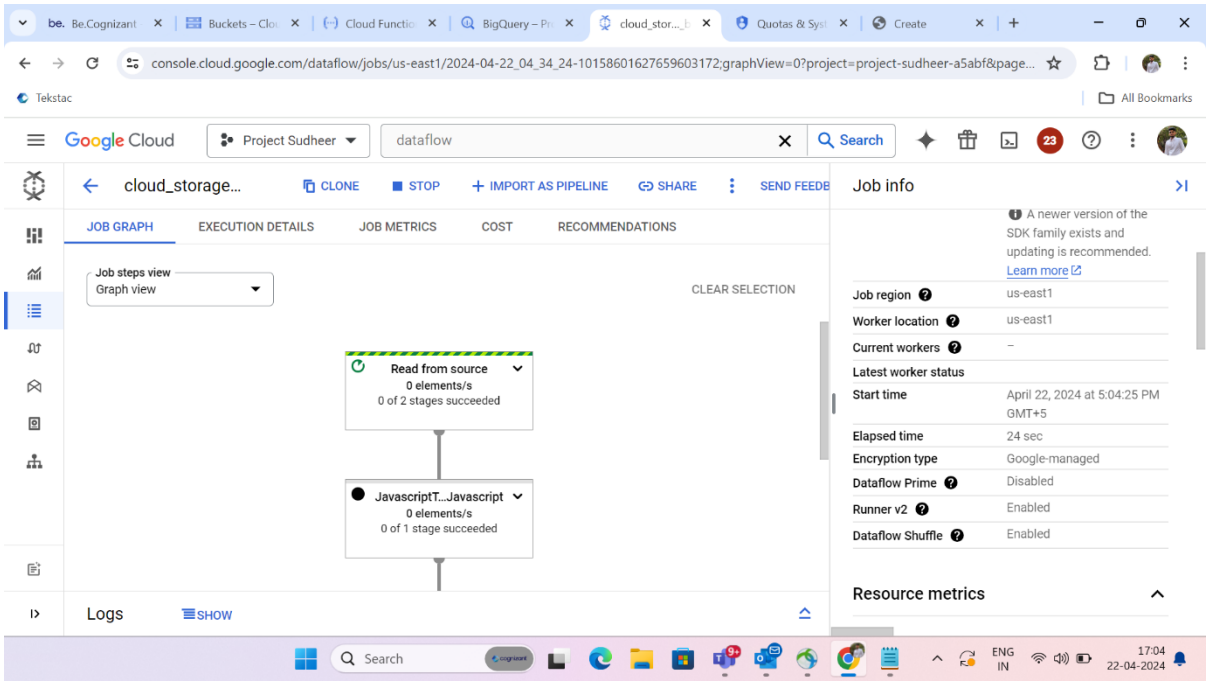

console.cloud.google.com/dataflow/createjob?project=project-sudheer-a5abf&walkthrough_id=dataflow_template_GCS_Text_to_BigQuery
Google Cloud
Project Sudheer
dataflow
Search
Create job from template
Required Parameters
Cloud Storage Input File(s) *
gs:// ayyas_destination_step1/OfficeData.csv
BROWSE
Path of the file pattern glob to read from. (Example: gs://your-bucket/path/*.csv)
Cloud Storage location of your BigQuery schema file, described as a JSON *
gs:// ayyas_destination_step1/dataschema1.json
BROWSE
JSON file with BigQuery Schema description. JSON Example: { "BigQuery Schema": [{ "name": "location", "type": "STRING" }, { "name": "name", "type": "STRING" }, { "name": "age", "type": "STRING" }, { "name": "color", "type": "STRING" }, { "name": "coffee", "type": "STRING" }] }
BigQuery output table *
BROWSE
Error: value is required
gs:// Temporary directory for BigQuery loading process *
BROWSE
Temporary directory for BigQuery loading process (Example: gs://your-bucket/your-files/temp_dir)
BigQueryCon...ToTableRow
Insert into Bigquery
LEARN
Tutorial
How to use this Dataflow template
Cloud Storage text to BigQuery
This template stages a batch pipeline that reads text files stored in Cloud Storage, transforms them using a Javascript User Defined Function (UDF) that you provide, and outputs the result to BigQuery.
Pipeline requirements
A JSON file that describes your BigQuery schema must exist in Cloud Storage. Ensure that there is a top level JSON array titled "BigQuery Schema" and that its contents follow the pattern {"name": "COLUMN NAME", "type": "DATA TYPE"}. For example:
{
"BigQuery Schema":[
{
Search
ENG
IN
17:02
22-04-2024

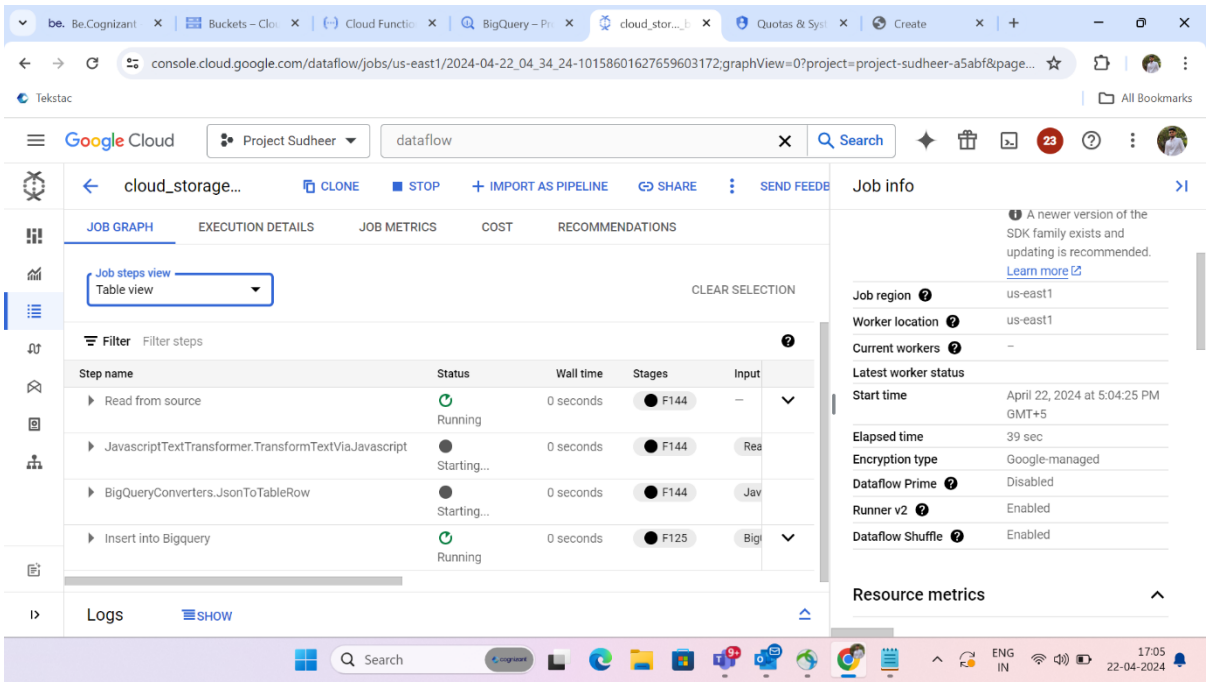

console.cloud.google.com/dataflow/createjob?project=project-sudheer-a5abf&walkthrough_id=dataflow_template_GCS_Text_to_BigQuery
Select table
Filter
projectid : project-sudheer-a5abf
Enter property name or value
Name
Description
Project
result_data
Dataset: Project_result
project-sudheer-a5abf
SELECT
CANCEL
LEARN
Tutorial
How to use this Dataflow template
Cloud Storage text to BigQuery
This template stages a batch pipeline that reads text files stored in Cloud Storage, transforms them using a Javascript User Defined Function (UDF) that you provide, and outputs the result to BigQuery.
Pipeline requirements
A JSON file that describes your BigQuery schema must exist in Cloud Storage. Ensure that there is a top level JSON array titled "BigQuery Schema" and that its contents follow the pattern {"name": "COLUMN NAME", "type": "DATA TYPE"}. For example:
{
"BigQuery Schema":[
{
Search
ENG
IN
17:02
22-04-2024

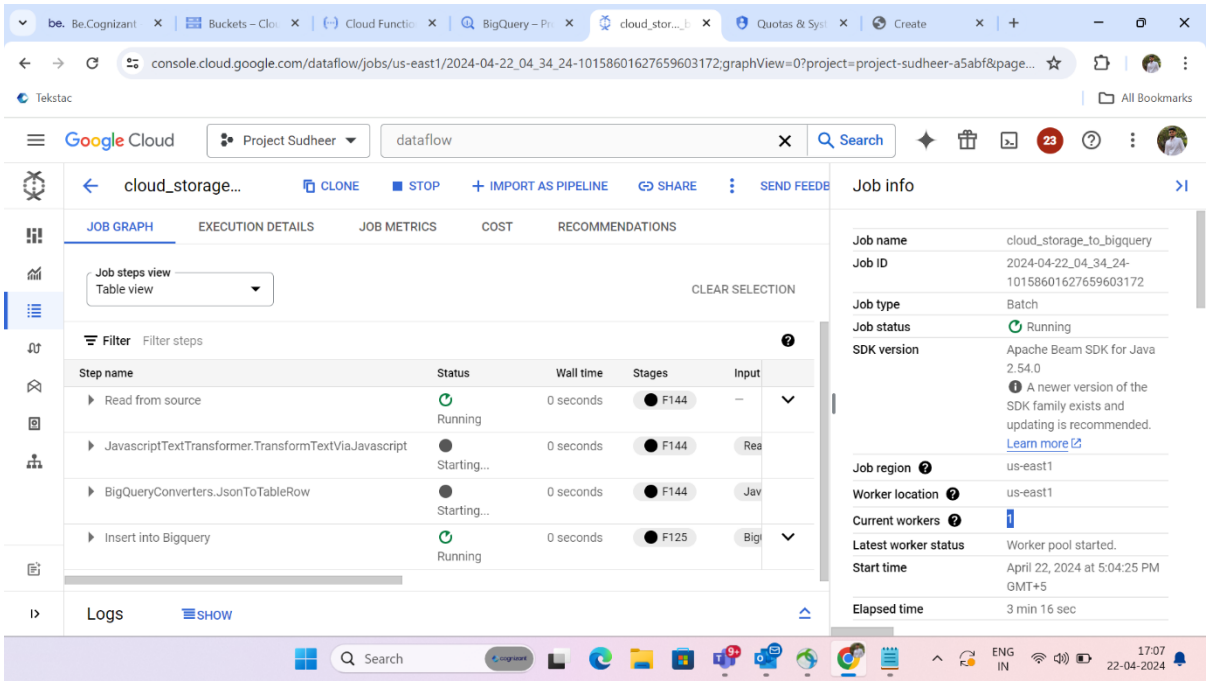

console.cloud.google.com/dataflow/createjob?project=project-sudheer-a5abf&walkthrough_id=dataflow_template_GCS_Text_to_BigQuery
Google Cloud
Project Sudheer
dataflow
Search
Create job from template
Optional Parameters
gs:// JavaScript UDF path in Cloud Storage
BROWSE
The Cloud Storage path pattern for the JavaScript code containing your user-defined functions.
JavaScript UDF name
The name of the function to call from your JavaScript file. Use only letters, digits, and underscores. (Example: transform_udf1)
Max workers
The maximum number of Google Compute Engine instances to be made available to your pipeline during execution, must be larger than 0
Number of workers
The initial number of Google Compute Engine instances to use, must be larger than 0
Worker region
LEARN
Tutorial
How to use this Dataflow template
Cloud Storage text to BigQuery
This template stages a batch pipeline that reads text files stored in Cloud Storage, transforms them using a Javascript User Defined Function (UDF) that you provide, and outputs the result to BigQuery.
Pipeline requirements
A JSON file that describes your BigQuery schema must exist in Cloud Storage. Ensure that there is a top level JSON array titled "BigQuery Schema" and that its contents follow the pattern {"name": "COLUMN NAME", "type": "DATA TYPE"}. For example:
{
"BigQuery Schema":[
{
Search
ENG
IN
17:02
22-04-2024

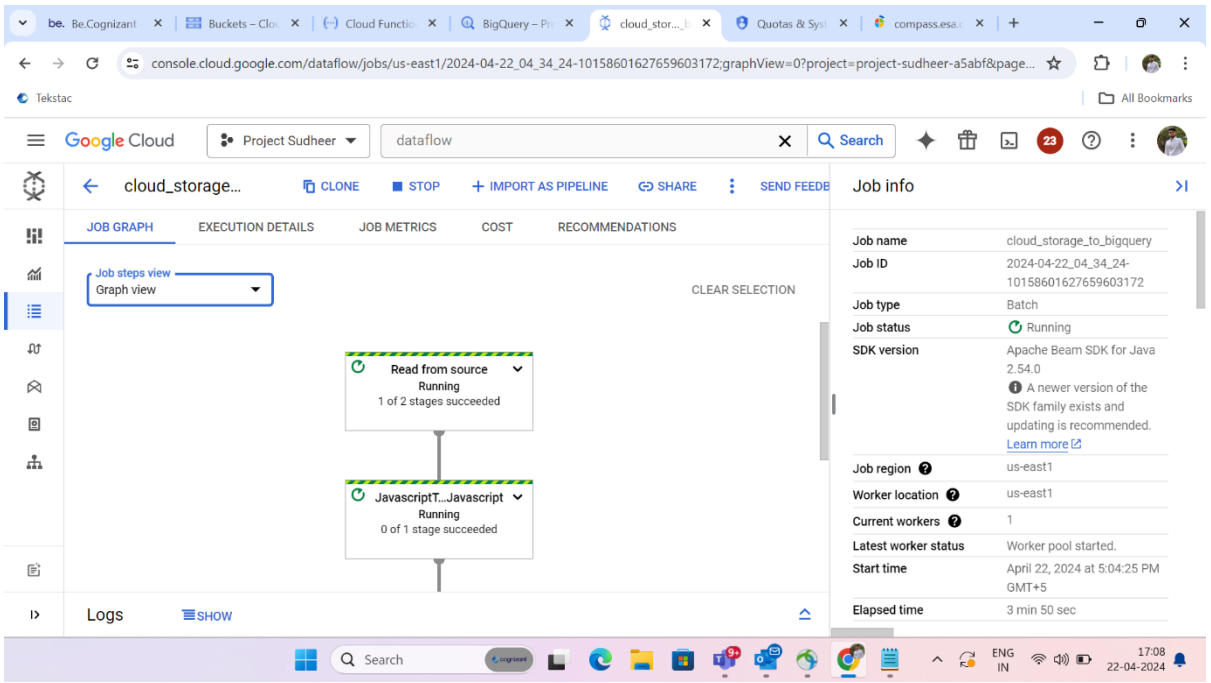

console.cloud.google.com/dataflow/createjob?project=project-sudheer-a5abf&walkthrough_id=dataflow_template_GCS_Text_to_BigQuery
Google Cloud
Project Sudheer
dataflow
Search
Create job from template
The maximum number of Google Compute Engine instances to be made available to your pipeline during execution, must be larger than 0
Number of workers
1
The initial number of Google Compute Engine instances to use, must be larger than 0
Worker region
us-east1
Worker zone
Use default machine type
The machine type for Google Compute Engine instances used in your pipeline execution. e.g., n1-standard-1. Learn more
Service account email
Compute Engine default service account
The email address of the service account to run the job as
Additional experiments
LEARN
Tutorial
How to use this Dataflow template
Cloud Storage text to BigQuery
This template stages a batch pipeline that reads text files stored in Cloud Storage, transforms them using a Javascript User Defined Function (UDF) that you provide, and outputs the result to BigQuery.
Pipeline requirements
A JSON file that describes your BigQuery schema must exist in Cloud Storage. Ensure that there is a top level JSON array titled "BigQuery Schema" and that its contents follow the pattern {"name": "COLUMN NAME", "type": "DATA TYPE"}. For example:
{
"BigQuery Schema":[
{
Search
ENG
IN
17:04
22-04-2024

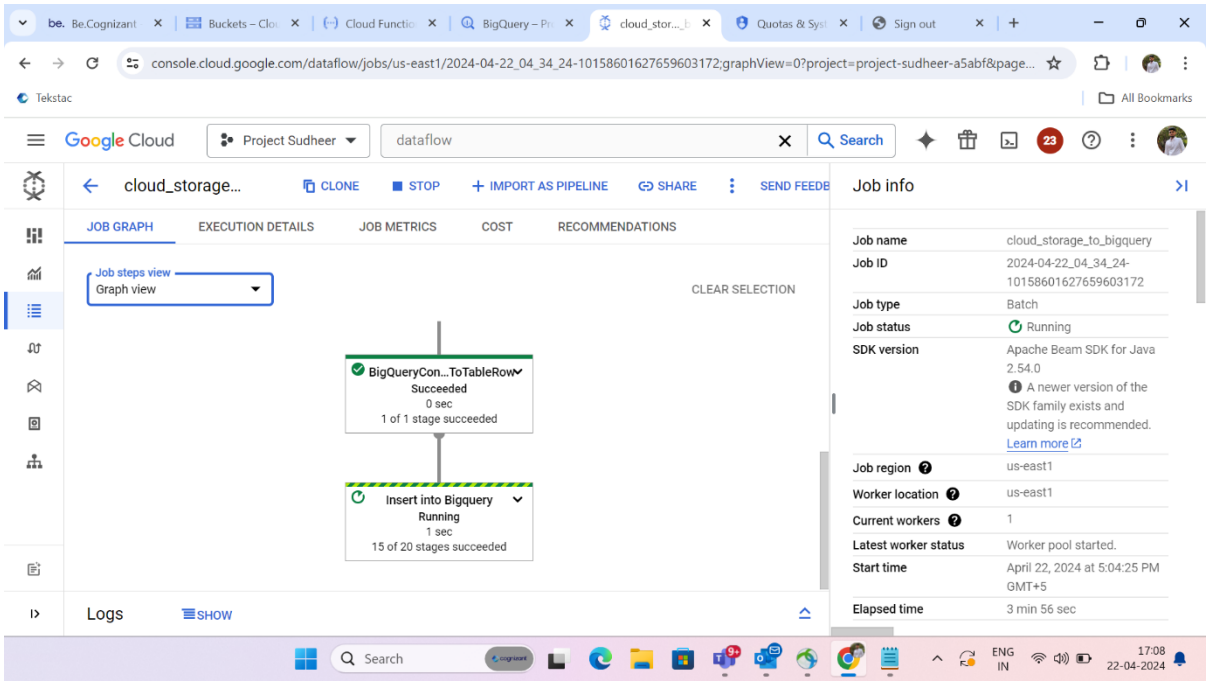

console.cloud.google.com/dataflow/createjob?project=project-sudheer-a5abf&walkthrough_id=dataflow_template_GCS_Text_to_BigQuery
Google Cloud
Project Sudheer
dataflow
Search
Create job from template
Unspecified
Dataflow workers can be configured to use public or internal IP addresses. Learn more
Network
Network to which workers will be assigned. If empty or unspecified, the service will use the network "default".
Subnetwork
Subnetwork to which workers will be assigned, if desired. Value can be either a complete URL or an abbreviated path. If the subnetwork is located in a Shared VPC network, you must use the complete URL. Learn more
Enable Streaming Engine
Dataflow Streaming Engine moves pipeline execution out of the worker VMs and into the Dataflow service backend. This setting only applies to jobs launched from a streaming template. For batch templates, you can set the "shuffle_mode=service" experiment flag.
RUN JOB
Equivalent REST or command line
LEARN
Tutorial
How to use this Dataflow template
Cloud Storage text to BigQuery
This template stages a batch pipeline that reads text files stored in Cloud Storage, transforms them using a Javascript User Defined Function (UDF) that you provide, and outputs the result to BigQuery.
Pipeline requirements
A JSON file that describes your BigQuery schema must exist in Cloud Storage. Ensure that there is a top level JSON array titled "BigQuery Schema" and that its contents follow the pattern {"name": "COLUMN NAME", "type": "DATA TYPE"}. For example:
{
"BigQuery Schema":[
{
Search
ENG
IN
17:04
22-04-2024

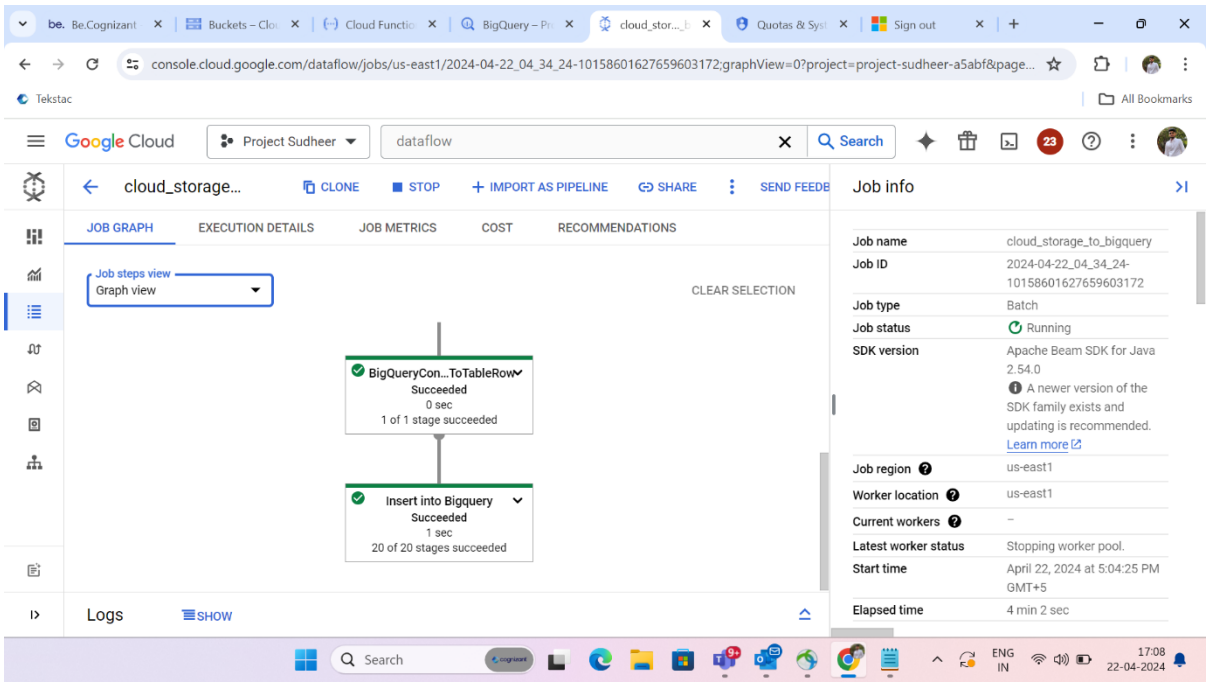

console.cloud.google.com/dataflow/jobs/us-east1/2024-04-22_04_34_24-10158601627659603172;graphView=0?project=project-sudheer-a5abf&page...
Google Cloud
Project Sudheer
dataflow
Search
cloud_storage...
CLONE
STOP
IMPORT AS PIPELINE
SHARE
SEND FEEDB
JOB GRAPH
EXECUTION DETAILS
JOB METRICS
COST
RECOMMENDATIONS
Job steps view
Graph view
CLEAR SELECTION
Read from source
0 elements/s
0 of 2 stages succeeded
JavascriptT...Javascript
0 elements/s
0 of 1 stage succeeded
Logs
SHOW
Job info
A newer version of the SDK family exists and updating is recommended.
Learn more
Job region us-east1
Worker location us-east1
Current workers –
Latest worker status
Start time April 22, 2024 at 5:04:25 PM GMT+5
Elapsed time 24 sec
Encryption type Google-managed
Dataflow Prime Disabled
Runner v2 Enabled
Dataflow Shuffle Enabled
Resource metrics
17:04
22-04-2024

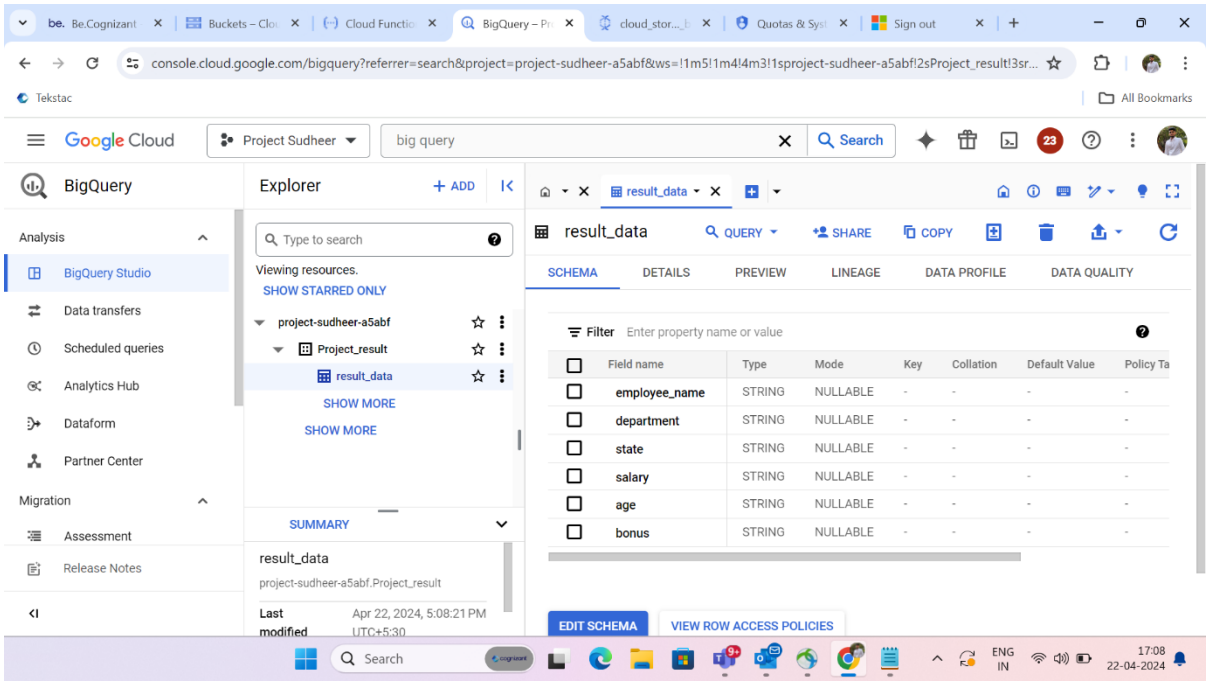

console.cloud.google.com/dataflow/jobs/us-east1/2024-04-22_04_34_24-10158601627659603172;graphView=0?project=project-sudheer-a5abf&page...
Google Cloud
Project Sudheer
dataflow
Search
cloud_storage...
CLONE
STOP
IMPORT AS PIPELINE
SHARE
SEND FEEDB
JOB GRAPH
EXECUTION DETAILS
JOB METRICS
COST
RECOMMENDATIONS
Job steps view
Table view
CLEAR SELECTION
Filter
Filter steps
Step name
Status
Wall time
Stages
Input
Read from source
Running
0 seconds
F144
JavascriptTextTransformer.TransformTextViaJavascript
Starting...
0 seconds
F144
BigQueryConverters.JsonToTableRow
Starting...
0 seconds
F144
Insert into Bigquery
Running
0 seconds
F125
Logs
SHOW
Job info
A newer version of the SDK family exists and updating is recommended.
Learn more
Job region us-east1
Worker location us-east1
Current workers –
Latest worker status
Start time April 22, 2024 at 5:04:25 PM GMT+5
Elapsed time 39 sec
Encryption type Google-managed
Dataflow Prime Disabled
Runner v2 Enabled
Dataflow Shuffle Enabled
Resource metrics
17:05
22-04-2024

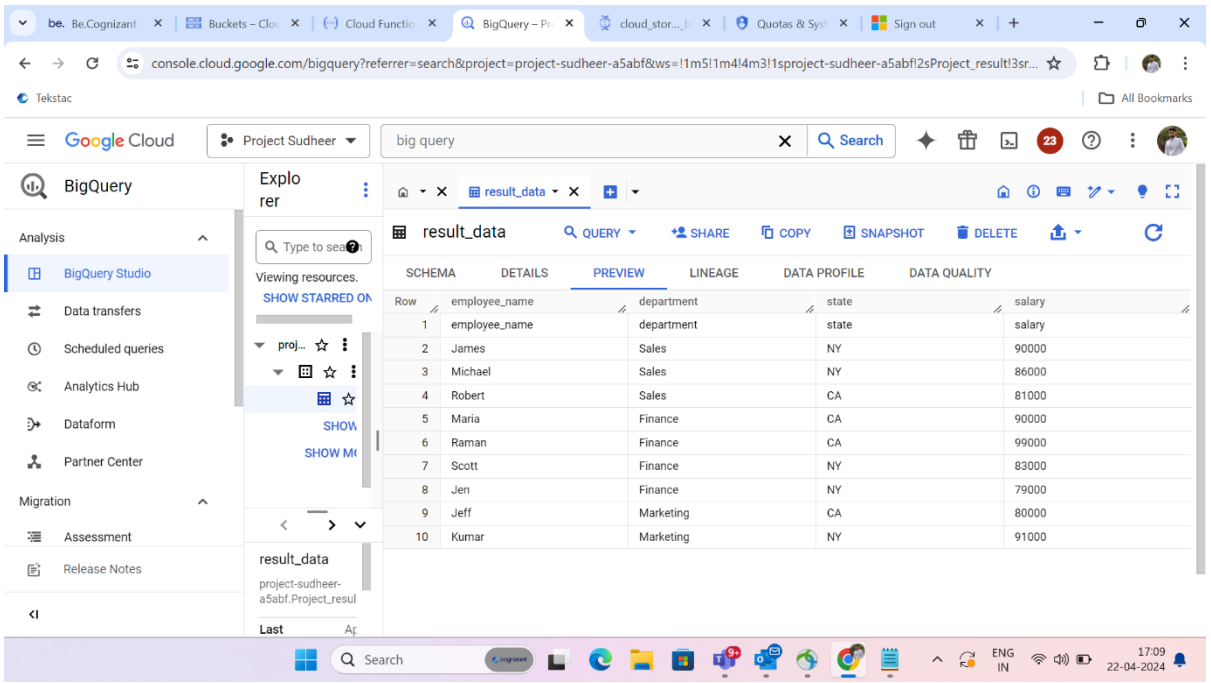

console.cloud.google.com/dataflow/jobs/us-east1/2024-04-22_04_34_24-10158601627659603172;graphView=0?project=project-sudheer-a5abf&page...
Google Cloud
Project Sudheer
dataflow
Search
cloud_storage...
CLONE
STOP
IMPORT AS PIPELINE
SHARE
SEND FEEDB
JOB GRAPH
EXECUTION DETAILS
JOB METRICS
COST
RECOMMENDATIONS
Job steps view
Table view
CLEAR SELECTION
Filter
Filter steps
Step name
Status
Wall time
Stages
Input
Read from source
Running
0 seconds
F144
JavascriptTextTransformer.TransformTextViaJavascript
Starting...
0 seconds
F144
BigQueryConverters.JsonToTableRow
Starting...
0 seconds
F144
Insert into Bigquery
Running
0 seconds
F125
Logs
SHOW
Job info
Job name
cloud_storage_to_bigquery
Job ID
2024-04-22_04_34_24-10158601627659603172
Job type
Batch
Job status
Running
SDK version
Apache Beam SDK for Java 2.54.0
A newer version of the SDK family exists and updating is recommended.
Learn more
Job region
us-east1
Worker location
us-east1
Current workers
1
Latest worker status
Worker pool started.
Start time
April 22, 2024 at 5:04:25 PM GMT+5
Elapsed time
3 min 16 sec
17:07
22-04-2024


console.cloud.google.com/dataflow/jobs/us-east1/2024-04-22_04_34_24-10158601627659603172;graphView=0?project=project-sudheer-a5abf&page...
Google Cloud
Project Sudheer
dataflow
Search
cloud_storage...
CLONE
STOP
IMPORT AS PIPELINE
SHARE
SEND FEEDB
JOB GRAPH
EXECUTION DETAILS
JOB METRICS
COST
RECOMMENDATIONS
Job steps view
Graph view
CLEAR SELECTION
Read from source
Running
1 of 2 stages succeeded
JavascriptT...Javascript
Running
0 of 1 stage succeeded
Logs
SHOW
Job info
Job name
cloud_storage_to_bigquery
Job ID
2024-04-22_04_34_24-10158601627659603172
Job type
Batch
Job status
Running
SDK version
Apache Beam SDK for Java 2.54.0
A newer version of the SDK family exists and updating is recommended.
Learn more
Job region
us-east1
Worker location
us-east1
Current workers
1
Latest worker status
Worker pool started.
Start time
April 22, 2024 at 5:04:25 PM GMT+5
Elapsed time
3 min 50 sec
17:08
22-04-2024


console.cloud.google.com/dataflow/jobs/us-east1/2024-04-22_04_34_24-10158601627659603172;graphView=0?project=project-sudheer-a5abf&page...
Google Cloud
Project Sudheer
dataflow
Search
cloud_storage...
CLONE
STOP
IMPORT AS PIPELINE
SHARE
SEND FEEDB
JOB GRAPH
EXECUTION DETAILS
JOB METRICS
COST
RECOMMENDATIONS
Job steps view
Graph view
CLEAR SELECTION
BigQueryCon...ToTableRow
Succeeded
0 sec
1 of 1 stage succeeded
Insert into Bigquery
Running
1 sec
15 of 20 stages succeeded
Logs
SHOW
Job info
Job name
cloud_storage_to_bigquery
Job ID
2024-04-22_04_34_24-10158601627659603172
Job type
Batch
Job status
Running
SDK version
Apache Beam SDK for Java 2.54.0
A newer version of the SDK family exists and updating is recommended.
Learn more
Job region
us-east1
Worker location
us-east1
Current workers
1
Latest worker status
Worker pool started.
Start time
April 22, 2024 at 5:04:25 PM GMT+5
Elapsed time
3 min 56 sec


console.cloud.google.com/dataflow/jobs/us-east1/2024-04-22_04_34_24-10158601627659603172;graphView=0?project=project-sudheer-a5abf&page...
Google Cloud
Project Sudheer
dataflow
Search
cloud_storage...
CLONE
STOP
IMPORT AS PIPELINE
SHARE
SEND FEEDB
JOB GRAPH
EXECUTION DETAILS
JOB METRICS
COST
RECOMMENDATIONS
Job steps view
Graph view
CLEAR SELECTION
BigQueryCon...ToTableRow
Succeeded
0 sec
1 of 1 stage succeeded
Insert into Bigquery
Succeeded
1 sec
20 of 20 stages succeeded
Logs
SHOW
Job info
Job name
cloud_storage_to_bigquery
Job ID
2024-04-22_04_34_24-10158601627659603172
Job type
Batch
Job status
Running
SDK version
Apache Beam SDK for Java 2.54.0
A newer version of the SDK family exists and updating is recommended.
Learn more
Job region
us-east1
Worker location
us-east1
Current workers
–
Latest worker status
Stopping worker pool.
Start time
April 22, 2024 at 5:04:25 PM GMT+5
Elapsed time
4 min 2 sec


console.cloud.google.com/bigquery?referrer=search&project=project-sudheer-a5abf&ws=!1m5!1m4!4m3!1sproject-sudheer-a5abf!2sProject_result!3sr...
Google Cloud
Project Sudheer
big query
Search
BigQuery
Explorer
+ ADD
Analysis
BigQuery Studio
Data transfers
Scheduled queries
Analytics Hub
Dataform
Partner Center
Migration
Assessment
Release Notes
Type to search
Viewing resources.
SHOW STARRED ONLY
project-sudheer-a5abf
Project_result
result_data
SHOW MORE
SHOW MORE
SUMMARY
result_data
project-sudheer-a5abf.Project_result
Last modified
Apr 22, 2024, 5:08:21 PM UTC+5:30
result_data
QUERY
SHARE
COPY
SCHEMA
DETAILS
PREVIEW
LINEAGE
DATA PROFILE
DATA QUALITY
Filter
Enter property name or value
Field name
Type
Mode
Key
Collation
Default Value
Policy Ta
employee_name
STRING
NULLABLE
department
STRING
NULLABLE
state
STRING
NULLABLE
salary
STRING
NULLABLE
age
STRING
NULLABLE
bonus
STRING
NULLABLE
EDIT SCHEMA
VIEW ROW ACCESS POLICIES
Search
ENG
IN
17:08
22-04-2024


console.cloud.google.com/bigquery?referrer=search&project=project-sudheer-a5abf&ws=!1m5!1m4!4m3!1sproject-sudheer-a5abf!2sProject_result!3sr...
Google Cloud
Project Sudheer
big query
Search
BigQuery
Explo
rer
Analysis
BigQuery Studio
Data transfers
Scheduled queries
Analytics Hub
Dataform
Partner Center
Migration
Assessment
Release Notes
Viewing resources.
result_data
project-sudheer-a5abf.Project_resul
result_data
QUERY
SHARE
COPY
SNAPSHOT
DELETE
SCHEMA
DETAILS
PREVIEW
LINEAGE
DATA PROFILE
DATA QUALITY
Row
employee_name
department
state
salary
1 employee_name department state salary
2 James Sales NY 90000
3 Michael Sales NY 86000
4 Robert Sales CA 81000
5 Maria Finance CA 90000
6 Raman Finance CA 99000
7 Scott Finance NY 83000
8 Jen Finance NY 79000
9 Jeff Marketing CA 80000
10 Kumar Marketing NY 91000
Search
ENG
IN
17:09
22-04-2024