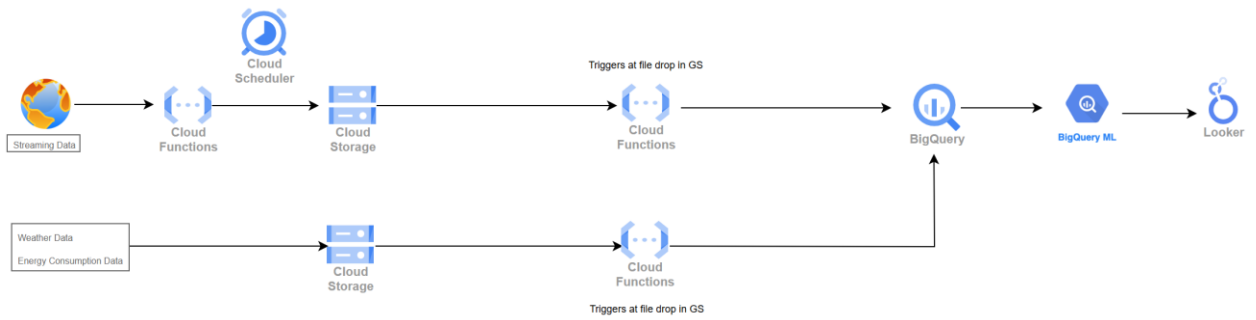


Pipeline:



1) Static monthly weather archival data loaded into cloud storage

capstone824

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							
Buckets > capstone824							
UPLOAD FILES UPLOAD FOLDER CREATE FOLDER TRANSFER DATA MANAGE HOLDS EDIT RETENTION							
DOWNLOAD DELETE							
Filter by name prefix only Filter Filter objects and folders Show Live objects only							
<input type="checkbox"/>	Name	Size	Type	Created	Storage class	Last m	
<input type="checkbox"/>	Static_MonthlyWeather_TrentonF...	24.3 KB	text/csv	Aug 12, 2024, 8:15:49 PM	Standard	Aug 1	Download More

2) Cloud function for a one time load of data from cloud storage to Big Query Table:

Cloud Functions

Function details

EDIT

DELETE

COPY

LEARN

K

archiveddata

2nd gen

(Deployed at Aug 12, 2024, 7:00:23 PM)

Powered by Cloud Run

archiveddata

URL: https://us-central1-probable-hash-285404.cloudfunctions.net/archiveddata

METRICS

DETAILS

SOURCE

VARIABLES

TRIGGER

PERMISSIONS

LOGS

TESTING

Runtime: Python 3.12

Entry point: archiveddata

EDIT

DOWNLOAD ZIP

main.py

requirements.txt

```

1 from google.cloud import bigquery
2 from google.cloud import storage
3 import pandas as pd
4 from io import StringIO
5
6 # Replace with your actual project ID
7 PROJECT_ID = 'probable-hash-285404'
8 DATASET_ID = 'Capstone824'
9 TABLE_ID = 'static_weather_trentonfalls_ny'
10
11 def archiveddata(event, context):
12     """Triggered by a change to a Cloud Storage bucket.
13     Args:
14         event (dict): The Cloud Functions event payload.
15         context (google.cloud.functions.Context): Metadata of trigger
16
17
18     # Get the file that triggered the function
19     file_name = event['name']
20     bucket_name = event['bucket']
21
22     # Instantiate clients
23     storage_client = storage.Client()
24     bucket = storage_client.bucket(bucket_name)
25     blob = bucket.blob(file_name)
26     blob_string = blob.download_as_text()
27
28     # Set chunk size (number of rows per chunk)
29     chunk_size = 10000
30
31     # Define the schema once for reuse
32     schema = [

```

3) Data loaded into Big Query table

Explorer

+ ADD

K

Search BigQuery resources

Viewing resources.

SHOW STARRED ONLY

Data canvases

External connections

Capstone824

energyconsumption...

static_energyconsu...

static_energyconsu...

static_joined

static_weather_trent...

static_weather_trent...

static_weather_trent...

bts

opensky

static_monthlyweather_t...

static_weathe...

QUERY

SHARE

COPY

SNAPSHOT

DELETE

EXPORT

SCHEMA

DETAILS

PREVIEW

TABLE EXPLORER

PREVIEW

INSIGHTS

PREVIEW

LINEAGE

DA

Row	STATION	NAME	LATITUDE	LONGITUDE	ELEVATION	DATE
1	USC00308578	TRENTON FALLS, NY US	43.2761	-75.1566	243.8	2016-01
2	USC00308578	TRENTON FALLS, NY US	43.2761	-75.1566	243.8	2016-02
3	USC00308578	TRENTON FALLS, NY US	43.2761	-75.1566	243.8	2016-03
4	USC00308578	TRENTON FALLS, NY US	43.2761	-75.1566	243.8	2016-04
5	USC00308578	TRENTON FALLS, NY US	43.2761	-75.1566	243.8	2016-05
6	USC00308578	TRENTON FALLS, NY US	43.2761	-75.1566	243.8	2016-06
7	USC00308578	TRENTON FALLS, NY US	43.2761	-75.1566	243.8	2016-07
8	USC00308578	TRENTON FALLS, NY US	43.2761	-75.1566	243.8	2016-08
9	USC00308578	TRENTON FALLS, NY US	43.2761	-75.1566	243.8	2016-09
10	USC00308578	TRENTON FALLS, NY US	43.2761	-75.1566	243.8	2016-10
11	USC00308578	TRENTON FALLS, NY US	43.2761	-75.1566	243.8	2016-11
12	USC00308578	TRENTON FALLS, NY US	43.2761	-75.1566	243.8	2016-12
13	USC00308578	TRENTON FALLS, NY US	43.2761	-75.1566	243.8	2017-01
14	USC00308578	TRENTON FALLS, NY US	43.2761	-75.1566	243.8	2017-02
15	USC00308578	TRENTON FALLS, NY US	43.2761	-75.1566	243.8	2017-03

4) Log extract of cloud function trigger from cloud storage file upload

Cloud Functions

Function details

archiveddata 2nd gen (Deployed at Aug 12, 2024, 7:00:23 PM)

URL: https://us-central1-probable-hash-285404.cloudfunctions.net/archiveddata

Powered by Cloud Run

METRICS DETAILS SOURCE VARIABLES TRIGGER PERMISSIONS LOGS TESTING

Logs Severity: Default Filter Search all fields and values

SEVERITY	TIMESTAMP	SUMMARY
> *	2024-08-12 18:51:36.760 PDT	Error converting Pandas column with name: "CSD_ATTRIBUTES" and datatype: "int64" to an appropriat...
> *	2024-08-12 18:51:36.777 PDT	Exception on / [POST] Traceback (most recent call last): File "/layers/google.python.pip/pip/lib...
> i	2024-08-12 18:59:34.681 PDT	namespaces/probable-hash-285404/servi... service-757496973353@gcf-admin-robot... (@type: type...
> i	2024-08-12 18:59:35.232 PDT	Cloud Functions UpdateFunction us-central1:archiveddata appalbadi@gmail.com (@type: type.googl...
> i	2024-08-12 19:00:13.966 PDT	namespaces/probable-hash-285404/servi... service-757496973353@gcf-admin-robot... (@type: type...
> i	2024-08-12 19:00:22.032 PDT	Default STARTUP TCP probe succeeded after 1 attempt for container "worker" on port 8080.
> i	2024-08-12 19:00:22.107 PDT	Cloud Run v1 archiveddata-00017-zuh (@type: type.googleapis.com/google.cloud.audit.AuditLog, ...
> i	2024-08-12 19:00:23.501 PDT	Cloud Run v1 archiveddata (@type: type.googleapis.com/google.cloud.audit.AuditLog, methodName...
> i	2024-08-12 19:00:23.846 PDT	Cloud Functions UpdateFunction us-central1:archiveddata appalbadi@gmail.com (@type: type.googl...
> i	2024-08-12 19:00:41.760 PDT	Default STARTUP TCP probe succeeded after 1 attempt for container "worker" on port 8080.
> i	2024-08-12 19:00:41.806 PDT	POST 200 130 B 3 s APIs-Google; (+https://developers.goo... https://archiveddata-vdugnrzaq-uc...
> *	2024-08-12 19:00:44.848 PDT	Successfully loaded data into Capstone824:static_weather_trentonfalls.ny from file Static_MonthlyW...
> i	2024-08-12 19:04:55.872 PDT	POST 200 331 B 683 ms APIs-Google; (+https://developers.goo... https://archiveddata-vdugnrzaq...
> *	2024-08-12 19:04:56.572 PDT	Exception on / [POST] Traceback (most recent call last): File "/layers/google.python.pip/pip/lib...
> i	2024-08-12 20:15:54.510 PDT	Default STARTUP TCP probe succeeded after 1 attempt for container "worker" on port 8080.
> *	2024-08-12 20:15:58.380 PDT	Successfully loaded data into Capstone824:static_weather_trentonfalls.ny from file Static_MonthlyW...

1) Energy archival data loaded into cloud storage

capstone824

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

Buckets > capstone824

UPLOAD FILES UPLOAD FOLDER CREATE FOLDER TRANSFER DATA DOWNLOAD DELETE

Filter by name prefix only Filter objects and folders Show Live objects only

	Name	Size	Type	Created	Storage class	Last modified
<input type="checkbox"/>	Static_MonthlyWeather_TrentonF...	24.3 KB	text/csv	Aug 12, 2024, 8:15:49 PM	Standard	Aug 12, 2024
<input type="checkbox"/>	Utility_Energy_Registry_Monthly...	1.6 KB	text/csv	Aug 12, 2024, 9:01:19 PM	Standard	Aug 12, 2024
<input type="checkbox"/>	weather_data_schema.json	5.9 KB	application/json	Aug 12, 2024, 12:27:55 PM	Standard	Aug 12, 2024

2) Cloud function for a one time load of data from cloud storage to Big Query Table:

✓ **weatherdata** 2nd gen (Deployed at Aug 12, 2024, 9:06:31 PM)

Powered by Cloud Run 
[weatherdata](#)

URL: <https://us-central1-probable-hash-285404.cloudfunctions.net/weatherdata>  

METRICS DETAILS **SOURCE** VARIABLES TRIGGER PERMISSIONS LOGS TESTING

Runtime : Python 3.12 Entry point : weatherdata

EDIT

DOWNLOAD ZIP

main.py
requirements.txt

```
1 from google.cloud import bigquery
2 from google.cloud import storage
3 import pandas as pd
4 from io import StringIO
5
6 # Replace with your actual project ID
7 PROJECT_ID = 'probable-hash-285404'
8 DATASET_ID = 'Capstone824'
9 TABLE_ID = 'static_energyconsumptionbyzip_NY'
10
11 def weatherdata(event, context):
12     """Triggered by a change to a Cloud Storage bucket.
13     Args:
14         event (dict): The Cloud Functions event payload.
15         context (google.cloud.functions.Context): Metadata of triggeri
16     """
17
18     # Get the file that triggered the function
19     file_name = event['name']
20     bucket_name = event['bucket']
21
22     # Instantiate clients
23     storage_client = storage.Client()
24     bucket = storage_client.bucket(bucket_name)
25     blob = bucket.blob(file_name)
26     blob_string = blob.download_as_text()
27
28     # Set chunk size (number of rows per chunk)
29     chunk_size = 10000
30
31     # Define the schema once for reuse
32     schema = [
```

3) Cloud function trigger log:

weatherdata 2nd gen (Deployed at Aug 12, 2024, 9:06:31 PM) Powered by Cloud Run

URL: <https://us-central1-probable-hash-285404.cloudfunctions.net/weatherdata>

METRICS DETAILS SOURCE VARIABLES TRIGGER PERMISSIONS **LOGS** TESTING

Logs Severity: Default Filter Search all fields and values

SEVERITY	TIMESTAMP	SUMMARY
> i	2024-08-12 21:00:59.605 PDT	Cloud Run v1 weatherdata-00004-nec {@type: type.googleapis.com/google.cloud.audit.AuditLog, m...
> i	2024-08-12 21:01:01.004 PDT	Cloud Run v1 weatherdata {@type: type.googleapis.com/google.cloud.audit.AuditLog, methodName:...
> i	2024-08-12 21:01:01.505 PDT	Cloud Functions UpdateFunction us-central1:weatherdata appalbadi@gmail.com {@type: type.google...
> i	2024-08-12 21:01:20.683 PDT	POST 200 331 B 288 ms APIs-Google; (+https://developers.goo... https://weatherdata-vdukgnrzaq-...
> *	2024-08-12 21:01:20.988 PDT	Exception on / [POST] Traceback (most recent call last): File "/layers/google.python.pip/pip/lib...
> i	2024-08-12 21:05:38.088 PDT	namespaces/probable-hash-285404/servi... service-757496973353@gcf-admin-robot... {@type: type...
> i	2024-08-12 21:05:38.730 PDT	Cloud Functions UpdateFunction us-central1:weatherdata appalbadi@gmail.com {@type: type.google...
> i	2024-08-12 21:06:20.024 PDT	namespaces/probable-hash-285404/servi... service-757496973353@gcf-admin-robot... {@type: type...
> i	2024-08-12 21:06:29.196 PDT	Default STARTUP TCP probe succeeded after 1 attempt for container "worker" on port 8080.
> i	2024-08-12 21:06:29.269 PDT	Cloud Run v1 weatherdata-00005-yit {@type: type.googleapis.com/google.cloud.audit.AuditLog, m...
> i	2024-08-12 21:06:30.658 PDT	Cloud Run v1 weatherdata {@type: type.googleapis.com/google.cloud.audit.AuditLog, methodName:...
> i	2024-08-12 21:06:31.077 PDT	Cloud Functions UpdateFunction us-central1:weatherdata appalbadi@gmail.com {@type: type.google...
> i	2024-08-12 21:06:59.011 PDT	POST 200 130 B 3.9 s APIs-Google; (+https://developers.goo... https://weatherdata-vdukgnrzaq-u...
> *	2024-08-12 21:07:02.945 PDT	Successfully loaded data into Capstone824:static_energyconsumptionbyzip_NY from file Utility_Energ...

4) Big Query table loaded through cloud function trigger

Explorer + ADD

static_en..._NY

static_energycons...

SCHEMA DETAILS **PREVIEW** TABLE EXPLORER **PREVIEW** INSIGHTS **PREVIEW** LINEAGE DATA PROFIL

Row	year	data_class	data_field_display_name	data_field	zip_city
1	2020	electricity	Business Consumption (SC+O)	2_nat_consumption	Caroga Lake
2	2020	electricity	Business Consumption (SC+O)	2_nat_consumption	Caroga Lake
3	2020	electricity	ICAP Capacity Tag (R)	1_ny_policy_support	Caroga Lake
4	2020	electricity	CCA-Ineligible Customer Count...	8_ny_policy_support	Caroga Lake
5	2020	electricity	ICAP Capacity Tag (R)	1_ny_policy_support	Caroga Lake
6	2020	electricity	ICAP Capacity Tag (T)	6_ny_policy_support	Caroga Lake
7	2020	electricity	ICAP Capacity Tag (T)	6_ny_policy_support	Caroga Lake
8	2020	electricity	ICAP Capacity Tag (SC+O)	5_ny_policy_support	Caroga Lake
9	2020	electricity	CCA-Ineligible Customer Count...	8_ny_policy_support	Caroga Lake
10	2020	electricity	ICAP Capacity Tag (T)	6_ny_policy_support	Glenfield
11	2020	electricity	ICAP Capacity Tag (R)	1_ny_policy_support	Glenfield
12	2020	electricity	CCA-Ineligible Customer Count...	7_ny_policy_support	Piffard
13	2020	electricity	ICAP Capacity Tag (SC)	2_ny_policy_support	Piffard
14	2020	electricity	CCA-Ineligible Customer Count...	8_ny_policy_support	Piffard
15	2020	electricity	CCA-Ineligible Customer Count...	8_ny_policy_support	Piffard
16	2020	electricity	Total Consumption (T)	3_nat_consumption	Piffard
17	2020	natural_gas	Business Consumption (SC+O)	5_nat_consumption	Castleton on Hudson
18	2020	electricity	Total Consumption (T)	3_nat_consumption	Castleton on Hudson
19	2020	electricity	ICAP Capacity Tag (R)	1_ny_policy_support	Castleton on Hudson
20	2020	electricity	CCA-Ineligible Customer Count...	7_ny_policy_support	Castleton on Hudson

static_energyconsumptionbyzip_NY
probable-hash-285404.Capstone824

Last modified Aug 12, 2024, 9:10:19 PM UTC-7

1) Streaming data load cloud function to call APIs and load the data into cloud storage:

Cloud Functions ← Function details [EDIT](#) [DELETE](#) [COPY](#)

✓ **streamdata** 2nd gen (Deployed at Aug 13, 2024, 11:34:08 AM) URL: <https://us-central1-probable-hash-285404.cloudfunctions.net/streamdata> [🔗](#) [?](#)

METRICS DETAILS **SOURCE** VARIABLES TRIGGER PERMISSIONS LOGS TESTING

Runtime: Python 3.12 Entry point: streamdata [EDIT](#)

main.py
requirements.txt

```

1  import functions_framework
2  import requests
3  import pandas as pd
4  from google.cloud import storage
5  from datetime import datetime
6  import os
7
8  # GCP Cloud Function entry point
9
10 def streamdata(request):
11     # Extract api_keys and areas from the request (query string parameters)
12     api_keys = ['7c8218f18550417496b43123242902', '7c8218f18550417496b43123242902', '7c8218f18550417496b43123242902', '7c8218f18550417496b43123242902']
13     areas = ['13354', '13304', '13352', '13435', '13438', '13469', '13438']
14
15     if not api_keys or not areas:
16         return "api_key and area parameters are required", 400
17
18     if len(api_keys) != len(areas):
19         return "The number of api_key and area values must be the same", 400
20
21     # Initialize Cloud Storage client
22     storage_client = storage.Client()
23     bucket_name = "capstone_streaming" # Replace with your bucket name
24     bucket = storage_client.bucket(bucket_name)
25
26     for api_key, area in zip(api_keys, areas):
27         # WeatherAPI URI
28         url = f"https://api.weatherapi.com/v1/current.json?key={api_key}&q={area}"
29
30         # Make the API request
31         response = requests.get(url)
32         if response.status_code != 200:

```

2) Cloud function to load data from cloud storage to BigQuery on the cloud storage trigger

Cloud Functions ← Function details [EDIT](#) [DELETE](#) [COPY](#)

✓ **streamdataload** 2nd gen (Deployed at Aug 13, 2024, 12:28:41 PM) URL: <https://us-central1-probable-hash-285404.cloudfunctions.net/streamdataload> [🔗](#) [?](#)

METRICS DETAILS **SOURCE** VARIABLES TRIGGER PERMISSIONS LOGS TESTING

Runtime: Python 3.12 Entry point: streamdataload [EDIT](#)

main.py
requirements.txt

```

1  from google.cloud import bigquery
2  from google.cloud import storage
3  import pandas as pd
4  from io import StringIO
5
6  # Replace with your actual project ID
7  PROJECT_ID = 'probable-hash-285404'
8  DATASET_ID = 'Capstone824'
9  TABLE_ID = 'streaming_data'
10
11 def streamdataload(event, context):
12     """Triggered by a change to a Cloud Storage bucket.
13     Args:
14         event (dict): The Cloud Functions event payload.
15         context (google.cloud.functions.Context): Metadata of triggering event.
16     """
17
18     # Get the file that triggered the function
19     file_name = event['name']
20     bucket_name = event['bucket']
21
22     # Instantiate clients
23     storage_client = storage.Client()
24     bucket = storage_client.bucket(bucket_name)
25     blob = bucket.blob(file_name)
26     blob_string = blob.download_as_text()
27
28     # Set chunk size (number of rows per chunk)
29     chunk_size = 10000
30
31     # Define the schema once for reuse
32     schema = [

```

3) Scheduler to run the API call function every 5 minutes

