Operational Excellence & Security

Cost Controls

To ensure this pipeline remains sustainable and within the Google Cloud Free Tier (or a minimal budget), the following controls are enforced:

- **Cloud Scheduler Frequency:** The trigger is strictly scheduled to minimize Cloud Function invocations and API calls
- **BigQuery Optimization:**
    - **Partitioning:** Tables are partitioned by timestamp. Queries only scan relevant date partitions rather than the full historical dataset.
    - **Bytes Billed Limit:** A "Maximum bytes billed" cap is set on queries to prevent accidental over-spending during ad-hoc analysis.
- **API Limits:** The request volume is calculated to stay under the Weatherstack Free Tier cap (1,000 calls/month), ensuring zero external data costs.

Failure Handling: Rate Limiting

The system includes specific logic to handle HTTP 429 (Too Many Requests) errors from the external API:

- **Detection:** The Cloud Function inspects the status_code of every response.
- **Fail Fast Strategy:** If a 429 is detected, the function logs a "Warning" severity alert and terminates immediately without retrying. This prevents infinite retry loops that would waste compute resources and further block the API key.
- **Monitoring:** Log-based metrics track these warnings to alert the team if the external quota is consistently exceeded.

Security: Secret Manager

Credential security is prioritized to prevent leakage in version control systems:

- **Storage:** The Weatherstack API Key is stored securely in **Google Secret Manager**, not in the source code.
- **IAM Access:** Only the specific Service Account attached to the Cloud Function is granted "Secret Manager Secret Accessor" permissions.
- **Runtime Injection:** The key is injected into the function's environment variables only at runtime.

---

3. Reproducibility Guide

Use the following Google Cloud CLI (gcloud) commands to rapidly deploy or decommission the infrastructure.

Spin Up (Deployment)

Prerequisites: Authenticated gcloud CLI and a valid GCP Project.

1. Set Environment Variables

```
export PROJECT_ID="your-project-id" export REGION="us-central1" gcloud config set project $PROJECT_ID
```

2. Create Infrastructure

Create Pub/Sub Topic

```
gcloud pubsub topics create weather-data-topic
```

Create BigQuery Dataset

```
bq --location=US mk -d weather_data
```

3. Deploy Cloud Function

```
gcloud functions deploy fetch_weather

--runtime python310

--trigger-topic weather-data-topic

--entry-point main

--region $REGION

--set-secrets 'WEATHER_API_KEY=projects/$PROJECT_ID/secrets/weather-api-key/versions/latest'
```

4. Start Scheduler

```
gcloud scheduler jobs create pubsub hourly-weather-job
```

```
--schedule "0 * * * *"

--topic weather-data-topic

--message-body "Trigger"

--location $REGION
```

Spin Down (Teardown)

Execute these commands to stop all billing immediately.

1. Delete Scheduler Job

```
gcloud scheduler jobs delete hourly-weather-job --location $REGION
```

2. Delete Cloud Function

```
gcloud functions delete fetch_weather --region $REGION
```

3. Delete Pub/Sub Topic

```
gcloud pubsub topics delete weather-data-topic
```

4. Remove BigQuery Dataset *(Warning: This permanently deletes all stored data)*

```
bq rm -r -f -d weather_data
```