Operations Runbook — Real-Time Weather Pipeline

Purpose

This runbook documents how to operate, monitor, validate, and recover the real-time weather streaming pipeline on Google Cloud Platform.

Pipeline:
Public Weather API → Cloud Function (2nd Gen) → Pub/Sub → Dataflow (Streaming) → BigQuery → BQML → Looker Studio

---

Monitoring

Dataflow
- Console path: Dataflow → Jobs
- Expected state: RUNNING
- Check Job Graph for stalled steps or errors
- Monitor worker count and throughput

Pub/Sub
- Console path: Pub/Sub → Subscriptions
- Backlog should remain near zero
- A growing backlog indicates downstream issues

BigQuery
- Validate streaming inserts by checking latest timestamps
- Ensure tables continue to append new rows

APIs & Metrics

- Console path: APIs & Services → Metrics
- Confirm activity for:
  - Pub/Sub API
  - Dataflow API
  - BigQuery API

Failure Recovery

Dataflow Job Failure
1. Stop the failed job
2. Relaunch the streaming template with the same parameters
3. Verify rows resume appearing in BigQuery

Cloud Function Failure
1. Check Cloud Logging for errors
2. Redeploy the function if needed
3. Confirm Pub/Sub receives new messages

Pub/Sub Backlog Growth
- Indicates Dataflow is stalled or stopped
- Restart Dataflow job if required

Data Validation

Freshness Check
- Query BigQuery for most recent timestamp
- Expected: new rows every ingestion interval

Volume Check
- Validate consistent row counts per minute
- Watch for sudden drops or spikes

Schema Check
- Ensure numeric fields remain valid
- Temperature, humidity, pressure, and wind fields should not be null

---

Cost Awareness
- Dataflow streaming jobs are the primary cost driver
- Stop the job when not required for demos or grading
- Monitor BigQuery streaming usage

Security & Access
- No secrets stored in code
- All access controlled via IAM
- Service accounts follow least-privilege principles
- Data is public and non-sensitive

Ownership
- Owners: Project Team
- Platform: Google Cloud Platform
- Data Source: Public Weather API

Notes
This runbook supports operational continuity and satisfies the project requirement for infrastructure operations and reliability documentation.