Comparison: Event-Driven STS DAGs for GCS to GCS Transfer

Use Case: GDW to APMF File Transfer with STS Triggering

This document compares two DAG implementations for Google Cloud Storage Transfer Service (STS) within a restricted GCP environment:

1. Basic Event-Driven Polling DAG
2. Deduplicated Job Management DAG (with job reuse)

Both DAGs avoid Pub/Sub and Cloud Functions, adhering to organizational constraints.

---

1. Basic Event-Driven Polling DAG

- Creates a new STS job on every run.
- Uses prefix filtering (e.g., 'incoming/').
- Uses file modified timestamp (last 10 min).
- Triggers STS run if new file is found.
- XCom used to communicate trigger signal.
- Risks: Creates duplicate STS jobs each time.

Recommended: Not ideal for production.

---

2. Deduplicated Create & Trigger DAG (Recommended)

- Reuses an existing STS job based on description.
- Creates a new job only if none exists.
- Uses prefix filter (e.g., 'incoming/').
- Uses file modified timestamp (last 10 min).
- Triggers STS run only if needed.
- Avoids duplicate STS jobs in console.

Recommended: Ideal for stable, production-safe environments.

Key Configuration Shared by Both

- Polling interval: every 10 minutes (can change to 5 mins).
- Prefix used: 'incoming/'.
- File updated time window: last 10 minutes.
- IAM roles: No admin, uses storagetransfer.admin, storage.objectViewer, storage.objectCreator.