# Infra Instructions: Event-Driven STS via Composer

Overview:

This document describes how to set up an event-driven GCS-to-GCS file transfer using Storage Transfer Service (STS), Pub/Sub, and Composer (Airflow). The transfer is initiated when a new file is uploaded to a bucket in Project 1, and files are transferred to a bucket in Project 2.

Architecture:

- GCS Bucket in Project 1 sends OBJECT_FINALIZE event to Pub/Sub
- Cloud Function (optional) filters and triggers Airflow DAG
- Composer DAG creates or runs an STS job to transfer files to Project 2

Required GCP Services (enable in both projects as needed):

- storage.googleapis.com (Cloud Storage API)
- storagetransfer.googleapis.com (STS API)
- pubsub.googleapis.com (Pub/Sub for event notifications)
- cloudfunctions.googleapis.com (if using Cloud Function)
- composer.googleapis.com (already used for Airflow)

IAM Permissions:

1. Composer Service Account (Project 1):

- roles/storagetransfer.user (to create and manage STS jobs)
- roles/storage.objectViewer (on source bucket)
- roles/storage.objectCreator OR storage.objectAdmin (on destination bucket)
- roles/composer.worker (to run DAGs)
- roles/pubsub.publisher (if Pub/Sub used to trigger Composer)

2. STS Transfer Service Account:

Format: service-<project-number>@gcp-sa-storagetransfer.iam.gserviceaccount.com

- roles/storage.objectViewer (on source bucket)

- roles/storage.objectCreator OR storage.objectAdmin (on destination bucket)

Must be manually added using:

```
gsutil iam ch serviceAccount:<STS_SA>:roles/storage.objectViewer gs://source-bucket
gsutil iam ch serviceAccount:<STS_SA>:roles/storage.objectCreator gs://destination-bucket
```

3. User or Triggering Service Account:

- roles/storagetransfer.user or storagetransfer.admin (to trigger or create STS jobs)

- Optional: storage roles on buckets (if not handled by Composer SA or STS SA)

Pub/Sub Setup:

- Create a topic in Project 1

- Configure GCS bucket to publish OBJECT_FINALIZE events:

```
gsutil notification create -t my-topic -f json -e OBJECT_FINALIZE gs://bucket-1
```

Cloud Function (Optional):

- Subscribes to Pub/Sub topic

- Validates event metadata

- Calls Composer DAG via API or workflow

STS Job:

- Can be pre-created (recommended for reuse)

- Composer DAG uses CloudDataTransferServiceRunJobOperator or CreateJobOperator

Notes:

- Use includePrefixes and objectConditions in STS job to ensure only new files are transferred
- Ensure Composer's VPC can reach both buckets and APIs
- Use custom roles in org-restricted setups to avoid admin access

This setup is aligned with Google Cloud?s best practices for scalable, event-driven data transfer pipelines.