

Infrastructure Setup for Cross-Project STS DAGs

This document outlines what the infrastructure team must configure to enable GCS-to-GCS transfers using Storage Transfer Service (STS) with Cloud Composer, including cross-project scenarios.

1. Enable the following APIs in both source and destination projects:

- Storage Transfer Service API (storage.googleapis.com)
- Cloud Storage API (storage.googleapis.com)

2. In the Composer environment (Airflow 2.5.3, Composer 2.6.5), add the following PyPI packages:

- `apache-airflow-providers-google==10.1.1`
- `apache-airflow-providers-amazon==6.1.0`

3. IAM Permissions (Grant to the Composer service account from Project 1):

A. On the Source Bucket (Project 1):

- `storage.objectViewer`

B. On the Destination Bucket (Project 2):

- `storage.objectCreator` or `storage.objectAdmin`

C. On Project 1 (where STS job runs):

- `storagetransfer.user` or custom role with:
- `storagetransfer.transferJobs.create`
- `storagetransfer.transferJobs.get`
- `storagetransfer.transferJobs.run`

4. Network:

- Ensure that the Composer VPC can access both buckets and is not blocked from accessing

*.googleapis.com

5. Event-Driven DAG (Polling):

- A DAG runs every 15 minutes to check for new files in Project 1 bucket.
- If new files exist, it creates a one-time STS job to transfer files to the Project 2 bucket.
- Each run creates a new job using `CloudDataTransferServiceCreateJobOperator`.

This setup supports event-driven file movement without requiring Pub/Sub or Cloud Functions and is suited for restricted organization environments.