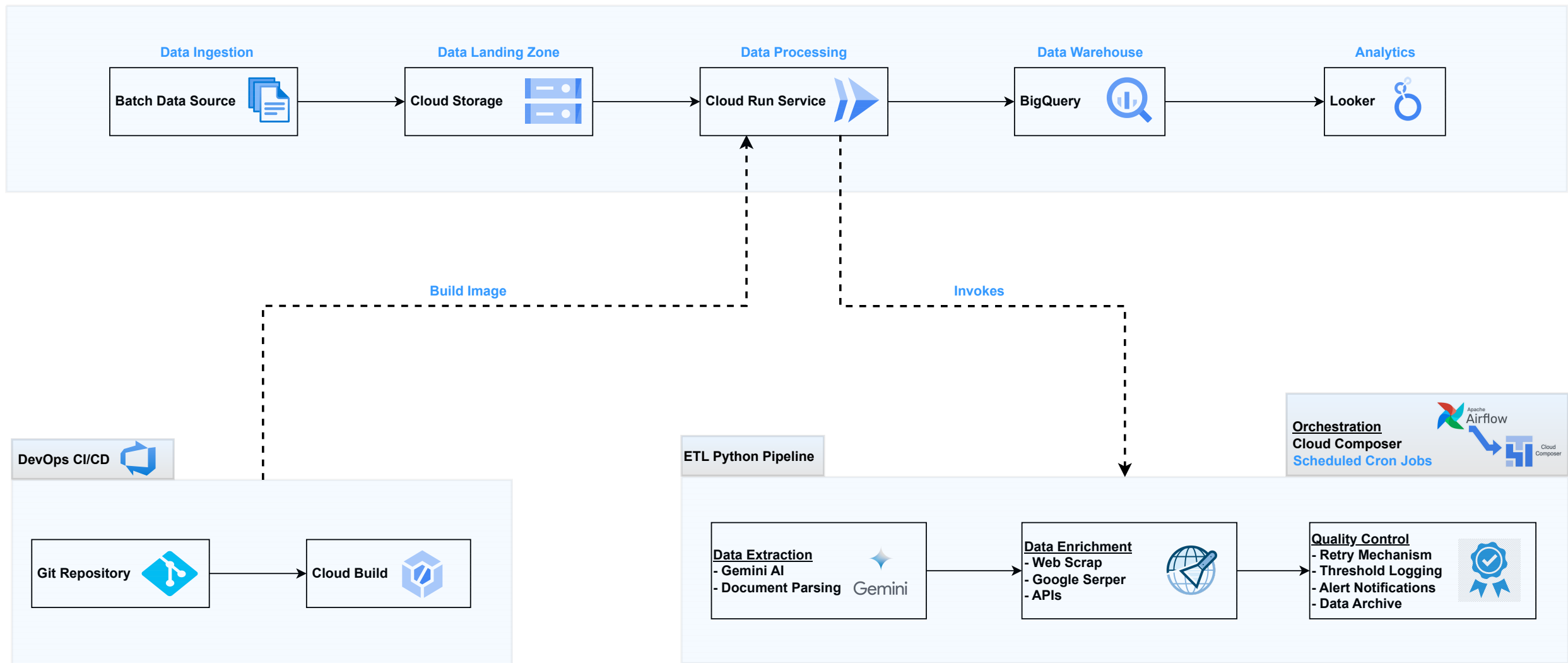


GCP Cloud Native Deployment Architecture



KEY CONSIDERATIONS

Assumptions

- Batch processing of 20-50 files per day (each < 10GB) from multiple business units (BUs) with scheduled data processing jobs.
- Non real-time and cost-optimized processing.
- Serverless and GCP-native services.

Design Overview

1. Serverless & Cost-Optimized:
 - Cloud Run:
 - Processed files only when they exist (Orchestration scheduled as required - 2x/day)
 - Scales to zero when idle (no VM costs)
 - Cheaper than Dataflow for small batches (<10GB)
 - Cloud Storage:
 - Auto-deletes archived files after 180 days from Archived folder (no manual cleanup)
2. Reliable Data Quality:
 - Failed records are logged and retried (via Cloud Logging/Alerts)
 - Threshold logging to ensure the data accuracy
 - Email alert notifications if threshold logs doesn't meet the limit
3. Compliance:
 - For least-privilege IAM, only BUs service accounts have write access to Cloud Storage folders