UseCase

By Dhandapani Yedappalli Krishnamurthi Sep 4, 2025

Problem Statement

The banking project focused on product sales analysis was initiated to solve a specific problem for the end-user: managing and focusing on a large portfolio of insurance and pension products.

Key aspects of this project include:

• The Core Problem: The end-user had thousands of insurance and pension products, which made it difficult for them to identify which ones were actually selling well. This prevented them from effectively targeting customers with the most successful products.

• The Proposed Solution: The goal was to create a Tableau dashboard. This dashboard would visualize only the products that were selling, enabling the business to concentrate its sales and marketing efforts on those specific items and target customers more effectively.

• Technical Architecture and Data Flow: The project utilized a cloud-based infrastructure, specifically Google Cloud Platform (GCP). The data flow followed a RAW, Processed, and Consumption model.

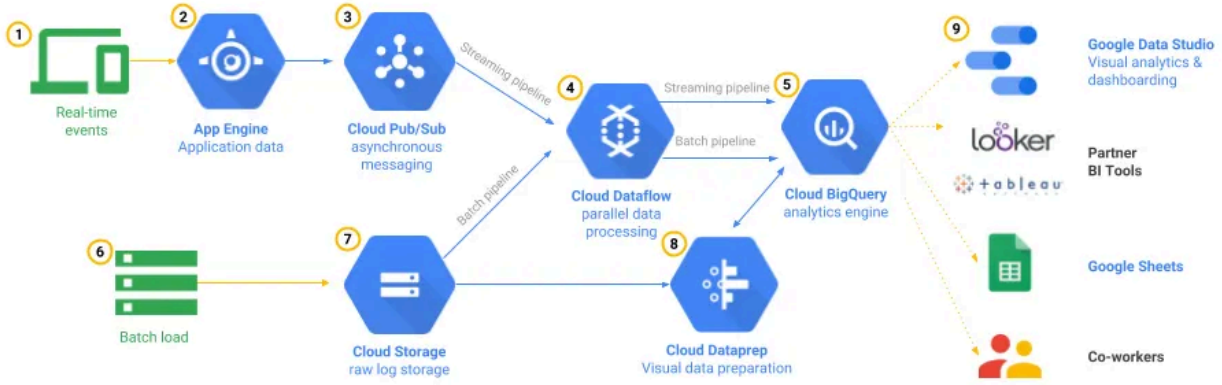
  ◦ Data Sources: Data was ingested from various file types, including CSV, TSV, and Flatfiles.

  ◦ Connectivity: Cloud Interconnect was used to connect on-premises storage with the cloud environment.

  ◦ Data Processing: The project used Cloud Composer for orchestration (likely managing Airflow workflows), Dataproc for data processing (likely running Spark jobs), and technologies like Spark, Python, and Airflow.

  ◦ Data Warehouse: Google BigQuery served as the data warehouse for the processed data.


Banking Project: Product Sales Analysis
End User Problem
1000s of Insurance/Pension Products
Unable to Target Selling Products
Need Visualization of Selling Products
Solution
Tableau Dashboard
Visualize Selling Products
Enable Customer Targeting
Data Sources
CSV
TSV
Flatfile
GCP Technologies Used
Data Processing
Cloud Composer
Dataproc
Data Warehouse
BigQuery
Connectivity
Cloud Interconnect (On-prem to Cloud)
Programming/Tools
Spark
Python
Airflow
Data Models
RAW Model
Processed Model
Consumption Model



What's in the diagram (left → right)

- **On-prem sources:** CSV/TSV/flat files + core policy/CRM systems.
- **Connectivity & ingestion: Cloud Interconnect/VPN**; **Storage Transfer Service** (and Transfer Appliance for bulk); optional **Pub/Sub** if you later add events.
- **Landing zones on GCS: RAW** buckets with versioning → **PROCESSED** (cleaned, partitioned Parquet/ORC).
- **Processing & modeling: Dataproc (Spark/PySpark)** for ETL; **BigQuery** for ELT and semantic marts; **Cloud Composer (Airflow)** orchestrates end-to-end.
- **Consumption: BigQuery** curated marts including a "Selling Products" fact; **Tableau** via BigQuery connector (ODBC/JDBC) powering dashboards that filter to selling products only.

Components

- **Governance & metadata: Dataplex + Data Catalog** for domains, tags, discovery, lineage.
- **Data quality:** Great Expectations or **CloudDQ** integrated in Composer.
- **Observability: Cloud Logging/Monitoring** with alerts and SLAs.
- **Networking & security: VPC**, Private Service Connect, **Cloud NAT**, firewall rules, **VPC Service Controls** for BQ/GCS; **IAM** least-privilege; **Secret Manager**; **CMEK (Cloud KMS)**.
- **DevOps: Cloud Build**, **Artifact Registry**, **Terraform** for IaC; unit tests and DAG tests.

Flow (brief)

On-prem files/systems → Interconnect/VPN → GCS **RAW** → Dataproc & BigQuery transforms → GCS **PROCESSED** → BigQuery **CONSUMPTION** marts → Tableau dashboards that spotlight the actual selling products for targeted campaigns.