



CASE STUDY

Health Care Organization Gains Robust GCP Data Lake to Streamline and Manage Data Ingestion



First Steps Toward Enhanced Ingestion of Customer Data

The Challenge

The customer data extracts provided to the organization by its members came from inconsistent data sources, requiring extensive manual review before ingestion into a Google data lake. The organization's business analysts were over-burdened, receiving multiple 100,000-record files with more than a million rows. The initial version of the customer project used a very manual process backed by hands-on support from Onix. It still took weeks to collect, clean and anonymize the data for analysis and visualization since it came from different data sources.

The Onix-Customer Relationship Reaped Early, Enriched Benefits

The Selection

The relationship with the customer began some years ago, when Onix helped the company develop its first data proof of concept. Onix completed various facility infrastructure migrations to GCP, and Onix has continued serving as a GCP and Workspace reseller for the organization over the years. Onix also has an existing Cloud Managed Services (CMS) relationship with the company.

Helping Hospitals Achieve Patient Outcomes and Financial Results

The Client

The customer, a non-profit community hospital organization helps its constituents remain community-operated and governed. The entity owns, manages and consults with hospitals across the United States, providing its members with the resources and experience needed to improve the quality of treatment outcomes, patient satisfaction and financial performance.



Version 2.0 Brings Numerous Improvements

The Solution

In Version 2.0, the process was revised so the organization's employees could manage the data ingestion. After automated receipt of the customer data extracts, validity scans were followed by Cloud Composer-orchestrated pipelines executing Python/Pandas code to transform the raw data files for consumption by the usable report formats in Connected Sheets.

The Onix implementation leveraged Cloud Composer and Google functions to reduce human error and manual manipulation. The transformation extract, transform, load (ETL) process shifted to management by Cloud Composer; Google Cloud Functions (GCF) handled file splitting. The raw CSV data files no longer required lookups to generate the reports; this was simplified using Cloud Composer.



BigQuery



Google Cloud Storage



Cloud Composer



Cloud VPN



Virtual Private Cloud



Cloud IAM



Logging

Significant Data Lake Time-saving Benefits Were Realized

The Results

Customer data extract files were uploaded to the drop zone of the standards-based Data Lake and initiated the automated ETL process. The Google Cloud Function split any consolidated data files into report data files, then the Cloud-Composer pipeline ingested the report data files into BigQuery automatically. The formerly weeks-long ingestion process was reduced to just minutes. In addition:

- Both an extension of the Onix CMS agreement to add DataOps for assistance with ingestion and pre-ingestion file review – and continuous development of the data ingestion pipelines developed by Onix are also possible.
- Exploration of additional ways to further automate the processing and report generation of healthcare data are also a possibility.

How Can a GCP Data Lake Benefit Your Organization?

Onix is pleased to provide your organization a consultation with a data expert.

[Request a Consultation NOW](#)

