Accelerating Cloud Platform Migration in Insurance

A Case Study of How Patterson Consulting Helped a Large Midwest Insurance Company Accelerate their Data Engineering Pipeline Development

Josh Patterson

Overview

Company Overview: Property Insurance Company

- Industry: Property Insurance
- Employees: > 10,000
- Headquarters: Midwest, United States
- Use Case: Platform Engineering, Data Engineering
- Stack: Google BigQuery, Google Cloud Composer, Kubernetes, Google Batch Processing, Python

Insurance Company's Challenge

This insurance company needed to migrate several data pipelines from AWS to GCP. This platform required not only data engineering to port the pipelines from AWS Redshift to GCP BigQuery, but also significant platform engineering to manage the inforamtion architecture.

Patterson Consulting's Solution

The Patterson Consulting team accelerated the migration by:

- Bringing in multiple data engineers to help the existing engineering team port data pipelines
- Adding in the right platform engineering expertise to make sure the new data pipelines would have good performance
- A professional engineering process that kept the customer's end goals in mind setting them up for future success

The Story

As a significant operator in the home and auto insurance space, this major insurance provider offers coverage for nearly half the states in North America. With an expansive portfolio of insurance products, this insurance provider is on the cutting edge of developing new machine learning models to help drive down costs for customers.

As any enterprise knows, building machine learning models requires a robust data engineering practice supported by a cloud platform and many data pipelines. These pipelines integrate many raw data sources such as raw text logs to exports from the data warehouse. Each data pipeline requires regular maintenance and upkeep as well.

As the insurance provider's management team kicks off new initiatives, the data science and data engineering teams have to not only maintain the current data pipelines, but also add new data pipelines for new projects. Additionally, the company decided to move some of the data pipelines from AWS Redshift to Google BigQuery. The ongoing workload plus new initiatives began to create a backlog of work tasks for the data engineering and platform engineering teams.

To help catch up the insurance provider's data engineering teams and their platform engineering teams, Patterson Consulting brought 8 engineers in to provide quality staff augmentation for 4 different groups. These engineers worked side-by-side with the insurance provider's data engineering and platform engineering teams for a year to get the data pipelines migrated and development tasks completed.

Why Patterson Consulting?

Patterson Consulting has consistently shown the ability to lead projects or augment existing engineering staff to help meeting project deadlines.

The insurance provider chose to work with Patterson Consulting over other vendors because of our team's track record with building robust and secure cloud data platforms. From data pipelines to kubernetes resource orchestration, Patterson Consulting has helped dozens of customers build more stable, secure, and performant data platforms.

Project Overview

Patterson Consulting's team started off by quickly getting the lay of the land for each working group, with quick and efficient onboarding so they could make an impact as efficiently as possible.

The platform engineering teams began working on an information architecture plan in Google Cloud Storage and Google BigQuery for moving data pipelines input, staging, and output datasets for machine learning modeling projects.

The Patterson Consulting data engineers quickly began documenting the AWS Redshift queries that needed to be transcribed into Google BigQuery SQL query dialect. Each data pipeline had specific requirements for input and outputs that had to be maintained for any port to another cloud platform and the Patterson Consulting team worked diligently to test that each data pipeline met the requirements.

Another team of engineers worked at building pipelines in Google BigQuery to process raw text data for a new project that used new datasets. Each set of data pipelines performed transforms to build features for downstream machine learning models.

Finally, the platform engineering team at Patterson Consulting supported the insurance company engineers in helping integrate Google Cloud Composer (Apache Airflow) with the new data pipelines for orchestration. For some tasks, Kubernetes integration was required for specific resources to be managed and orchestrated.