

Deluxe Corporation: Streaming Data Platform for Machine Learning

Overview

Deluxe Corporation is the company that invented the checkbook. Founded in 1915, they've evolved with their customers over the last century from "the check company" into a partner that deepens customer relationships through trusted, technology-enabled solutions, for everything from web marketing to payroll services to remote deposit capture and integrated receivables. Their ability to keep up with the times has led them now to boast 4.8 million active small business customers and more than 4,600 financial institution clients.

To keep up with their continual evolution and the data-driven products that made clients successful, Deluxe knew they needed to employ advanced analytics, machine learning, and real-time data availability. As they began exploring these technologies, they realized their existing ETL solution needed to be replaced.

Challenge

With 300+ disparate data sources, Deluxe was finding it impossible to leverage all available data for important tasks like advanced analytics. "When you need to join different sources coming from so many different locations, using classic ETL is just impossible," said Jean-Marie Bertonecelli, big data and analytics architect.

In addition, Deluxe didn't have the ability to capture data when it changed in real time. Since advanced machine learning requires

refreshed versions of data over time, they couldn't build the advanced machine learning models they needed.

Data availability was slow and the ability to use the data the way they wanted was difficult. Answering a request from a customer or delivering a data product often took up to a few days. For example, Deluxe has a specialized business line of data products that provide advanced marketing solutions. The customer provides some initial data and Deluxe augments the data to make it more precise for end-user targeting. In a world where personalized, real-time marketing and customer experience rules, providing data that is even a few minutes old taints results and misses an opportunity for action. Jean-Marie said, "It wasn't sustainable. Not having the type of data we need to answer a business request right away was a huge problem."

Solution

Having used StreamSets at a previous company, Jean-Marie spun up an instance of StreamSets Data Collector on Microsoft Azure Cloud to demonstrate how StreamSets could address these challenges. They were able to quickly build data pipelines and leverage automation for ingest and change data capture of click-stream data and web and product usage metrics, providing an end-to-end solution.

“StreamSets will ingest data coming from our 300+ sources, and clean and augment the data in flight. This definitively helps us by centralizing the data and letting us build the data warehouse in the Azure Cloud, on our big data platform,” said Jean-Marie. He went on to share that StreamSets is a key component of their database modernization, with StreamSets’ smart data pipelines for moving data into and out of the data platform. And, StreamSets automation and change data capture (CDC) abilities let them build important machine learning models.

Result

Deluxe’s evolution to serve their customers continues. They’re in the midst of a ‘New Day’ initiative, breaking down legacy silos, building new platforms and new approaches to deliver the right information at the right time to their business partners.

“Deluxe is data-centric, meaning most of our business is data-based. Without data we are nothing. By moving to the cloud with StreamSets and by leveraging their big data ecosystem, we can finally use all of our data and break down the silos we’ve struggled with in the past.”

ABOUT STREAMSETS

StreamSets built the industry’s first multi-cloud DataOps platform for modern data integration, helping enterprises to continuously flow big, streaming and traditional data to their data science and data analytics applications. For more information, visit www.streamsets.com.