CASE STUDY

Datalake Enrichment Layer at a National Retail Chain

Introduction

The client is a nation-wide retail chain in India. They are expanding their small format stores by 25x over the next few years to 10000 stores. They are also switching these stores to being membership-based (ala Costco, Metro et al) and its associated business model. The client organized their efforts around the notion of personas that they associate with every end-customer. Given their scale, they needed accurate understanding of the personas, their distribution, demand and context at a fine granularity so that they can plan assortment, incentives, training, and marketing while also addressing other operational issues.

The client needed a platform that could look at easily pick the relevant sources of data, allow for the laying of data pipelines (with integrity and quality checks), and then perform the following functions:

1. Codify these personas, classify customers, and enrich all entities
2. Model the accuracy of the personas, and help evolve them with modern ML techniques
3. Integrate with their existing BI and order management systems

The solution is being used to solve for multiple opportunities, including:

A. Improved, real-time inventory tracking and management:

Enrich matches EOD inventory with sales data as it is coming in and shows where the gaps are, to prevent potential stockouts in near realtime

B. Improved, real-time inventory tracking and management:

Enrich tracks SKU level product feedback, discounts, wastage, sales etc. to project ROI, quality and timeliness guarantees, etc.

C. Identification of shopping trends and cross-selling opportunities:

Monitoring uptake by SKU and making stocking recommendations, Lookalike models based on enriched data.

D. Personalized recommendations and multi-level reward programs based on purchase preferences:

Building an ever richer, dynamic profile of each customer, and providing internal tiers of rewards.

The platform accesses core business data and has critical business processes dependent on it. This requires the solution be simple, robust, secure, auditable, evolvable, and operated in close cooperation with the data lake team.

Existing solution alternatives suffered from cost and complexity (internal data infrastructure, some third party enterprise data platform vendors), over-engineering and obsolescence (some enterprise data platform vendors), and roll out timeframes (analytics vendors). They needed a solution that is non-intrusive, operates at the granularity of individual customer, and cost effective as it scales.

Scribble Enrich was selected as the candidate solution to sit (architecturally) above the client's data lake.

Solution

Scribble Enrich was deployed behind the firewall in the client's private network. Enrich was customized with pipeline components, tasks and background services to address the client challenges. It integrated with the client's S3-based data lake, and processed transactions and other datasets. The enriched datasets were made available to end-users through the data lake and existing BI. No change was required for the client's data governance mechanisms.

ENRICH PLATFORM CAPABILITY	BENEFIT
Scalable enrichment pipeline	Handle upto 1M customers, upto 5GB/day, Auditable and reproducible
Automated consumption	Feed datasets to Data lake and BI system
Custom tasks	Built-in backup and notifications
Background services	Real time monitoring of the data lake for availability, quality
Enrich Software Development Kit (SDK)	Rapid development and rollout
Ad hoc Data Commands	Auditable and reproducible ad hoc data analysis

End Result

This is a live implementation, and positive ROI was realised in three months. The number of usecases that the solution helps solve within the client organization is growing each month.

About Scribble Data

Scribble Data is a Machine Learning powered decision analytics company. Our mission is to empower organizations to solve persistent business problems with data they can trust. With the Enrich Intelligence Platform, we enable businesses to make high-impact decisions fast, with reliable and trustworthy data. Our seamless data transformation and pre-built app store frees up valuable time for developers and analysts to focus on critical tasks and collaborate effectively when it comes to solving a multitude of data use cases.