Insurance

Manage data size and complexity with a single data delivery platform

fractal
INTELLIGENCE FOR IMAGINATION

The Big Picture

A leading insurer was faced with a data delivery challenge. It had an internal system for collecting policy quotes data across various lines of business. Although the data was very rich for analytical purposes, it was not usable due to its sheer size (>67 TB compressed) and complexity (>12k nested fields in XML).

The company needed to set-up a single data delivery platform that would enable different analytical and product user groups to query the data in an automated way. Another challenge the company faced was how to handle the various, disparate extractors that different teams across the organization had built, which could not handle the scale of the data and required additional man hours.

There were many other challenges. There were inconsistent business rules due to conflicts with the new data strategy. There was less granular data due to normalization. There was difficulty leveraging all of the historical data for insight generation. Manual intervention was required for data generation. A robust governance mechanism was needed. There was also no production grade system existing, which could be extendible.

Transformative Solution

Addressing the company's challenges meant solving three key problems:

- Data ingestion and management: Ingesting and collecting big data (~68TB and storing it on Hadoop).

- Data harmonization: De-personalizing sensitive (PII) fields and partitioning the data in a distributed environment. Optimizing the stored data using Avro and parquet containers.

- Data extraction: Creating a robust flask based UI to query the data in an automated manner and generate analytics-ready data easily, using map reduce and spark.

A single data delivery platform was created, aiming to streamline the process of data extraction, while consuming it in the purest form (XMLs). The platform had three main components:

- Partitioning the data in a Hadoop environment on the basis of certain parameters. The parameters were identified upon thorough assessment of business requirements.

- Building a Custom UI (User Interface) that would facilitate requests for data extraction based on the parameters identified, and rectangularize the nested raw data to CSV files, easing the analytical consumption of data.

- End-to-end integration of the platform with the existing framework, further enhancing and improving easy scale up and operationalization.

The developed platform acted as a single source of truth for any data related needs, which was scalable to the enterprise (usable across multiple business units, functions, and teams). The platform was developed, keeping the business rules in mind in alignment with the data strategy. Highly granular data was made available due to direct interface with the raw layer, and data points were made available to the lowest degree of granularity. Optimized storage and pre-processing resulted in the historical data being available for extraction. A fully-automated platform eliminated manual intervention and enabled on-demand data extraction. A robust governance and security mechanism was built in using security groups and Kerberos for limiting data access. The platform and application developed was per industry and enterprise standards and flexible for new enhancements.

The Change

As a result of this process, the company received a first-of-its-kind big data platform for data hosting, MDM and custom data access, and business intelligence handling ~1TB of data. This provided a single source of information to different analytical teams, offered support for different filtering requirements, and delivered a streamlined process by abandoning the existing multi-layered architecture.

Visit https://fractalanalytics.com/ or email info@fractalanalytics.com. Follow us: