



Insurance

Track customer journeys across insurance functions

The Big Picture

A major P&C insurer wanted to develop an early alert system for tracking deviations from its goals and benchmarks in its A leading insurer wanted to track customer journeys across insurance functions and also optimize contact center operations. The company's existing data was relational in nature, and even though it captured the customer history, a standard data structure (table) would not be able to capture the history in its entirety. The data the company housed was too big to be transformed or extracted using traditional tools.

More specifically, the company needed a flexible and robust ETL mechanism in place to convert relational parent attributes to a key value-based data structure that tracks consumer lifecycle journeys. Manual intervention was required for data generation, and different pipelines needed to be created to merge data, which was becoming a tedious process. Effort was needed to replicate similar functionalities across different data pipelines, resulting in additional effort and no standardization. A better security and governance mechanism was needed. Also, data was manually validated and loaded to Hive tables.

Transformative Solution

Solving the company's data challenges meant addressing three key focus areas:

- Developing a raw layer: The approach scheduled data ingestion from different structured and semi-structured sources using Airflow, an open source scheduler.
- Developing an integration layer: the approach merged raw data in Spark to create a key value-based data structure that contained the entire historical information with scalability to add more data points. Parquet and Avro-based compression were used for optimized storage.
- Developing the blended layer: The approach used Airflow-driven, use-case data generation, which was scheduled and free from manual intervention.

A data lake was proposed, developed, and implemented that would act as a single source of truth for any analytical data-related needs in the client vertical. A multi-layered architecture was developed and implemented, leveraging Airflow for automation and scheduling. The raw layer served as the holding area for the historical data and any new data that was ingested. The ingestion and validation were then automated and scheduled using Airflow.

The integration layer served as the merging area for performing transformations, as well as converting data to a merged key-value format, containing end-to-end consumer history from different parent attribute tables. The blended layer contained data for various use cases and was scheduled using Airflow. The merged data was stored in a compressed format, allowing the layered architecture to make incremental additions to the data lake. The system was governed and secure, using Unix and Kerberos-based access to the data.

The Change

- As a result of the engagement, the company attained:
- A layered architecture: This was scalable, robust, and secure, enabling standardized development.
- Modular reusable components: These provided plug and play components for data ingestion, validation and loading. These were developed in a standardized manner across different data teams, saving time and effort.

- Workflow manager driven automation and scheduling: This enabled a process with no manual overhead. Data generation could be scheduled as well as triggered manually from a user interface.
- Automated ingestion, validation, and data loads: Using Airflow resulted in faster parallel and error-free execution.

The data lake, by means of leveraging a key value-based data structure, allowed for transforming data to track customer histories and journeys across functions.planning. ssssss

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

© 2018 Fractal Analytics Inc. All rights reserved | Confidential and proprietary Information of Fractal Analytics Inc. Fractal is a registered trademark of Fractal Analytics Limited.