StreamSets in Health Care Services

Overview

Data is the lifeblood of the health care industry, but the world of data is exploding, offering new opportunities and challenges in the form of pervasive intelligence, which is the ability of myriad analytics groups to self-service ready-to-use health care data. Machine learning and artificial intelligence, applied by groups throughout an organization, can help in a variety of areas from predicting sepsis rates, driving better accuracy in claims prediction, and delivering precision medicine; the end result being better patient outcomes and more cost-effective service delivery.

Challenges

In the health care industry, things are changing all the time. Bringing in all available data regardless of size, shape, or structure and landing it in a centralized data hub is the goal of many organizations, but how do they achieve that without soaking up resources that distract them from their core focus of health care innovation? Priorities for managing health care data include:

Leveraging New Data Sources:
Health care companies around the world are tasked with digitizing the wide breadth of medical records, leveraging instrumented medical devices for predictive capabilities, and optimizing health care delivery at facilities. New standards like HL7 provide a framework for the exchange, integration, sharing, and retrieval of electronic health information but there are variations in how different organizations implement it.

Establishing Patient 360:
Health care companies want to have a full 360 degree picture of a patient's health but oftentimes that information spans systems and is siloed in separate locations.

Enabling a Unified Data Lake:
To overcome data silos, as well as issues of growing data volume and variety, a centralized enterprise data lake provides many benefits. However, ingesting data into the lake usually requires custom coding and specialized skills, translating to high costs and lengthy projects that delay getting data to data scientists and analysts.

Data Ingestion into the Cloud:
As many companies in this industry step away from legacy on-prem systems (including mainframes) to explore new use cases and different ways to leverage their data, migrating to the cloud becomes an integral part of their data strategy. However, it often becomes a challenge due to limited IT resources: teams are unsure of what will be compatible in the cloud, and even though there are many tools available, they have limited knowledge when it comes to understanding how to build and debug data pipelines. This is also a challenge when organizations go through mergers and acquisitions.

Complex Data Transformation (ETL):
Many health care organizations are implementing solutions that will increase the ability to apply real-time analytics and expand data access with self-service. In order to achieve these outcomes, they must first extract, transform, and load their data into a destination where it can be analyzed. Organizations with a fully on-prem environment usually have limited resources to handle this process.

Cybersecurity:
As companies move to modern data platforms, the inspection, detection and disposition of personal health information (PHI) is a challenge, especially when data is in motion. HIPAA requirements are increasingly unforgiving and often evolve at a pace that is difficult for companies to adapt in practice. This involves protecting data at origin, in-flight, and at its destination.

Solution

Streamsets helps health care companies accelerate drug research and development, improve patient outcomes, and make new data sources like bedside equipment, patient records and radiology images available to analytics teams and applications, while ensuring data is protected in-motion.

The StreamSets Platform helps companies design and run batch and streaming pipelines in a fraction of the time using a drag-and-drop environment that minimizes coding and facilitates collaboration. It also detects and handles data drift, which may manifest as added fields or changed data types that can occur without notice when data sources

are upgraded. Hundreds of complex data flow topologies can be managed with StreamSets, giving you end-to-end visibility into your data movement. In addition, users can create data processing pipelines that execute on any Apache Spark environment or on Snowflake. Using a simple to use drag and drop UI, pipelines for performing ETL, stream processing, and machine learning operations can be created with ease.

StreamSets Benefits

StreamSets enables health care organizations to:

- Develop a DataOps practice and manage the health, delivery, and security of critical pipelines feeding the organization's analysis and discovery.

- Gain quicker, governed access to data generated across multiple sources.

- Achieve compliance with industry regulations and data security, governance, and integrity guidelines.

- Quickly analyze and act on insights generated from large data volumes spread across disparate sources.

Impact

While the problems that health care companies face are indeed complex in nature and transformative in scope, the business impact and strategic advantages are paying off for those companies that decide to leverage data in a way that addresses future compliance requirements.

Companies like Availity use StreamSets to ingest data into a vast repository which helps lower cost and increase data discovery. Pharma companies like GlaxoSmithKline use StreamSets to accelerate the drug development process from 10 years down to 2.

Closing

StreamSets is helping health care companies leverage the data they need to realize business value from a growing list of strategic use cases. By providing a platform that intelligently handles the eminent evolution of data systems and practices, monitors and reports on data pipeline operations, and removes barriers from leveraging complex data types, health care companies can build groundbreaking functionality that impacts the health and well-being of our society.

Use Cases

Predict Readmission
Hospitals and care facilities want to understand admission rates to staff and allocate resources effectively. Detecting patterns in anomalous readmission rates for certain patient segments can lead to actions that improve patient outcomes.

Predict Sepsis and Rapid Response
Oftentimes caregivers only know a patient has sepsis when it is already affecting their health. Monitoring connected devices can detect early warning signs before a patient is affected and trigger action.

Medical Image Repository
Creating a data repository for scanned images and applying deep learning, NLP, and text analysis to the images. "StreamSets counts 4 of the top 20 health care companies in the U.S amongst its customers."

Genomics and Precision Medicine
Companies are using big data platforms and large corpi of data to map the human genome. These activities can give way to delivering precision medicine based on genetic predispositions.

Real World Evidence for Clinical Trials
Practitioners and patients agree that it takes too long to bring life saving drugs to market. Using more data and implementing simulation and predictive modeling companies can shorten the time to market for crucial drugs.

Clinical Data Lake and Member 360
Creating a single view of customer health that can be utilized by caregivers, insurance providers, and claims adjusters.

Pharma Supply Chain Optimization
Understanding areas by geo and cohort to deliver better stocks and educational material about available pharmaceutical solutions.

Health care Plan Fraud
Sometimes patients, providers and facilities mis-report services and charges to health care providers. This results in lost capital that is categorized as fraud. By looking at greater trends and detecting anomalies health care providers can identify potential fraud before payment is made.

Medicaid and TRICARE Services
Federal efforts to create operational efficiency in delivering and paying for medical care. Limited budgets and an abundance of empirical data make these initiatives paramount for government agencies.

About StreamSets

At StreamSets, a Software AG company, our mission is to ensure data engineering teams thrive in today's world of constant change. We do this by embedding the DataOps philosophy of "continuous data for the connected enterprise" into the StreamSets DataOps Platform. StreamSets empowers data engineers to build, run, monitor, and manage smart data pipelines for the modern data ecosystem.

StreamSets is the only data integration platform that provides a single design experience for all design patterns for 10x greater developer productivity; smart data pipelines that are resilient to change for 80% less breakages; and a single pane of glass for observing and monitoring all pipelines to eliminate blind spots and control gaps. With StreamSets, you can deliver continuous data for the modern data ecosystem and hybrid integration in a world of constant change. **For more information, visit streamsets.com**