StreamSets

Availity®

How a real-time data repository increases business value while lowering costs.

Overview

Established in 2001, Availity is a healthcare IT company whose platform helps nationwide health plans and providers share administrative, clinical, and financial information with each other in real time. Its platform isn't just processing transactions --Availity's platform has really become a destination for health plans and providers to collaborate, offering streamlined workflows, an intuitive user experience, and critical insights. The increased interaction and collaboration between healthcare stakeholders results in better care and outcomes for patients, and optimized financial performance for Availity clients.

Challenges

Data is core to Availity's business, and the dynamic nature of healthcare data continues to drive Availity to evolve its data strategy. As a leading Healthcare IT company, Availity continues to ensure its people, process and technology are enabling an innovative platform to improve the outcomes of its customers. The company is focused on implementing solutions that will increase the ability to apply real-time analytics, expand data self-service interfaces and apply a DevOps culture.

Availity's strategy focuses on enabling transformation whereby they "use the tool that meets the requirements (scale, reliability, performance, etc.) for each task vs. forcing all requirements into a traditional database with ETL tools," according to Jeff Currier, Director of Data

Management and Analytics at Availity. His team sought new technologies and culture that would deliver

- Low TCO

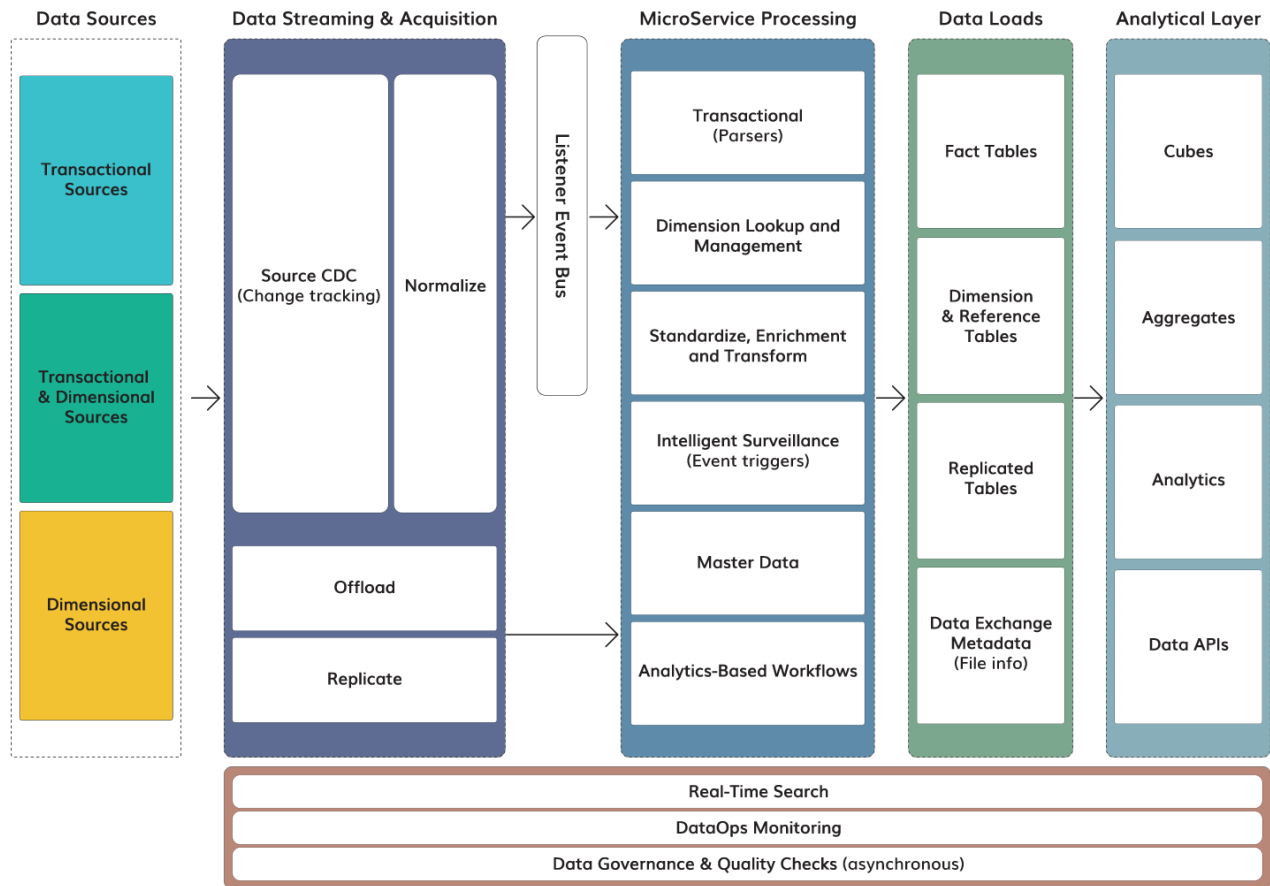- Performance and reliability at scale

The new platform needed to provide an economic way to move, store, and analyze hundreds of terabytes (TB) of data while providing leading and innovative solutions on the Availity platform.

Solution

A significant amount of data processing workloads are moving to a stream-first architecture, leveraging StreamSets, as part of Availity's "real-time data repository" built on a Hadoop-based Cloudera platform. StreamSets Data Collector is a simple, customizable solution that facilitates a streaming data architecture. Availity leverages a wide array of origins and destinations such as: Apache Kafka, Oracle, Apache Kudu, Apache Hbase, Elastic Search, Apache Solr and HDFS.

The capabilities delivered by StreamSets align with the culture that Currier and team are creating around "DataOps," or DevOps for data teams. StreamSets Data Collector allows Availity to break down components of data flow pipelines into micro services and then watch or tweak performance at a granular level. StreamSets Dataflow Performance Manager (DPM) provides visibility and end-to-end control over all pipelines.

Data Platform with StreamSets

Data Sources	Data Streaming & Acquisition			MicroService Processing	Data Loads	Analytical Layer
Transactional Sources	Source CDC (Change tracking)	Normalize	Listener Event Bus	Transactional (Parsers)	Fact Tables	Cubes
Transactional & Dimensional Sources				Dimension Lookup and Management	Dimension & Reference Tables	Aggregates
				Standardize, Enrichment and Transform		
				Intelligent Surveillance (Event triggers)	Replicated Tables	Analytics
Dimensional Sources	Offload			Master Data		
	Replicate			Analytics-Based Workflows	Data Exchange Metadata (File info)	Data APIs

Real-Time Search

DataOps Monitoring

Data Governance & Quality Checks (asynchronous)

Results

Availity's real-time data repository is delivering on the strategy to increase value while lowering costs through, in part, cost-efficient data movement at scale.

StreamSets can sync two systems in real time with throughput of thousands of records per second, whereas alternative solutions performed cyclical loads and would end up out of sync.

StreamSets is also facilitating the DataOps culture that Currier and team are cultivating. "Everybody is excited about StreamSets," said Currier. "It's simple and easy enough that they don't need to find a StreamSets developer, but more general data engineers/developers, to create their own data pipelines."

Long-term Benefits

By offering a self-service data repository, the company will be able to innovate faster. Availity expects customer satisfaction to rise as a direct result of deeper insights delivered faster and with greater reliability from the new platform.

Meanwhile TCO of its data management infrastructure is lower as they continue to evolve their use of Big Data solutions.