



Implementation of Enterprise Data Lake for Big Data Aggregation & Analytics

Success Story

CitiusTech designed and developed an enterprise data lake, based on Hadoop distribution, to provide a single point of secure access for Big Data and create a unified view of archived and current data

Business Impact

A leading benefits management company leveraged CitiusTech's healthcare Big Data, EDW and analytics expertise to create a single source of truth across a diverse range of data sources – to drive advanced analytics.

22+ TB

member data
ingested

65%

reduction in data
processing time

80%

reduced TAT in data
load / ingestion

\$30M+

savings in PA auto
adjudication

Client Requirement

Client is a leading medical benefits management organization covering 100 million lives. It contracts with 140+ health plans and commercial clients

The client needed to create a single source of truth across all enterprise data sources using Hadoop components and allow access through a Big Data analytics hub. It also the ability to extract and standardize data from multiple sources and ingest it into the data lake.

Given its expertise in healthcare Big Data, EDW and analytics, CitiusTech was selected to design and build the solution.

The CitiusTech Advantage

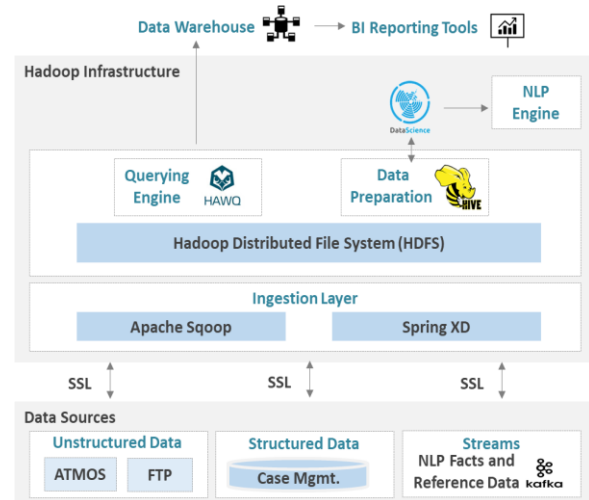
CitiusTech designed and implemented an advanced analytics solution using Hadoop-based scalable infrastructure. The solution included:

- A repeatable solution to ingest Big Data (structured, semi-structured and unstructured data) from multiple sources into the data lake
- Integration of NLP to process clinical notes and extract intelligence for clinical decision making
- Lower deployment costs, enhanced quality and improved time-to-market for its new solutions by leveraging CitiusTech's global delivery model

Technology Overview

CitiusTech designed, developed and implemented a holistic Big Data environment that served as a single, secure point of access for all enterprise data and provided a unified view of data to downstream systems. The solution included:

- Hadoop cluster using Hortonworks distribution and maintenance
- Framework for ingestion of data from RDBMS and streaming platforms
- Ability to process data in any form and feed it to downstream DW/BI systems
- Data federation fast and efficient access to data
- Workflow design for processing paid claims
- Defined strategy for disaster recovery



Solution Schematic

Technologies & Standards Used

- Pivotal Hortonworks distribution platform for Big Data implementation
- Pivotal HAWQ MPP data base on Hadoop for storage, data access and complex querying
- Pivotal Spring-XD for batch ingestion and Apache NiFi for real-time data processing
- Oozie for job orchestration and SQOOP for data movement
- Hive for data aggregation and JSON processing
- Security hardening using Kerberos and Ranger
- Integration of Hadoop cluster with Pivotal Greenplum and Microsoft Polybase

About CitiusTech

CitiusTech is a specialist provider of healthcare technology services and solutions to healthcare technology companies, providers, payers and life sciences organizations. With over 4,000 professionals worldwide, CitiusTech enables healthcare organizations to drive clinical value chain excellence - across integration & interoperability, data management (EDW, Big Data), performance management (BI / analytics), predictive analytics & data science and digital engagement (mobile, IoT). CitiusTech helps customers accelerate innovation in healthcare through specialized solutions, platforms, proficiencies and accelerators. With cutting-edge technology, world-class service quality and a global resource base, CitiusTech consistently delivers best-in-class solutions and an unmatched cost advantage to healthcare organizations worldwide.

Princeton | New York | Rochester | Dallas | Boston | Philadelphia | Mumbai | Bengaluru | Chennai | Dubai | Singapore | London