Client Requirements

Client is a leading diagnostic services provider that assists patients and physicians in making better healthcare decisions to improve patient outcomes. In order to enhance the quality of its analytics and decision support systems, client wanted to replace their legacy data warehousing solution with a data lake that could manage high volumes of data (5-10 million messages per day) and provide deeper interactive visualizations (historical and longitudinal reporting, trend analysis, etc.) and analytical queries on massive healthcare datasets.

CitiusTech was selected by the client to develop the infrastructure, given its strong expertise across healthcare workflows, data mining and standardization, big data processing and advanced analytics use cases in healthcare.

CitiusTech Solution

Requirement Analysis
CitiusTech put together a team of domain experts and big data professionals to do a detailed analysis of the client's existing data environment. Based on its analysis, CitiusTech designed the solution based on Cloudera Hadoop distribution, MapReduce, Spark Streaming and other Hadoop/ Big data technologies. The solution also

included building a flexible and configurable data model to represent healthcare data, and enable advanced analytics.
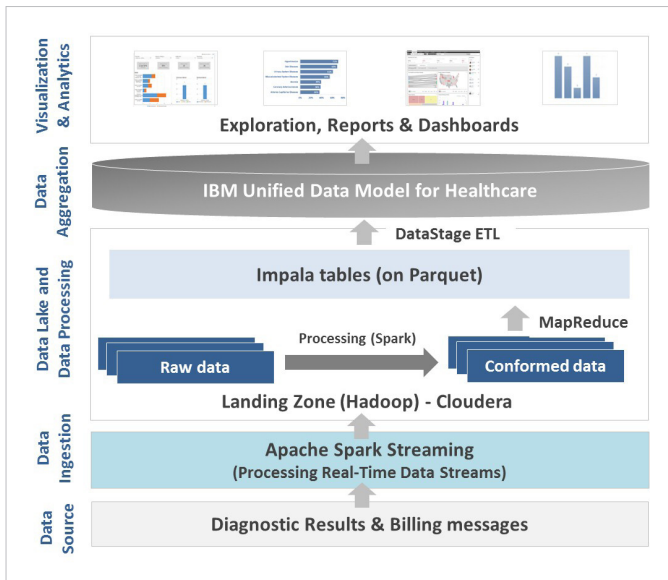
Solution Design

- **Data Ingestion and Parsing:**
  - Ingesting and storing HL7 V2.5.1, V3 (Spark Streaming) and BFE (billing data flat files) messages in the data lake (HBase + HDFS) via Spark, jobs that are capable of processing 20,000+ records/ second
  - Deduplicating, cleansing and standardizing (LOINC, SNOMED) data using MapReduce jobs
  - Parsing using custom HL7 and BFE parsers and storing data in HBase tables

- **Data Processing and Integration:**
  - Processing data in HBase using MapReduce and converting the messages into Parquet format that can be queried using Cloudera Impala
  - Exposing the Parquet data using Impala tables (compliant with the IBM healthcare data model) using IBM DataStage for ETL

- **Data Aggregation and Storage:** Populating and storing messages from Impala tables into IBM healthcare data model on IBM PDA

- **Data Presentation:** Using big data lake as the source of information to downstream analytics applications, providing visualizations using BI/ analytics tool that leverage multi-dimensional models like Cognos

Value Delivered

By partnering with CitiusTech, the client was able to:
- Leverage CitiusTech's expertise on Hadoop and big data technology to build a configurable data model that processes high volumes of data efficiently
- Reduce the dependency on the legacy system and build a data repository with standardized, cleansed and filtered data that is used for downstream consumption
- Leverage BI and analytics to provide powerful reports with rich visualizations to providers
- Drive key big data processing and advanced analytics use cases such as point of care decision support, population health reporting, etc.

About CitiusTech

CitiusTech is a specialist provider of healthcare technology services and solutions to medical technology companies, providers, payers and life sciences organizations, with over 2,700 professionals worldwide. CitiusTech's services and solutions include healthcare software development, healthcare interoperability, regulatory compliance, BI/analytics, consumer engagement, care coordination and population health management. CitiusTech helps customers accelerate innovation in healthcare through a number of solutions and accelerators for clinical quality reporting, healthcare big data, cloud computing, mobile health and predictive analytics. With cutting-edge technology expertise, world-class service quality and a global resource base, CitiusTech consistently delivers best-in-class solutions and an unmatched cost advantage to healthcare clients worldwide.

Princeton | Rochester | Dallas | Toronto | London | Dubai | Mumbai | Bengaluru | Singapore

CitiusTech