

Case Study: Big Data Analytics Using Natural Language Processing

Client Requirements

Client is a leading provider of Health Information Management (HIM) solutions, serving over 3,000 hospitals and integrated delivery networks across the U.S. and helping them to improve efficiencies and streamline operations.

Client needed to integrate big data capabilities within its solutions to process and manage massive amounts of unstructured electronic medical records and clinical data that it receives from various sources and prevent loss of meaningful information.

Client leveraged H-Scale, CitiusTech's big data platform, and CitiusTech's expertise in clinical workflows and healthcare interoperability to implement a robust, scalable solution that supported large volumes of unstructured data. The solution also provided an explicit structure to facilitate efficient data search, ad-hoc analytics and reporting to generate actionable insights.

scalability. Natural Language Processing (NLP) was used to process the unstructured data to ensure persistence of relevant clinical content. CitiusTech team conducted a detailed assessment of the client's requirements including:

- Performing gap analysis to customize H-Scale according to client's requirements
- Analysis of EMRs, clinical notes and sample data feeds
- Identification of language semantics and patterns for NLP training data

Solution Implementation

CitiusTech put together a team of big data, clinical informatics and interoperability professionals to design, develop and integrate the solution.

Data Ingestion and Parsing:

- Ingesting and storing of data feeds in a data lake (HBase + HDFS)
- De-duplicating, cleansing and standardizing data
- Parsing data with pre-built Hadoop parsers

Data Processing:

- Creating models to represent unstructured data and training the NLP system with the models
- Implementing NLP service using Apache cTAKES to process the clinical data, extract clinical content and persist in transaction repository

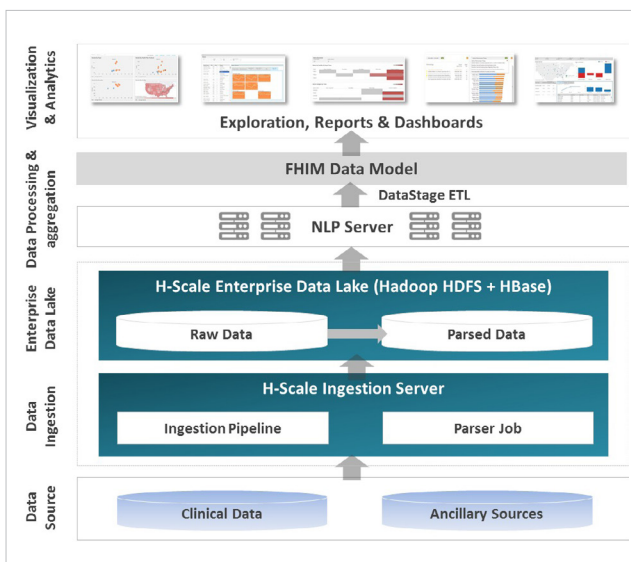
Data Aggregation and Storage: Mapping and populating the output to FHIM data model

Data Presentation: Leveraging BI/analytics tools for interactive visualizations, ad-hoc analytics and reporting

Value Delivered

By partnering with CitiusTech, the client was able to:

- Leverage CitiusTech's big data platform, H-Scale, and Hadoop expertise to build a flexible and robust healthcare data lake that could store and process high volumes of data
- Develop NLP capability to process unstructured data and prevent loss of information using Apache cTAKES on Hadoop
- Accelerate deployment and go-to-market by reusing existing technology platforms and components



CitiusTech Solution

H-Scale and NLP for big data

Client leveraged H-Scale and Hadoop technologies to build an enterprise data lake to store large volumes of incoming data from multiple sources and to ensure

All product and company names mentioned herein are trademarks of their respective owners

About CitiusTech

CitiusTech is a specialist provider of healthcare technology services and solutions to medical technology companies, providers, payers and life sciences organizations. 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.

Seattle | Dallas | Princeton | London | Mumbai | Airoli | Singapore