Geisinger Health System, based in Danville, PA, is one of the largest health services organizations in the United States, serving more than three million residents throughout northeast, north-central, and south-central Pennsylvania, and southern New Jersey.

Geisinger is one of America’s leading rural healthcare organizations. Its integrated, physician-led system includes 30,000 employees, nearly 1,600 physicians, 12 hospital campuses, two research centers, a 530,000-member health plan, children’s hospital, and an alcohol and chemical dependency treatment facility. Founder Abigail Geisinger and her first chief of staff, Dr. Harold Foss, opened the health system’s first hospital, Geisinger Medical Center, modeled after the Mayo Clinic, in 1915. Over the past century, Geisinger has built a national reputation for clinical innovation, research, medical education, and patient experience.

**Improving The Patient Experience**

Geisinger’s reputation for outstanding patient experience is due in large part to its focus on innovations to enhance patient care, integrated health system approach, and an overall focus on caring and compassion. Geisinger was an early adopter of the Electronic Health Record (EHR) and has implemented the EHR throughout the health system. This digital integration connects the system’s hospitals, 40 community practice sites, and the primary and specialty care physicians and extenders who serve patients throughout the Geisinger network.

---

**CASE STUDY**

**GEISINGER HEALTH SYSTEM**

**KEY HIGHLIGHTS**

- Onboarded 30TB of data
- Saved $2MM in EDW replacement costs
- Saved $500k in annual maintenance costs
In addition, with its Geisinger Health Plan, Geisinger Health System is both a payer and provider. Witnessing first-hand the evolution of healthcare and changes in payment models, Geisinger Chief Informatics Officer Alistair Erskine, MD, and former Chief Data Officer Nicholas Marko, MD, understood the need for Geisinger to develop modern data applications to enhance both the efficiency and effectiveness of the healthcare delivery network, including a 360-degree view of its patients. Yet Geisinger faced many technological challenges along the way.

The quality of Geisinger’s integration between EHRs and its delivery system depends on its ability to make data quickly and easily available to caregivers. To that end, Geisinger’s physicians and data scientists wanted to combine the terabytes of data already in the Epic EHRs with data from clinical department systems (including radiology and cardiology), as well as patient data from Health Information Exchanges (HIEs) and patient satisfaction surveys. The data diversity challenge took on more urgency as other hospitals and clinics joined Geisinger Health System. Those additions brought new patients with different data profiles, and Geisinger needed to ensure it could provide the same quality of care to all of its patients.

In addition to its existing data, Geisinger anticipated a trove of new data from devices that weren’t even invented when the health system architected its EHRs. Geisinger aimed to proactively prepare for real-time ingestion of new data. Devices within its hospitals and lab relayed critical patient vitals useful for detailed monitoring of every step of the patients’ journeys, but that detail was neither available quickly enough for near real-time decisions, nor stored with sufficient granular detail to surface subtle opportunities for improvement. Geisinger recognized that through deep analysis of both Internet of Things (IoT) data and manual physician notes, it could innovate and author the newest chapter in healthcare delivery.

Health system leaders faced two challenges before they could turn that vision into reality. First, they needed to wrangle Geisinger’s data-in-motion under control more quickly and efficiently. Second, they needed to reduce the cost to store all data so data scientists could derive deep insight from comparison of raw data spanning years and millions of doctor-patient interactions.

### Assembling A New Healthcare Delivery System

In the fall of 2015, Geisinger began transitioning its architecture to meet those business and clinical needs. The responsibility for architecting and building the next generation platform rested with Bipin Karunakaran, Vice President for Enterprise Data Management, and Data Management Directors Joseph Scopelliti and Mark Mossel. They turned to Apache Hadoop® and Hortonworks Data Platform (HDP®) to consolidate the structured and unstructured data. The initial use focused on filling the gaps not met by their Enterprise Data Warehouse (EDW), and enriching that patient data with financial data such as billing records and data from the Centers for Medicare Services (CMS). Recognizing the economies of scale with an open-source approach, Geisinger soon turned to Hortonworks Connected Data Platforms to meet these challenges and speed the delivery of actionable insight from both data-in-motion and data-at-rest.

The health system was soon onboarding over 30TB of important patient data. At the same time, initiatives for precision medicine and MyCode, Geisinger’s genomic marker analysis for patients, were generating significant amounts of data and processing needs. Fortunately, the decision to use HDP ensured that all relevant data would be captured and consolidated, no matter the source or schema.

### Immediate Advantages And Savings

#### CONSOLIDATING DATA

Geisinger immediately began the process of archiving and processing its 30 terabytes of patient data from its EDW into HDP. For most organizations, especially those in healthcare, the associated storage costs are a prohibitive factor. Geisinger, however, was able to save $2 million in EDW replacement costs and $500,000 in annual maintenance costs by eliminating the need to continue and expand its EDW platform. Geisinger also leveraged the ability to retain its existing EDW queries and use them in the Hadoop SQL Workbench, which saved time and eased the transition.

#### QUERYING UNSTRUCTURED DATA

After Geisinger successfully on-boarded its structured data, attention turned to its unstructured data. A vast trove of medical records and doctor notes came into HDP in non-structured text format, but then had to be queried. Through the use of Apache Solr, the powerful open-source search platform that ships with HDP, Geisinger now runs queries on its unstructured data to derive analytical insights. Analysts now look at the sequence of patient visits, prescriptions, and medical records. Using Solr, clinicians and non-clinicians are able to search through 200 million patient note records in seconds to find relevant conditions and medications, which helps them analyze the success of treatments, identify areas of improvement, and determine ways to save time and money for both patients and providers.
HADOOP FOR EVERYBODY

With the merging of structured and unstructured data, Geisinger tapped into a previously untapped well-spring of thought and innovation. Its initial trial of EDW offload into Hadoop was limited to a few select users who only ran one SQL query. Soon there was a much larger demand for a wider variety of queries.

Geisinger also took advantage of the data governance and security features of HDP. This ensured success and compliance for every user who required it, and also ensured that data scientists who wanted to leverage similar functionalities such as R, but with the increasing scaling and performance of Apache Spark, could do so. The doctors and other hospital users who wanted to use Solr for querying unstructured data could accomplish this uninhibited, but according to their permissions.

This level of data governance permits a fluid distribution of data to different users, so that they can each make their own query without affecting the queries of others. For instance, an ad-hoc analytical study using Spark can instantly be spun up and down with no downtime or hassle.

Prescribing the Future of Healthcare

The next major leap for Geisinger is Hortonworks Data Flow (HDF), powered by Apache NiFi, Apache Kafka, and Apache Storm, an integrated system for real-time dataflow management and streaming analytics on premise or in the cloud. With data streaming in through HDF, Geisinger believes it will be able to provide real-time alerts and recommendations to its patients, as well as prescribe more precise treatments. Geisinger looks forward to all of this while continuing to benefit from deep historical analysis and machine learning already available in its existing data lake.

About Geisinger

Geisinger Health System, based in Danville, Pennsylvania, is one of the largest health service organizations in the United States, serving more than 3 million residents throughout Pennsylvania and southern New Jersey. Geisinger is one of America's leading rural healthcare providers, with an integrated, physician-led system that includes 30,000 employees, nearly 1,600 employed physicians, 12 hospital campuses, and two research centers. Abigail A. Geisinger and Dr. Harold Foss opened the system's first hospital in 1915, modeled off of the Mayo Clinic. For more than a century since then, Geisinger has built a national reputation for clinical innovation, research and patient experience.

About Hortonworks

Hortonworks is a leading provider of enterprise-grade, global data management platforms, services and solutions that deliver actionable intelligence from any type of data for over half of the Fortune 100. Hortonworks is committed to driving innovation in open source communities, providing unique value to enterprise customers. Along with its partners, Hortonworks provides technology, expertise and support so that enterprise customers can adopt a modern data architecture. For more information, visit hortonworks.com.

Contact
For further information, visit hortonworks.com

© 2011-2018 Hortonworks Inc. All Rights Reserved.
Privacy Policy | Terms of Service

© 2018
05/2018