



Improving healthcare through advanced analytics

Introduction

Healthcare is among the industries that may benefit most from advanced data analytics. Providers, payers, researchers, medical device designers, and pharmaceutical developers increasingly are improving patient outcomes through an enhanced understanding of data available today across the healthcare community.

Data is emerging as an essential tool for predicting patient events, detecting epidemics, enabling genomic analysis, and delivering on the promise of molecular medicine. The ability to mine extremely large data sets with robust analytical platforms and tools provides an unsurpassed capability to uncover and resolve a limitless array of global healthcare concerns.

That's why more than one hundred health and life sciences organizations, including the leading health systems, data providers, health IT companies, and pharmaceutical organizations are engaged in this modern medicine revolution using Cloudera solutions.

Health Systems/Providers

Health providers benefit from data analysis in several ways. Patient outcomes are being improved and hospital stays reduced with precision medicine approaches, facilitated by big data research. Applications range from principal investigators doing research on specific diseases to regional hospitals working to minimize readmission rates or understand approaches to care and patients in the clinic.

Streaming data from the bedside is increasingly used in real time to report to caregiver stations, which results in reduced patient risk and more timely care, while contributing to optimized provider productivity. This kind of data analysis is helping global healthcare providers organize facilities and staffing resources to ensure they deliver the most effective and efficient care.

One prominent children's hospital, for example, has leveraged previously discarded data to design new, difficult-to-treat symptoms. Implementing a Cloudera system in a do-it-yourself fashion and incorporating streaming data from bedside sensors, the center now is a leader in mining data for improving pediatric patient care. One early result has helped determine optimal post-treatment care for retinopathy in premature infants. Analysis of data from multiple infants led to the identification of the most effective procedures that have been implemented across the board with significant, positive results.

Pharmaceutical Companies and Research Institutions

Genomic research and real-world evidence studies, by their nature, yield extremely large data sets. Likewise, stores of medical imaging data are growing continually, as is research data and information from wearable body sensors and other Internet of Things (IoT) devices. These sources in combination are invaluable to pharmaceutical companies and healthcare providers for drawing insights into trending health issues, correlations of symptoms, and identification of new and effective treatments.

Stored genomic data could be as large as 40 billion gigabytes (GB) by 2025.

Organizations worldwide have embraced big data analysis, as evidenced by companies that employ Cloudera solutions producing five of today's most widely used drugs. Additionally, the Broad Institute, which maintains a 19-petabyte (PB) repository of genomic data, uses Spark, an open source project, for development of future versions of Genome Analysis Toolkit (GATK). Cloudera, with the most Spark customers being supported, collaborates with this and other research organizations around the globe.

Health Plans

Healthcare plans focus on care coordination, improved outcomes, cost optimization, and many other areas of member health via data science and informatics. They process enormous numbers of claims daily that have significant financial impacts on their patients, providers, and their own organizational performance. For health plans, one of their primary concerns is ensuring payment integrity. In addition, with ever-increasing healthcare costs, these organizations require methods to ingest and analyze data that help reduce unnecessary expenses and predict future population healthcare needs.

Cloudera is actively engaged with four of the five largest U.S. health insurance providers to support a number of their informatics requirements, including detection of potential fraud after payment toward streaming analytics to catch improper payments at the outset.

Healthcare Data Companies

The medical community's conversion to electronic health records (EHR) has resulted in a treasury of data that can be de-identified and shared with healthcare providers and insurers to optimize their operations. Aggregated EHR data is critical to understanding population health, making predictive analyses, and, most importantly, protecting and improving patient care. Data analytics enable definition of the most cost-effective outcomes, who needs the highest level of care, patient pathways, and other insights. With a goal of bringing together a large portion of health data into a comprehensive health platform, Cerner turned to Cloudera and has succeeded in bringing more than 2 petabytes of data onto their enterprise data hub while insuring the security and data integrity of the platform.

Cerner, the largest global EHR provider with more than 500 million anonymous patient records, employs 2,000 nodes of Cloudera to power its real-time sepsis prediction agent.

The Future of Healthcare Analytics—What's Next?

Access and insight to the ever-increasing quantities of health and genomic data generated will lead to further inroads against the most important killers—cancer, heart attack, and more. Advanced data analysis can be used to expand public awareness, create targeted prevention, and address healthcare at the molecular level. In the near-term, the healthcare community can use the:

- Proliferation of remote monitoring devices and the data they collect for analysis and treatment design, including sensors and other new sources of electronic signals toward patient-centered and clinic-centered Internet of Things
- Data analysis platforms and tools to determine preventive measures and effective cures for long-standing maladies, including various forms of cancer, to emerging illnesses like microcephaly from the Zika virus
- Worldwide application of predictive analytics for signal detection and to identify potential pandemics and mitigate disease outbreaks at the source
- Applications for newer technologies such as Hyperledger for patient-controlled access and centralization of a patient's medical record

The Cloudera platform is available as open-source software and, as such, it continues to help make these healthcare advances available on a global scale. Data scientists in developing countries now can participate and contribute to the international community dedicated to improving healthcare in their countries and around the world. By employing an open-source approach, physicians, data analysts, and programmers can add their talent, experience, and creativity to this most important global initiative.

For more information on how Cloudera supports health and life sciences professionals, please visit: <http://www.cloudera.com/solutions/healthcare.html>