



Saving Lives with Big Data Analytics that Predict Patient Conditions

Overview

Cerner Corporation, a longtime leader in the **healthcare** IT space, is in the midst of an evolution. Today, its solutions and services are utilized in over 14,000 medical facilities around the world, such as hospitals, integrated delivery networks, ambulatory offices, and physicians' offices.

But Cerner's goal is to deliver more than software and solutions. The company is expanding its historical focus on electronic medical records (EMR) to help improve health and care across the board. Cerner aims to assimilate and normalize the world's healthcare data in order to reduce cost and increase efficiency of delivering healthcare, while improving patient outcomes.

The firm is accomplishing this by building a comprehensive view of population health on a **Big Data** platform that's powered by a **Cloudera enterprise data hub** (EDH).

The Challenge

As **David Edwards**, Vice President and Fellow at Cerner, described, "Our vision is to bring all of this information into a common platform and then make sense of it – and it turns out, this is actually a very challenging problem."

When the Cerner team set out to build this platform, they established a few key objectives. It would need to be:

- Capable of bringing together all of the world's health data
- Secure, traceable, and audited
- Catalogued and explorable
- Usable for any need at any time

"In our first attempts to build this common platform, we immediately ran into roadblocks," said **Ryan Brush**, Senior Director and Distinguished Engineer at Cerner. "Most tools available at the time weren't really a great fit for the magnitude or complexity of the global healthcare data challenge we were trying to address. We started out somewhat modestly —building search indexes for medical records—but even this required huge amounts of computational power."

Key Highlights

Industry

Healthcare

Locations

- Headquarters: Kansas City, MO, USA
- Facilities worldwide

Business Application Supported

- Comprehensive view of population health data to improve patient treatments and outcomes

Impact

- Unprecedented ability to coalesce data from unlimited sources
- Improved patient diagnosis delivers measurable increase in lives saved
- Enhanced operational and financial performance in healthcare facilities

Technologies in Use

- Hadoop Platform: Cloudera Enterprise, Data Hub Edition
- Components in Use: Apache Crunch, Apache HBase, Apache Hive, Apache Kafka, Apache Oozie, Apache Storm, Cloudera Manager, MapReduce
- Servers: HP
- Data Mart: HP Vertica
- BI and Analytics Tools: SAP Business Objects, SAS

Big Data Scale

- More than 2 PB on Cloudera today
- Multi-tenant environment supports several hundred clients

To move forward, Cerner needed an achievable way to throw significant CPU power at a very large dataset without compromising agility.

“We needed the ability to iterate quickly on our search processing algorithms,” continued Brush. “Our original event-driven prototypes allowed us to run everything through the event pipeline, but it clearly wasn’t optimized for what we wanted to do.”

Solution

The Cerner team looked for up-and-coming technologies that would help overcome the restrictions found with traditional approaches, and **Apache Hadoop** presented a viable alternative. Brush had downloaded the raw open source code from Apache in 2009 and was delivering impressive results in support of several different use cases already. In 2010, he had adopted Cloudera’s Distribution Including Apache Hadoop (CDH) because of its integrations with **Apache HBase** for random access to the data in Hadoop.

When Cerner decided to build its comprehensive population health platform on Hadoop in 2013, the team knew it would need a partner. “Hadoop is a very sophisticated ecosystem of technologies, and we are a healthcare company, not an infrastructure company,” Edwards explained. “We decided to find a partner that could help take care of the infrastructure, allowing us to focus on the real, healthcare-related problems that we’re trying to solve.”

Edwards’ team identified a few leading commercial Hadoop providers and came up with the following evaluation criteria, on which each offering was scored:

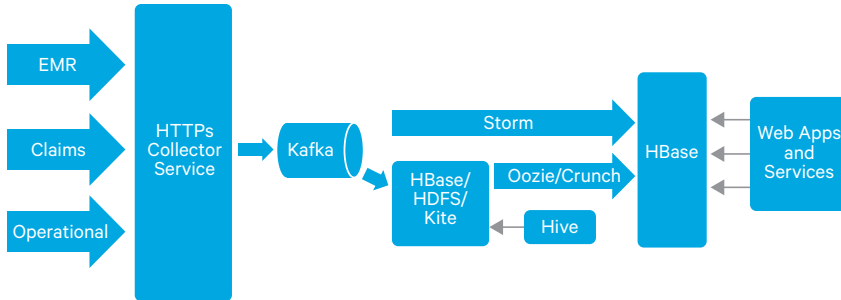
- Management tooling
- Scalability and performance
- Support quality and options
- Training options
- Ability to accommodate data science and ad hoc analysis
- Involvement and leadership within the open source Hadoop community
- Partner integrations
- Security
- High availability and disaster recovery options
- Data management tooling
- Price

Cloudera earned the highest score. Cerner also already knew and liked CDH, and had developed respect for the **Cloudera engineers** who actively contribute to the open source community, which made the decision to move forward even easier.

Brush added, “In addition to the quantitative results from our assessment, we felt that the Cloudera approach was more aligned with our own philosophies, such as building simpler, more prescriptive libraries that broaden the audience for the platform. We also appreciated seeing key open source projects—such as **Crunch** and Kite—being driven by engineers from Cloudera.”

Today, Cerner’s enterprise data hub contains more than two petabytes (PB) of data in a multi-tenant environment, supporting several hundred clients. “Our deployment was rapidly expanding, so we brought everything under the control of **Cloudera Manager**,” noted Edwards. “It provides a holistic view of our whole environment, and allows us to manage multiple clusters from a central point.”

The platform ingests multiple different EMRs, HL7 feeds, Health Information Exchange information, claims data, and custom extracts from a variety of proprietary or client-owned data sources, uses **Apache Kafka** to ingest real-time data streams, and then pushes data to the appropriate HBase or HDFS cluster using **Apache Storm**. Cerner is exploring adding other real-time components to the platform as well, such as **Apache Flume**, **Apache Samza** (incubating), and **Apache Spark**.



“We’re able to achieve much better outcomes, both patient-related and financial, than we ever could by just looking at pieces of the puzzle individually. It all comes down to bringing everything together and being able to extract value for any requirement. The enterprise data hub topology allows us to do exactly that.”

Ryan Brush, Senior Director and Distinguished Engineer, Cerner

Data moves from the Cloudera environment to Cerner’s **HP Vertica** data marts via bulk loads, giving data scientists, **SAP Business Objects** users, and **SAS** users the ability to interact with Hadoop data for broad reporting and analysis using tools they’re familiar with. This helps them understand the most significant risks and opportunities for improvement across a population of people. For instance, Cerner computes quality scores for managing a number of chronic conditions, and analysts can use Business Objects to see which conditions could gain the most by improving those scores. The end result: better use of health resources.

Cerner is starting to leverage **SAS** on Hadoop for deep data science as well, for instance, to build prediction models for avoidable hospital readmissions. These use cases are expected to grow over time.

The Cerner team is also evaluating tools such as **Cloudera Search** and **Impala** to allow users across the organization to interact directly with the data in Cloudera, and is considering connecting **SAS** directly to Cloudera using **SAS/ACCESS Interface to Impala**.

Cerner has also taken steps to ensure the security and data integrity of its Big Data platform. In the healthcare space especially, a technical solution must provide a mechanism for threat mitigation in order to be considered a viable data management technology. Edwards said, “Our Cloudera environment holds actual patient data, so it’s imperative that everything be completely protected. We’ve designed the infrastructure to ensure that all the information is secured behind multiple firewalls, with multiple levels of authentication being required in order to just begin to get access.”

Cloudera advised Cerner’s approach to encrypting data at rest and on its Kerberos integration, and Edwards’ team values Cloudera’s dedication to enhancing security on Hadoop. Cerner is actively evaluating tools like **Apache Sentry** (incubating) to complement what the team has already built.

“Our clients are reporting that the new system has actually saved hundreds of lives by being able to predict if a patient is septic more effectively than they could before.”

Ryan Brush, Senior Director and Distinguished Engineer,
Cerner

Impact: Improved Insights Save Lives

Traditional healthcare IT solutions tended to be limited in scope and restricted to a particular source of data. What is unique about Cerner’s EDH is that it brings together data from an almost unlimited number of sources, and that data can be used to build a far more complete picture of any patient, condition, or trend. As a result, “We’re able to achieve much better outcomes, both patient-related and financial, than we ever could by just looking at pieces of the puzzle individually,” said Brush. “It all comes down to bringing everything together and being able to extract value for any requirement. The enterprise data hub topology allows us to do exactly that.”

Each of the projects running on Cerner’s data hub delivers a unique value proposition. “There are obviously a lot of systemic problems relating to healthcare. We’re working to mitigate them by using data to build a complete picture of what’s going on, and then applying that knowledge to solve specific issues,” commented Brush. “For example, in addition to empowering our suite of offerings around population health management, the centralized hub now gives us the ability to predict the probability of a discharged patient being re-admitted for the same or a similar condition.”

Using the same strategy, Cerner can accurately determine the probability that a person has a bloodstream infection. “Our clients are reporting that the new system has actually saved hundreds of lives by being able to predict if a patient is septic more effectively than they could before,” Brush added.

Edwards summarized, “Our real aim is to get the technology out of the way so all that users see is the value that comes from their efforts. We really want the focus to be on the outcomes and results, not on what it takes to deliver them. The Cloudera platform is the technology that’s driving the value and it’s allowing us to build applications that help healthcare systems improve how they manage the chronic conditions of their populations. We’re now able to aggregate the information, stratify it, and offer the opportunity to look at this data in a way that has never been possible before.”

About Cloudera

Cloudera is revolutionizing enterprise data management by offering the first unified Platform for Big Data, an enterprise data hub built on Apache Hadoop. Cloudera offers enterprises one place to store, process and analyze all their data, empowering them to extend the value of existing investments while enabling fundamental new ways to derive value from their data. Only Cloudera offers everything needed on a journey to an enterprise data hub, including software for business critical data challenges such as storage, access, management, analysis, security and search. As the leading educator of Hadoop professionals, Cloudera has trained over 40,000 individuals worldwide. Over 1,400 partners and a seasoned professional services team help deliver greater time to value. Finally, only Cloudera provides proactive and predictive support to run an enterprise data hub with confidence. Leading organizations in every industry plus top public sector organizations globally run Cloudera in production. www.cloudera.com.

cloudera.com

1-888-789-1488 or 1-650-362-0488

Cloudera, Inc. 1001 Page Mill Road, Palo Alto, CA 94304, USA

© 2015 Cloudera, Inc. All rights reserved. Cloudera and the Cloudera logo are trademarks or registered trademarks of Cloudera Inc. in the USA and other countries. All other trademarks are the property of their respective companies. Information is subject to change without notice.

cloudera-casestudy-cerner-105

