National Children's Hospital

Applies Enriched Data Analysis to Improve Pediatric Care & Outcomes

Overview

One of the leading pediatric healthcare facilities in the United States knows children are not just small adults. They need specialized diagnosis, treatment, equipment, and support. Most importantly, they require doctors, nurses, and specialists who understand these differences.

This national children's hospital relies on a big data platform to better understand its patients, their conditions, and the quality of care they receive in support of its mission: to make kids better today and healthier tomorrow.

The Challenge

Hospitals' intensive care units (ICUs) have bedside monitors that continuously collect data streams on patient vitals such as respiration, heart rate, and blood pressure. The amount of information is vast, leading healthcare facilities to take hourly snapshots and then discard the data after several days.

This national children's hospital identified two projects that would benefit from longer-term data storage and analysis:

Bedside alarm study: "We have bedside monitors which connect back to central servers that store data for three days," explained the hospital's manager of business intelligence and data warehousing. "Any clinician can come in and analyze three days' worth of patient information. But after that, because of the sheer volume that is generated, data is overwritten. For any study to be meaningful, we needed to capture and retain all of the data on a minute by minute basis before we could consider analyzing it."

For instance, the organization wanted to know how environmental factors, such as ambient light and noise, impact the quality of care and patient outcomes of infants in the neonatal ICU; these patients are perhaps the most sensitive to factors of their environment, but least capable of communicating what they're experiencing.

Asthma research and discovery: Asthma is one of the chief reasons that patients come to the hospital's emergency rooms and are admitted, so understanding the precursors and potential triggers to asthmatic events is critical. The organization has access to 20 years of air quality data from the Environmental Protection Agency (EPA), but, as the BI manager explained, "The task of correlating this with our asthma research was beyond the scope of our existing enterprise data warehouse environment."

cloudera®

The organization's legacy infrastructure consists predominantly of relational and multi-dimensional databases. These systems serve critical purposes, but could not cost-effectively accommodate the volumes and types of data needed to support the complexity of analysis associated with these projects. The BI manager noted, "Our enterprise data warehouse was operating at capacity for both storage and processing."

The organization set out to find a big data solution that could be implemented quickly to relieve the EDW, without requiring significant resources or budget. Apache Hadoop surfaced as one solution, promising three key benefits:

1. A single, fault tolerant platform for storage and analysis
2. Low costs, both upfront and in scaling up
3. An active and growing ecosystem

The hospital turned to commercial vendors in search for an easy-to-use, pre-packaged distribution of Hadoop. CDH seemed to be the most pervasive, open source distribution—the organization even had an in-house resource already familiar with it—with good online documentation and an active user community. The hospital opted to build a workstation-based cluster as a proof of concept (POC) on CDH, and the test configuration was put together with just a few hundred dollars' investment.

Solution

The hospital's BI manager reported, "From start to finish, we were able to produce the output that our research collaborative needed in three weeks."

Based on this early success, the organization decided to move its big data platform into production and knew it would need vendor support to ensure system reliability and stability. The organization signed up for a Cloudera Enterprise subscription in September 2013.

The national children's hospital also adopted Cloudera Manager for system administration and management of the Hadoop cluster. "Cloudera Manager takes much of the day-to-day burden of managing the environment and puts it into an easy-to-use, friendly web format. You can quickly identify where it elevates your understanding of what is going on 'under the hood.'"

In the next several months, the organization accumulated nearly two terabytes (TB) of data on the Cloudera platform—including ICU vital sign sensor data, .wav audio files, and air quality sensor data—and volumes continue to grow by 50 gigabytes (GB) each week.

The bedside vitals data is ingested into Cloudera and made available for end-user analysis via Apache Hive or Cloudera Impala, which delivers interactive analysis to healthcare practitioners through its integration with Qlik. The organization uses Apache Sqoop to move output data into and out of the enterprise data warehouse on a daily basis.

The next phase of the bedside monitor project is to use Apache Flume and Apache Spark (incubating). Flume enables ingestion of HL7 interface data in near real time. Spark makes it easy to develop fast, unified Big Data applications combining batch, streaming, and interactive analytics on the data in CDH. The national children's hospital also plans to implement Kerberos for network authentication.

Impact: More Data, More Answers

In its initial POC with CDH, the hospital was assessing the neonatal intensive care unit for environmental factors. Ambient light and noise data was correlated with patient vitals collected from the bedside monitors. It was trying to understand how the environment of care affected the children and the potential impact on patient outcomes.

Prior to creating the Hadoop environment, full fidelity vital records were only available for three days. After three days, only hourly snapshots were available. "Everyone knew that this represented a miniscule percentage, but we had no way of handling a larger volume of data." With the processing and storage capacity of Hadoop, the hospital can collect

cloudera®

> "Improving patient care is always first and foremost in our minds. We are answering the questions that previously were unresolved, and we hope to see improved outcomes because of it."

Manager of business intelligence and data warehousing

continuous vital records. This supports analysis and correlation with medical records. The organization's patient care studies have been enriched by having access to the full data set, instead of relying on a small sample.

Now, the hospital is constantly collecting streams of data in granular detail without throwing any of it away. The organization will also soon bring historical electronic medical record (EMR) data from 1985 on into the Cloudera environment, which will enable even deeper, more comprehensive analysis.

Impact: Peak Performance

Although the national children's hospital stores most of its bedside monitor data and asthma data in Apache Hive tables, the day-to-day research and data validation is done using Impala, using Hue as the interface for both.

"Impala is incredibly fast," said the organization's BI manager. "When I do an advanced query, I don't have to wait for 30 minutes to see if my numbers look right. When viewed through Hue, Impala is a fantastic tool."

And, once Spark is in production, "We'll be able to push records to our data warehouse in near real time and alert clinicians to significant changes in bedside monitor data based on clinical criteria," he explained.

Impact: Delivering Value

This national children's hospital has already realized a return on its Cloudera investment. "Our entire Cloudera Enterprise platform, including hardware and license fees, cost less than three processor licenses for our data warehouse," said its BI manager.

And the organization expects the system to continue adding value through the new insights it delivers, for example, via tools like Cloudera Search, which the organization plans to implement before the end of 2014. By some estimates, "80% of all healthcare knowledge is contained in patient notes," explained the hospital's BI manager. "The ability to extract discreet data elements from these notes is critical. Traditional data warehouse technologies don't do a great job with that, but implementing Search will enable us to tap into previously inaccessible information."

He concluded, "Improving patient care is always first and foremost in our minds. We are answering the questions that previously were unresolved, and we hope to see improved outcomes because of it."

cloudera-blind-childrenhospital-casestudy-Q414-101