

## Streamlining Healthcare Connectivity with Big Data

### KEY HIGHLIGHTS

#### INDUSTRY

Healthcare

#### BUSINESS APPLICATIONS SUPPORTED

- > Complex processing & archiving of claims & remittance data
- > Storage & analysis of transactional log data

#### HADOOP IMPACT

- > Providers collect payment faster through expedited messaging with payers
- > Industry standard hardware is 10x less expensive than alternative technologies while enabling analytics on stored data
- > Simple deployment, minimal ongoing maintenance

### Company Overview

The connectivity and information technology subsidiary of a major pharmaceutical company was created to simplify how the business of healthcare is managed while making the delivery of care safer and more efficient. As more and more of the US' healthcare system goes electronic, this organization meets challenges and opportunities through an open network that supports future growth via interoperability among organizations, systems and solutions.

### Business Challenges Before Cloudera

With regulations such as the Health Insurance Portability and Accountability Act of 1996 (HIPAA), healthcare organizations are required to store healthcare data for extended periods of time. This health IT company instituted a policy of saving seven years' historical claims and remit data, but its in-house database systems had trouble meeting the data retention requirement while processing millions of claims every day.

A software engineer at the company explained, "All of our systems were maxed out. We were constantly having database issues. It was just too much data for what they were meant to handle. They were overworked and overloaded, and it started to cause problems with all of our real-time production processing."

Further, the organization sought a solution that would allow users to do more than just store data. The manager of software development at the company explained, "In today's data driven world, data really is this huge asset. We wondered, 'What framework, what platform will allow us to optimize the data that we have?'"

The team set out to find a new solution. "We could have gone the SAN route, but it's expensive and cumbersome," said the software engineer. They did some searching online and came across Hadoop, MongoDB, and Cassandra. "We analyzed them and came up with a prototype for each one. In the end, we decided Hadoop was what we wanted."

Initially the organization downloaded Hadoop from Apache and configured it to run on 10 Dell workstations that were already in house. Once the small Hadoop cluster showed its functionality and demonstrated value, the team decided to make a commitment to the platform, but would need support to do so. When evaluating various Hadoop distributions and management vendors, they recognized that Cloudera was different: its Hadoop distribution — CDH — is 100% Apache open source. This allows Cloudera customers to benefit from rapid innovations in the open source community while also taking advantage of enterprise-grade support and management tools offered with the Cloudera Enterprise subscription.

**KEY HIGHLIGHTS****TECHNOLOGIES IN USE**

- > Hadoop Platform: Cloudera Enterprise
- > Hadoop Components: Flume, Sqoop, MapReduce, HDFS
- > Data Warehouses: Oracle, IBM Netezza

**BIG DATA SCALE**

- > 30 TB across 10 CDH nodes in 4 months
- > Archiving 7+ years' historical data
- > 24x7 feeds from source systems; processing ~1 TB per day

**ADVICE TO NEW HADOOP USERS**

- > Step back and evaluate your architecture to take advantage of Hadoop's capabilities
- > Take Hadoop training

**"If you look at the margin that the average hospital has, it's between 2–3%. So their cash flow is very tight. Anything you can do to reduce the time to get paid is very valuable to a healthcare provider."**

MANAGER, SOFTWARE DEVELOPMENT

**Use Case**

When deciding to deploy CDH, the team set out to identify applications that were already seeing performance issues in production. "One of the big advantages of Hadoop has been to be able to segregate big data from transactional processing data and allow smoother processing of information. Basically, it allows us to offload a lot of stress from the database," said the company's manager of software development.

They quickly identified two areas that were a strong fit for Hadoop:

- Archiving seven years' claims and remit data, which requires complex processing to get into a normalized format
- Logging terabytes of data generated from transactional systems daily, and storing them in CDH for analytical purposes

Today the health IT organization uses Flume to move data from its source systems into the CDH cluster on a 24x7 basis. The company loads data from CDH to an Oracle online transaction processing (OLTP) database for billing purposes. This load runs once or twice each day via Sqoop.

**Impact: Helping Providers Collect Payment Faster through Operational Efficiencies**

"If you look at the margin that the average hospital has, it's between 2-3%," stated the manager of software development. "So their cash flow is very tight. Anything you can do to reduce the time to get paid is very valuable to a healthcare provider."

Since deploying Cloudera Enterprise, the organization has reduced the time it takes for healthcare providers to get paid by streamlining their transfer of messages to payers. The ability to expedite this process is especially valuable when regulatory changes come into play, such as the recent conversion from HIPAA 4010 to HIPAA 5010.

"We assist with the conversion and processing of these messages," said the company's manager of software development. "For example, 4010 messages came in and we'd convert them to 5010 to allow seamless processing. The providers didn't have to upgrade any of their systems when the regulations went into effect. We gave them a bit of a buffer to implement changes. And since we do a lot of electronic processing, we can do basic sanity checks on the messages as they come in and let providers know what adjustments need to be made in order to get paid faster."

**Impact: Low Cost + Greater Analytic Flexibility**

Because Hadoop uses industry standard hardware, the cost per terabyte of storage is, on average, 10x cheaper than a traditional relational data warehouse system. "One of my pet peeves is: you buy a machine, you buy SAN storage, and then you have to buy licensing for the storage in addition to the storage itself," explained the manager of software development. "You have to buy licensing for the blades, and it just becomes an untenable situation. With Hadoop you buy commodity hardware and you're good to go. In addition to the storage, you get a bigger bang for your buck because it gives you the ability to run analytics on the combined compute and storage. The solutions that we had in place previously really didn't allow for that. Even if the costs were equivalent, the benefit you get from storing data on a Hadoop type solution is far greater than what you'd get from storing it in a database."

### Impact: Simple Deployment & Administration

After deciding on the Cloudera solution, “the deployment process into production with Hadoop was actually quite easy,” said a software engineer at the company. “Cloudera Manager really helped us a lot. It’s as easy as just clicking a few buttons, and you’re up and running. It’s really simple. And the support staff at Cloudera have been great. They really helped us out with a couple of issues we had along the way.”

Several employees enrolled in Cloudera University training as well, which was “very beneficial” according to one software engineer. And with Cloudera Manager, the team spends very little time managing the cluster.

Further, this health IT organization appreciates the proactive customer support offered by Cloudera Enterprise. “We ask questions and we get them answered very quickly,” commented their manager of software development. “Not only do the Cloudera Support folks answer the question, they come back and say, ‘Do you have any other questions? Is there anything else we can help you with?’ It’s very different. The people that are on Cloudera’s Support team — you can definitely tell they are Hadoop Committers. Not only will they find you the answer but they can tell you, ‘This may not be the best practice, you may want to change the way you’re doing your development to take advantage of other features.’ The Cloudera Support organization is world class.”