Explorys Medical: Improving Healthcare Quality & Costs Using a Big Data Platform

Company Overview
In 2009, Explorys’ co-founders – Stephen McHale, CEO; Charlie Lougheed, president & CTO; and Doug Meil, chief software architect – set forward on a mission to integrate clinical, financial and operational data in order to improve delivery of care, resulting in better health outcomes and reduced costs associated with delivering medical services. Explorys has built a cloud-based Enterprise Performance Management (EPM) computing platform for healthcare providers that enables real-time exploration, performance and predictive analytics of clinical data. Several of the United States’ largest integrated delivery networks have deployed Explorys’ platform including Cleveland Clinic, MedStar, University Hospitals, St. Joseph Health System, Catholic Health Partners and Summa Health System.

Use Case
Explorys “bet the company on Hadoop in 2009,” stated Meil. Explorys’ Big Data platform, known as DataGrid, is based on Cloudera Enterprise and offers three software-as-a-service (SaaS) applications:

- EPM: Explore provides sub-second ad-hoc search across populations, providers and care venues.
- EPM: Measure is an integrated application and framework for constructing and viewing key performance metrics across providers, groups, care venues and locations.
- EPM: Engage is an integrated application and framework for coordinating rules-driven registries, prioritized patient and provider outreach, and messaging.

Explorys uses MapReduce to process, organize and curate large batches of incoming data very quickly into the Hadoop Distributed File System (HDFS) and HBase. Data sources include HL7, claims data and third party feeds from virtual health records (VHRs) such as Epic, Eclipsys, Amalga, McKesson and Cerner. That data is materialized into structures that can be rapidly accessed and analyzed via off-the-shelf tools or custom-built applications. Those applications then generate data back into HBase for other processing needs. Analytics powered by Hadoop facilitate clinical quality measure generation, measure calculations for registries, proactive care management and other critical tasks.

For example, Explorys uses Hadoop-powered analytics for purposes similar to the following scenario: In order to better serve its community, a county hospital might want to explore why people utilize the emergency room (ER) rather than going to a primary physician for care. This practice is not beneficial for either the patient or hospital because it is expensive for the hospital and the patient doesn’t get the continuity of care that a primary physician would provide. Additionally, it is important to analyze a bigger problem than just the ER record system. With Hadoop, Explorys can analyze what factors are demographically different about a given population, where patients live and whether or not care is available in their neighborhoods. Explorys can run these analytics daily in order to take immediate action, such as sending a letter to a patient the day after she or he has visited the ER that includes information about local healthcare providers and instructions to help prevent readmittance.

Business Challenges Before Hadoop
“With a clinical enterprise performance management system built on a relational data warehouse, if healthcare practitioners want to understand something specific about a population or segment of data, they have to go to their IT departments and then wait days or weeks to get that information back. We wanted to provide a platform that would give them an answer as fast as if they were searching on Google,” said McHale.

Rather than managing clinical, financial and operational data in three data silos, Explorys sought to bring all of that data together. Lougheed explained, “It’s about merging the three elements and telling a story about how an organization is doing, because ultimately what we want to do is improve healthcare and do it at a lower cost. With over 17% of the nation’s GDP being spent on healthcare services, we’ve got to find a better way to deliver healthcare at a cheaper cost point.”

Business Applications Supported
- Complex data processing & curation of clinical, financial & operational data
- Clinical quality measure analytics
- Proactive care management analytics

Big Data Scale
- Unlimited data retention; never purge patient data
- Ability to curate 40+ billion clinical & operational data points over 13 million individuals
- 35 TB multi-structured data; growing to 70 TB by 2013

Hadoop Impact
- Time to insight reduced from days to minutes
- Increased business agility
- Reduced TCO from open-source model

Advice to New Hadoop Users
- Leverage lessons learned from early adopters and Apache community
“We had lots of experience working with traditional database platforms in our careers in banking and telecom, and we just didn’t believe they were going to scale economically or from a performance standpoint,” added McHale.

The variety of data would also present a challenge. “Electronic Health Record (EHR) platforms are just now becoming prevalent across the medical landscape,” noted Lougheed. “There are more and more devices that generate massive amounts of data. Patients are providing data and feedback on how they’re doing. They have devices in the home that provide data. There’s a ton of data variety coming in and it’s more than the healthcare space can really handle.”

Explorys needed to find a cost-efficient technology that would help the company address these big data challenges. Hadoop met both of these criteria, and Cloudera stood out to Explorys as the most credible company delivering an enterprise-ready Hadoop solution. McHale, Lougheed and Meil attended the first Hadoop World conference hosted by Cloudera in 2009, and were sold both on the value of Hadoop and on Cloudera’s “ability to deliver.” They also appreciated Cloudera’s contingent of on-staff Hadoop committers, signifying Cloudera’s ability to support production deployments and drive the direction of the technology.

After deploying the DataGrid platform on CDH, Explorys decided to take advantage of the Cloudera Enterprise subscription offering that bundles Cloudera Support with Cloudera Manager, a software application that delivers visibility into and across Hadoop clusters for streamlined management, monitoring and provisioning. Cloudera Enterprise gave Explorys the peace of mind to continue operating its mission critical Hadoop environment with a lean staff that could focus on the core competencies of the business.

**Hadoop Impact**

Explorys relies on CDH to provide the flexibility, scalability and speed necessary to answer complex questions on the fly. The window for analysis shrunk from 30 days using traditional technologies to seconds or minutes with Explorys’ CDH-powered system. Meil explained, “We didn’t invent clinical quality measures, but the ability to generate those measures on a rapid basis, support hundreds of them, implement complex attribution logic and manifest that with a slick user interface on top – that’s revolutionary. Cloudera provides the technology that allows us to address traditional challenges in medicine and in operational reporting with a radical new approach.”

With traditional systems, it is cumbersome to make changes to the data and then reprocess it. “The ability to have that data ready within a day is game changing,” noted McHale. “Without Hadoop, it would be extremely challenging to do and certainly not cost effective.”

Because Hadoop uses commodity hardware, the cost per terabyte of storage is, on average, 10x cheaper than a traditional relational data warehouse system. “We’d be spending literally millions more dollars than we are today on relational database technologies and licensing,” said Lougheed. “Those technologies are important, but they’re important in their place. For big data analytics, storage and processing, Hadoop is a perfect solution. It has brought opportunities for us to re-channel those funds, that capital, in directions of being more innovative, bringing more products to bear, and ultimately hiring more people for the company as opposed to buying licenses.”

On Cloudera Enterprise, Meil noted, “We have a non-trivial system operations team and any help we can get in terms of operational support and tooling for the management of our cluster is critical. I definitely see the value in Cloudera Enterprise.”

Lougheed summarized, “We were forced to think about a less expensive, more efficient technology from the beginning, and in hindsight, I’m glad that we were. For us – needing storage, high capacity index and analytics – Hadoop really made a lot of sense. And from a partner perspective, Cloudera is one of the most influential that we’ve had throughout the years.”

**About Cloudera**

Cloudera, the leader in Apache Hadoop-based software and services, enables data driven enterprises to easily derive business value from all their structured and unstructured data. Cloudera’s Distribution including Apache Hadoop (CDH), available to download for free at www.cloudera.com/downloads, is the most comprehensive, tested, stable and widely deployed distribution of Hadoop in commercial and non-commercial environments. For the fastest path to reliably using this completely open source technology in production for Big Data analytics and answering previously un-addressable big questions, organizations can subscribe to Cloudera Enterprise, comprised of Cloudera Manager software and Cloudera Support. Cloudera also offers training and certification on Apache technologies, as well as consulting services. As the top contributor to the Apache open source community and with tens of thousands of nodes under management across customers in financial services, government, telecommunications, media, web, advertising, retail, energy, bioinformatics, pharma/healthcare, university research, oil and gas and gaming, Cloudera’s depth of experience and commitment to sharing expertise are unrivaled.