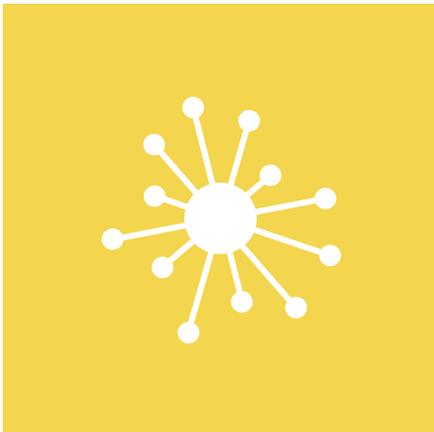




Intel and Cloudera Use Predictive Analytics to Help a Large Hospital Group Reduce Readmission Rates

Intel and Cloudera ingest socioeconomic data and electronic medical records (EMR) to identify high-risk patients and reduce readmission rates.



Why Intel and Cloudera

Intel and Cloudera take the guesswork out of Hadoop. Using a unique collaborative approach, we deliver the best performance, security, and quality distribution, built on open standards. Working with more vendors across the ecosystem, only a solution built on CDH can ensure freedom from lock-in, enabling you to build a robust big data solution to meet the needs of your business today and into the future.

- Uniquely aligned product roadmaps for software and hardware to drive innovation faster, providing more industry firsts than any other Hadoop alternative.
- Deep partnerships with virtually every provider in the data center, streamlining the process for building Big Data solutions.
- Proven track records of identifying the driving industry standards, so you don't run the risk of stranding yourself on an island.

For a large Hospital Group hoping to reduce readmission rates by 3% annually, predictive analytics with Cloudera distribution of Hadoop (CDH) is helping doctors pinpoint patients with high readmission risk. Hospital staff then administer additional medical care to these patients and thereby reduce readmission rates.

Results

The Hospital Group generates an individualized prediction of a patient's readmission rate at the time of diagnosis.

Using the derived predictions from the analysis, the Hospital Group reaps the following annual savings:

- Reduces 6,000 occurrences of patient readmission.
- Avoids \$4 million in potential Medicare penalties.
- Saves approximately \$72 million in medical service costs.
- Utilizes resources more efficiently by providing extra care to high-risk patients.
- Improves hospital rating based on lower readmission rate and increased patient satisfaction.

Business drivers

Recent changes in federal legislation have made hospitals restructure the way they manage patients to save money and avoid government penalties. Section 3025 of the Affordable Care Act added section 1886(q) to the Social Security Act, which took effect October 1, 2012. It established the Hospital Readmissions Reduction Program, which requires the Centers for Medicare & Medicaid Services (CMS)—a federal agency whose mission is to ensure effective healthcare coverage and to promote quality care for Americans—to reduce payments to hospitals with excess readmissions.

Consequently, hospitals have been seeking ways to reduce readmission rates across the board. Doing so would not only reduce unnecessary costs, it would help hospitals avoid CMS-levied payment penalties.

The Hospital Readmissions Program Accuracy and Accountability Act¹ requires CMS to account for patient socioeconomic status when calculating risk-adjusted readmission penalties. Holding all other factors constant, socioeconomic conditions—such as poverty, low literacy, limited English proficiency, minimal social support, poor living conditions, and limited community resources—likely have direct and significant impacts on avoidable hospital readmissions. Adjusting for

1. https://www.aamc.org/download/382510/data/the_hospitalreadmissionaccountabilityandimprovementact.pdf

these factors would improve accountability and quality of care.

Per capita healthcare costs in the US are the highest in the world and have trended upward for several decades. *Figure 1* shows \$25 billion spent annually in hospital readmissions alone. Reducing the number of unnecessary readmissions by even a few percent could create huge savings.

Figure 1 Healthcare's wasted dollars. The graph below shows that hospital readmissions cost \$25 billion annually. Source: Price Waterhouse Coopers Health Research Institute (2008)



A readmission is defined as a hospitalization that occurs approximately 30 days after a previous hospital stay. Readmissions are often the result of a patient's initial problem not being resolved. They can also be caused by a patient's mismanagement of the original condition, misunderstanding how to manage the condition, or lack of access to additional medical services or medications.²

2. Silow-Carroll, Sharon; Edwards, Jennifer N.; and Lashbrook, Aimee; Health management associates. "Reducing hospital readmission: lessons from top-performing hospitals" synthesis report. April 2011. http://www.commonwealthfund.org/~media/Files/Publications/Case%20Study/2011/Apr/1473_SilowCarroll_readmissions_synthesis_web_version.pdf

Solution details

According to a recent study, a patient's socioeconomic conditions can have direct and significant impacts on avoidable extended hospital stays.³ Adjusting for these factors improves accountability and quality of care.

Socioeconomic data along with electronic medical records (EMR) provide a patient's history and living standard for the model. This data is collected in an enterprise data hub (EDH) using Cloudera. The EDH facilitates data loading, cleansing, and association or linking between different datasets, such as EMR and socioeconomic data, for every patient.

Using the "Random Forests" algorithm, Intel helped the Hospital Group build models based on this linked dataset. These models predict a readmission risk score during the admission process for each patient based on his/her EMR and socioeconomic data. The Hospital Group assigns each patient who is

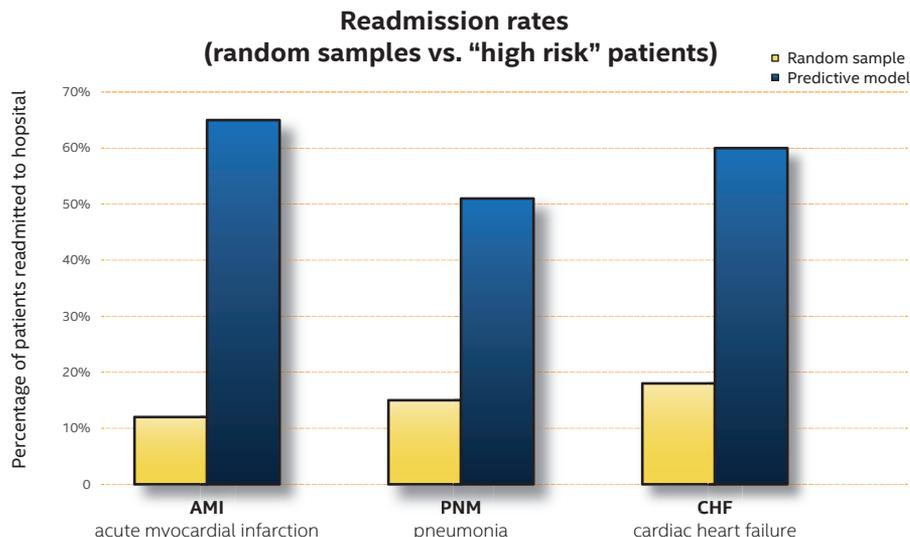
3. Rice, Sabriya. "Bill would adjust readmissions for socioeconomic factors." June 20, 2014. <http://www.modernhealthcare.com/article/20140620/NEWS/306209964/bill-would-adjust-readmissions-for-socio-economic-factors>

admitted a risk score, thus creating a high-risk bin, which consists of the top 5% of patients by risk score. With an accurate prediction, hospital administrators can suggest a special care plan for patients identified as high risk.

Figure 2 shows the Hospital Group's readmission rates for patients diagnosed with acute myocardial infarction (AMI), cardiac heart failure (CHF), and pneumonia (PNM). Using random samples of these patients does not yield high accuracy in predicting which of these patients would be readmitted—less than 20%. However, those patients identified by the Intel/Cloudera predictive models as "high risk" proved far more likely to be readmitted for each condition.

By focusing on patients in the high-risk bin, the Hospital Group can target patients with a higher likelihood of readmission for additional care during their first visit, and thus reduce the readmission rate of these patients. As a side benefit, this frees up resources they could use to help an additional 300 to 500% more patients.

Figure 2 Readmission rates. A random sample of patients (yellow bar) correctly identifies less than 20% of patients with acute myocardial infarction (AMI), cardiac heart failure (CHF), and pneumonia (PNM) that are actually readmitted to the hospital. The predictive model (blue bar) ranks patients and identifies the top 5% of patients with the highest risk score. By checking for socioeconomic high-risk factors, the predictive model correctly identifies patients that are actually readmitted with an accuracy of 50 to 65%.



The requirements necessary for the data storage solution include analysis of a patient's readmission prediction, scope of the patient's medical condition, accuracy of the prediction, growing population data, easy ingestion of diverse data, fault tolerance, low cost, and security of the data.

Working with Intel, the Hospital Group selected several data sources for the predictive model, including EMR from a relational model and additional socioeconomic data such as housing prices and availability of healthcare within the immediate area of each hospital.

Cloudera Enterprise

The Hospital Group selected Cloudera to assist them with their data analysis for better predictability to lower readmission rates for the following reasons:

- **Security.** The Hospital Group values security and protected health information (PHI) compliance. Cloudera offers the most secure and enterprise-ready Hadoop distribution. An important factor is CDH's support for transparent encryption in Hadoop Distributed File System (HDFS), which allows security of personally identifiable information (PII) data that is encrypted all the way to the client, securing data both when it is at rest on a disk and in transit. Furthermore, Cloudera Navigator Key Trustee makes encryption key management extremely easy.
- **Flexibility.** Cloudera is flexible and has large dataset storage. A schema-on-read architecture for data ingestion is the only way to support the objectives, which eliminates traditional databases as a storage solution. In open source technologies, many data storage solutions offer schema on read and an ability to store a variety of data formats and scalability.

Furthermore, the Hadoop ecosystem offers tools to ingest data using Sqoop and Flume, data cleanse and prep tools like Pig, and analytical libraries like Mahout and R.

- **Maturity.** The Hadoop ecosystem is mature and broad enough to accommodate data warehousing, analytics, and NoSQL, along with the traditional focus on storage.
- **Support.** Cloudera Enterprise provides an experienced technical support team. Cloudera's contribution to the Hadoop ecosystem and open source community demonstrates that Cloudera has the experience to assist during the life cycle of application development. Cloudera Support offers the industry's only predictive and proactive support capabilities.

Summary

By using predictive analytics with Cloudera, the Hospital Group takes advantage of more unconventional data sources to produce more accurate readmission predictions. Cloudera has the power to ingest unrelated, unstructured, and semi-structured data sources, which the Hospital Group uses to enrich existing medical data.

The readmission predictive models Intel helped create for the Hospital Group are very successful at identifying patients who are a high risk for readmission after an initial hospital stay. With these more accurate predictions, hospital administrators can now suggest special care plans for high-risk patients.

Let us help your business too.

Spotlight on 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, access, process, secure, 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.

Cloudera's open source Big Data platform is the most widely adopted in the world, and Cloudera is the most prolific contributor to the open source Hadoop ecosystem. As the leading educator of Hadoop professionals, Cloudera has trained over 40,000 individuals worldwide. Over 1,600 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.

For more information, visit www.cloudera.com.

Meeting your needs

We look forward to meeting with you to define your requirements and meet your objectives.

- **Accelerate time to value:** Achieve real-time cost savings, respond to market trends, and drive innovation.
- **Secure Big Data:** Deploy a sustainable Big Data program that doesn't put your organization, or you, at risk.
- **Maintain control:** Work with a partner who educates your team so you become self-sufficient.
- **Increase business potential:** Create and execute a plan that helps you adapt now, and in the future.

Contact us

Contact your sales rep or e-mail us.
Intel.com/bigdata/services

Hadoop sizing guide

		Cluster size		
		Small	Medium	Large
CPU		Intel® Xeon® Processor E5 v3		
Storage (TB)		<72 TB	72 to 570 TB	>570 TB
Node count	Master	2 to 3	4 to 7	≥8
	Slaves	<12	12 to 95	≥ 96
Memory (GB)	Master	64 GB	128 GB	≥256 GB
	Slaves	48 GB	96 GB	≥128 GB
Network		1 Gbps	10 Gbps	10 Gbps

Hardware configuration is highly dependent on workload. A high storage density cluster may be configured with a 4 TB JBOD hard disk, while a compute intensive cluster may be configured with a higher memory configuration.



Intel does not control or audit the design or implementation of third-party benchmark data or websites referenced in this document. Intel encourages all of its customers to visit the referenced websites or others where similar performance benchmark data is reported and confirm whether the referenced benchmark data is accurate and reflects performance of systems available for purchase.

Copyright © 2015 Intel Corporation. All rights reserved.
 Intel and the Intel logo are trademarks or registered trademarks of Intel Corporation in the United States and other countries.