Intel and Cloudera Help a Large Hospital Group Allocate Resources by Predicting Patient Length-of-Stay

With a more accurate length-of-stay prediction model, hospitals can allocate personnel and resources more efficiently, thus reducing operational costs.

Why Intel and Cloudera

Intel and Cloudera take the guesswork out of Hadoop. Using a unique collaborative approach, we delivery excellent 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.

Reducing patient length-of-stay allows a large Hospital Group to save $15 million in service operational costs while increasing facility utilization by 5% annually. Predictive analytics with Cloudera distribution of Hadoop (CDH) is helping hospital administrators accurately predict discharge dates. By taking advantage of more unconventional data sources, administrators reduce the variance in patient occupancy and thereby reduce facility costs due to lengthy patient stays.

Results

The Hospital Group generates an individualized prediction of each patient's length of stay at the time of hospital admission. Using the more accurate predictions from the Cloudera-based analysis, the Hospital Group reaps the following annual benefits:

- Improves scheduling for approximately 30,000 patients (15 to 20% of all patients).
- Increases facility utilization by 5%, which allows hospitals to potentially serve an additional 10,000 patients annually.
- Saves $120 million in annual costs (about $12,000 per patient).
- Saves about $15 million in medical service costs (about $500/patient for last day of stay).

Business drivers

Recent legislation for Medicare reimbursement for hospital care, specifically the prospective payment system (PPS), encourages shorter hospital stays for patients by standardizing payments for procedures performed, regardless of the number of days a patient spends in the hospital.

This standardization in Medicare reimbursement has forced hospitals to use resources—such as hospital beds—more efficiently to maximize Medicare revenue. To improve bed utilization, an early prediction of a patient's discharge is required. This prediction also determines a patient's duration of stay at the time of admission.

Forecasting the length of stay helps the Hospital Group:

- Identify patients who might be at risk for an extended length of stay.
- Alter a patient's treatment plan from the moment of admission based on the risk analysis to reduce the hospital stay.
- Improve patient satisfaction by meeting expectations set regarding the hospital stay during admission.
- Improve resource management by efficiently allocating resources and accommodating more patients with the same resources.

Solution details

According to a recent study, a patient's socioeconomic conditions can have direct and significant impacts on avoidable extended hospital stays.[1] 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 patient's length of stay during the admission process, based on the patient's EMR and socioeconomic data, and on the condition being treated. With an accurate prediction of a patient's discharge date, hospital administrators can better allocate resources with higher confidence.

Because traditional methods of predicting length-of-stay could not address these socioeconomic factors, hospital administrators—using an average length of stay (ALOS) and a two-day margin for error—were only hitting the mark about 60% of the time. The yellow bars in *Figure 1*, for example, show the percentages of patients whose actual hospital stay fell within two days of the traditional predicted length of stay.

1. 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*

Leveraging the insights gleaned from nontraditional data sources and socioeconomic data for patients within each hospital's vicinity, Intel helped the Hospital Group create a model that identified high-risk patients outside that 60% average range, tagging factors that would indicate longer stays and skew the accuracy of the "typical average" predictions.

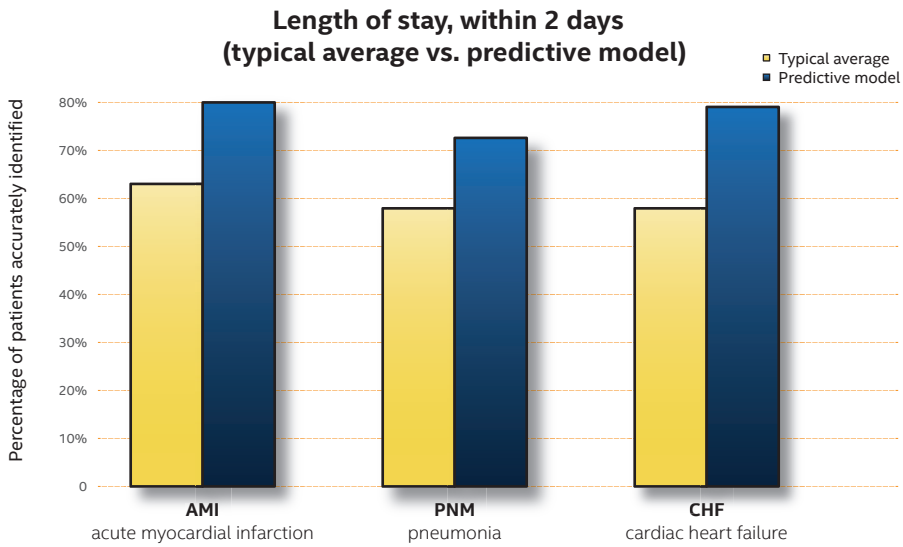
Using the predictive model, the Hospital Group can improve predictions of their collective patients' lengths of stay by 25 to 40% when compared to the typical average for each condition. The Hospital Group's predictions within the two-day window now approach 80% success.

This improvement in length-of-stay prediction accuracy in itself does not *reduce* the length of stay. (Although the modeling does allow data architects to fine-tune their analyses to pinpoint root causes that might be responsible for the longer stay, and address those issues to reduce overall length of stay for individuals in the high-risk pool.) It does, however, provide administrators with a more accurate measure of the resources they will need at any given time to serve the patients in the queue, which aids in planning and allocation.

The requirements necessary for the data storage solution include analysis of a patient's length-of-stay 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. The Hospital Group also wanted to be cautious about overzealous length-of-stay reduction policies that might contribute adversely to higher readmission rates—another area of concern for most hospitals—so it was important for the modeling to account for both goals.

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.

Figure 1 **Resource allocation based on accurate length-of-stay predictions.** Accurately predicting a patient's length of stay in the hospital has consequential benefits to hospital operations. Because administrators can better allocate beds, staff, supplies, and other resources, they can plan more accurately. Using the models Intel helped develop, the Hospital Group was able to improve their length-of-stay prediction accuracy by an average of more than 30% for the three conditions listed below.

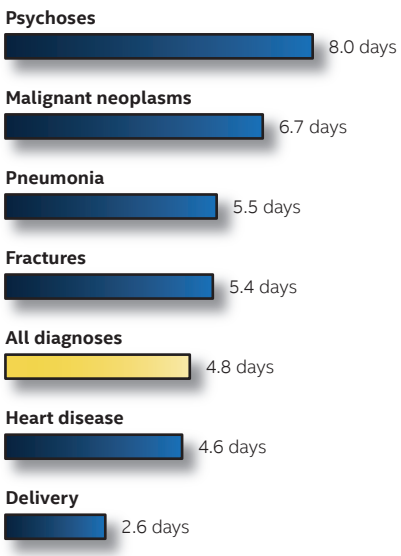


Cloudera Enterprise

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 a highly secure enterprise-ready Hadoop distribution. An important factor is CDH's support for transparent encryption in Hadoop Distributed File System (HDFS), which helps secure personally identifiable information (PII) data that is encrypted all the way to the client, both 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 predictive and proactive support capabilities.

Figure 2 **Hospital lengths of stay per condition.** The graph below shows the average length of stay (ALOS) for hospitals in the US, for diagnostic categories with at least 1 million hospital discharges. *Source: DeFrances, Hall, Podgornik. 2003. National Hospital Discharge Survey. Centers for Disease Control and Prevention.*



Summary

By using predictive analytics with Cloudera, the Hospital Group takes advantage of more unconventional data sources to produce more accurate length-of-stay 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 length-of-stay predictive models Intel helped create for the Hospital Group are very successful at predicting patients' hospital stays at the time of their admission. With these more accurate predictions, hospital administrators can now plan and staff more efficiently.

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**.

cloudera®

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.

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.

Contact us

Contact your sales rep or e-mail us.

Intel.com/bigdata/services

332574-003