Intel and Cloudera* Generate Positive ROI for an ISV and its Healthcare Provider Customers

Intel and Cloudera help an independent software vendor upgrade its proprietary healthcare analysis architecture and improve its customers' return on investment.

Why Intel and Cloudera

Intel and Cloudera take the guesswork out of Apache* Hadoop*. Using a unique collaborative approach, we deliver excellent performance, security, and quality distribution, built on open standards. Because we work with hundreds of vendors across the ecosystem, a solution built on Cloudera Enterprise 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 many industry firsts with Hadoop*.
- 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.

An independent software vendor (ISV) that provides diagnostic insights to healthcare organizations through Big Data analytics asks Intel to rebuild their SaaS solution on Cloudera Enterprise* to leverage Hadoop*'s ability to ingest structured and unstructured data to provide new insights in near-real-time.

Results

After Intel installs Cloudera Enterprise in the Company's datacenter, the Company realizes the following benefits:

- The Company's analytics generate positive ROI for all of their customers—healthcare providers and individual payers.
- The Company transcends the limitations of its traditional RDBMS for search and machine learning applications.
- The Cloudera Enterprise* technology stack creates a data processing and cognitive computing pipeline for diverse data sources with multiple connectors and processing requirements.

Business drivers

Despite the trillions of dollars spent each year on healthcare and the creation of about a billion clinical care documents annually in the US, most healthcare providers still have difficulty accessing their patients' complete medical histories. This is due, in part, because nearly 80 percent of this data is unstructured, which makes it a challenge just to access the data let alone make it usable. Even the structured data in healthcare is often "noisy" and biased, which has driven healthcare organizations to third-parties, such as the Company, to help them better understand their patients.

The Company's healthcare provider customers have access to accurate, reliable, and comprehensive data on their patients, which is required to optimize the healthcare delivery system and manage patients and payment, but the data is diverse and is stored in diverse locations.

Using their customers' patients' healthcare records, the Company runs proprietary data extraction tools and machine learning algorithms to provide better healthcare and simple, straightforward workflows that reduce errors and improve efficiency for their customers. But the Company was only scratching the surface of the available data. They wanted to funnel very large data processing loads of structured and unstructured data—including clinical notes, records, and images—through a dynamically scalable pipeline.

To compute the "true state of a patient," the Company would need to process textual clinical data, image data, and structured clinical data, including annotation, semantic mapping of structured data, and machine learning.

Throughput for such compute-intensive processing of scanned documents and text drove the requirement for a scalable architecture. Traditional RDBMS architectures cannot scale with this type of application; Big Data on a Cloudera technology stack, however, can. The Company knew a parallel computing technology such as Hadoop* was the best candidate for this job.

Solution details

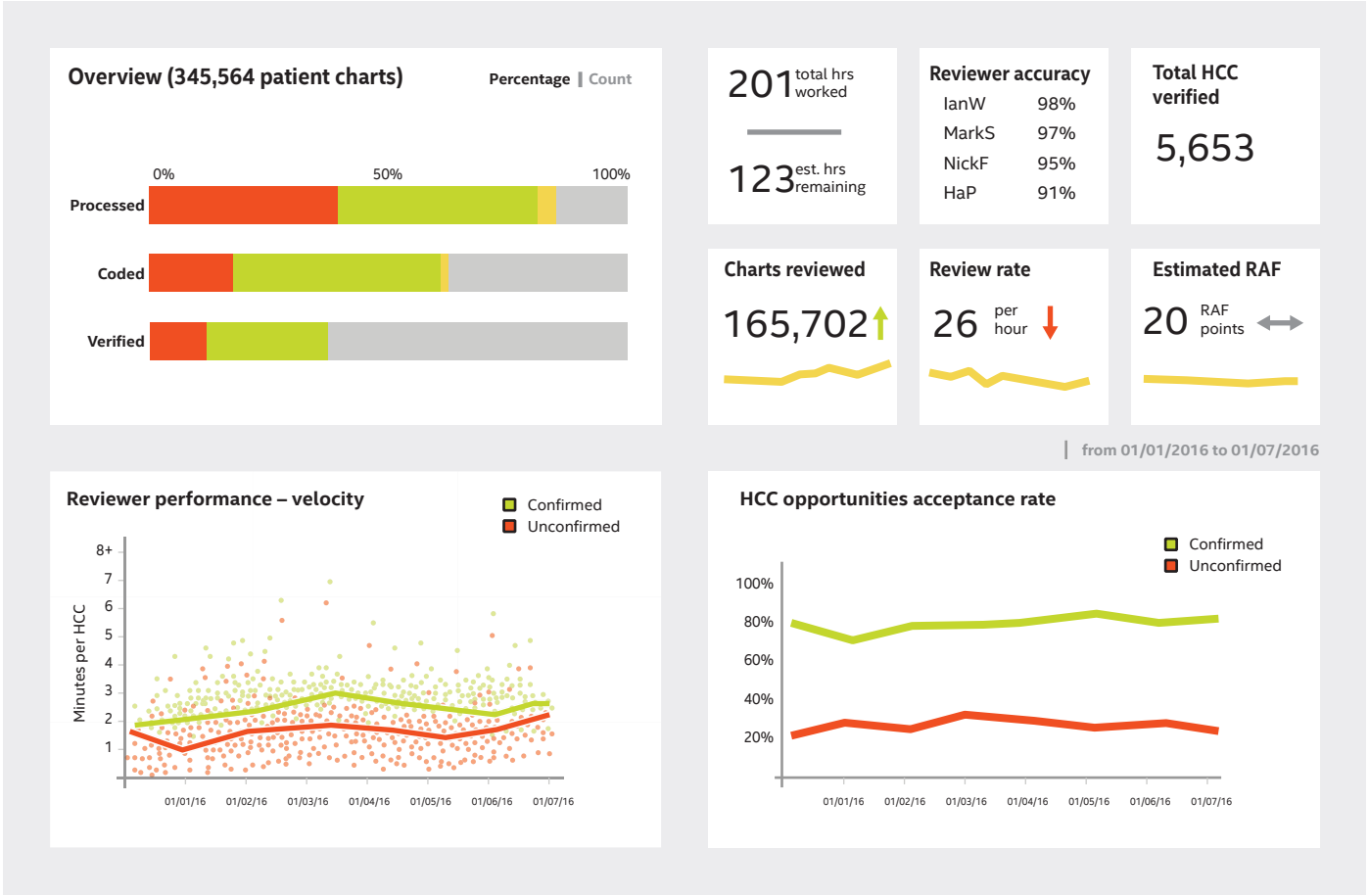
The Company's patented cognitive computing platform delivers advanced insights to healthcare providers, using data extraction tools, analytics, semantic concept extraction, computer learning infrastructure, flexible microservice APIs, and optimized expert review workflows in a reiterative process:

- **Map and ingest.** The Company uses tools such as Informatica*, Altova MapForce*, and their own custom innovations to incorporate a range of data types and formats.
- **Extract, transform, and load.** The ETL layer uses Cloudera to extract usable information from ingested data and writes results to a repository built on Apache* Cassandra*, a distributed database management system.
- **Analyze intelligently.** The intelligence layer applies the Company's sophisticated algorithms to the extracted data, looking for previously uncoded conditions.
- **Convert transparently.** The Company converts structured and unstructured data into practical clinical and business value.
- **Present simplified workflows and dashboards.** The platform presents dashboard graphics and concise, summarized results as workflows that human coders use to assess the validity of newly identified condition codes. User responses are folded back into the intelligence layer, "training" the system and enabling the algorithms to learn.

The Company's healthcare provider customers require personalized output for their patients as well as overall dashboard analytics of how they are performing in terms of processing patients (*Figure 1*).

Operating on top of Cloudera Enterprise now allows the Company to leverage the information from all of their customers' patient histories for analytical purposes, a foundation of more than 2 million patient-years of data. Their earlier architecture could not scale to such a massive pipeline of data.

Figure 1 **Sample dashboard.** Comprised of the latest natural language processing (NLP) and machine learning technologies, the Company's analytics engine uses the patient model to profile healthcare, determine risk and quality, assess care decisions, and quantify performance. This screenshot shows one output screen for one of the Company's healthcare provider customers.



Cloudera Enterprise

Centralized data management presents tremendous opportunity for enterprises looking to unlock the value of their data, but it also represents a challenge: namely, how to protect the sensitive business data and associated artifacts in a manner that satisfies compliance and meets internal and customer-driven security mandates.

In terms of patient confidentiality, for example, Cloudera's Navigator Encrypt* provides high-performance transparent encryption for compliance regulations such as HIPAA. Navigator Encrypt only allows authorized database accounts with assigned rights, connecting from applications on approved network clients, to access cardholder data stored on a server. Operating system users without access to Navigator Encrypt keys cannot read the encrypted data. So Cloudera Enterprise protects the integrity of patient data.

In addition to enhanced security features, Cloudera addressed all of the Company's needs and requirements for a Big Data solution:

- A scalable compute infrastructure that can respond to variable workloads for customer data ingest.
- A flexible, scalable pipeline to run ETL and analytics on diverse datasets.
- High performance No-SQL data storage model working in conjunction with the compute infrastructure.
- A platform that is consistent with other open source components, such as Apache Cassandra*.

Summary

Big Data analytics will continue to permeate the healthcare industry, and an increasing number of physicians will gain access to powerful healthcare datasets they never had access to before. Data analysis that enables real-time, data-driven feedback on clinical decisions will become commonplace for physicians, and patient care will continue to improve and benefit from the insights uncovered by Big Data analytics. Using clinical data, text, and scanned documents will help more and more healthcare providers compute the true state of their patients.

The Company had already developed a successful software-as-a-service platform for their healthcare provider customers, built upon the insights from analyzing more than 500 million patient documents over several years. But their original architecture could not take advantage of the wealth of new and legacy patient data available to them. The Company currently processes ongoing and historical clinical data for millions of patients—including structured data, text, and scanned documents.

They needed Intel's help maximizing their work on a massive Big Data scale, and they now have the power to unlock untold value from existing healthcare data and drive proactive, targeted healthcare. We made their SaaS work on Cloudera Enterprise and helped them improve patient data integrity and data security along the way.

Let us help your business too.

Spotlight on Cloudera

Cloudera is revolutionizing enterprise data management by offering a 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 1,900 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 at: Hadoop-services@intel.com.

Intel.com/bigdata/services

cloudera®

intel®

333587-001