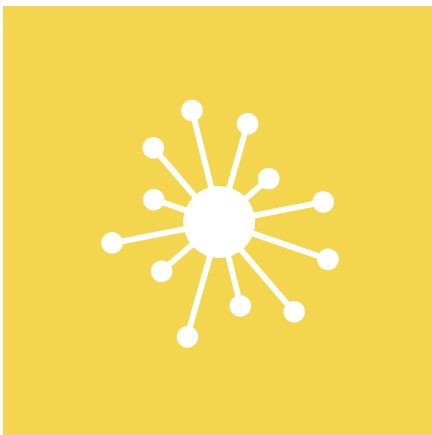


# Intel and Cloudera Reduce Insurance Fraud and Dramatically Improve Time to Access Claim Data

Intel and Cloudera helped reduce fraud in the financial services industry and exponentially improved Big Data analytical performance.



A major insurance company maintains over 60 years of adjuster claim notes in structured and unstructured forms. With Cloudera Enterprise, the Company can interactively research adjuster claim notes in real-time, analyze them to detect key phrases and patterns that identify potential fraud. Users can notify the legal claim department if a claimant engages an attorney, or refine underwriting and loss models with characteristics such as weather or other contributing factors at the time of a claim.

- The system processes unstructured, semistructured, and fully structured data easily.
- Business users conduct their own real-time research and analysis.
- The IT department saves over 1,000 man-hours per year spent on custom coding and batch deployment of analysis/reporting processes.
- Claims adjusters can access notes history in the field and can attach photographs, audio, video, and other documents.

## 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.

## Results

- Users can search results of all data interactive, with most queries completing in under a second. Text-mining results complete within minutes.
- Average response time improves from 72 hours to less than 5 seconds—a factor of more than 50,000.
- New models are more than twice as effective at predicting fraud, with reduced false-positive rates.
- Data scientists can apply analytics to data routinely, as scheduled production tasks.
- The system identifies claimants who have engaged legal counsel, which may affect the way the case is handled.

## Business drivers

Fraud costs the P&C (Property & Casualty) industry \$30 billion each year in the US alone, a figure that is projected to rise, as both opportunistic fraud and professional fraud continue to grow. Reducing fraud contributes to the Company's bottom line immediately and reduces risk exposure.

Big Data capability helps the Company's data analysts find and identify key patterns from free-form text to expose new fraud opportunities. Model enhancements can proactively identify fraud before the Company issues policies or a customer files a claim.

The Company can now “text-mine” their claim notes, which helps them identify individuals or networks of participants quickly and accurately, allowing for timely investigations and appropriate legal actions. Text-mining claim notes for these claim characteristics also enhances the accuracy of detection processes and ultimately claim outcomes.

The Company’s analytic staff can create more accurate and reliable models using and refining skills gained from applying more powerful tools to all available data.

Legal involvement from claimants can adversely affect profits. Early detection that a claimant has retained an attorney has a tremendous impact on the way cases are handled, for example, whether to pay the claim, deny the claim, or settle for some negotiated amount.

With the new system, the Company has a process to identify and alert the legal claim department of a claimant’s involvement of an attorney, based on notes data.

**Solution details**

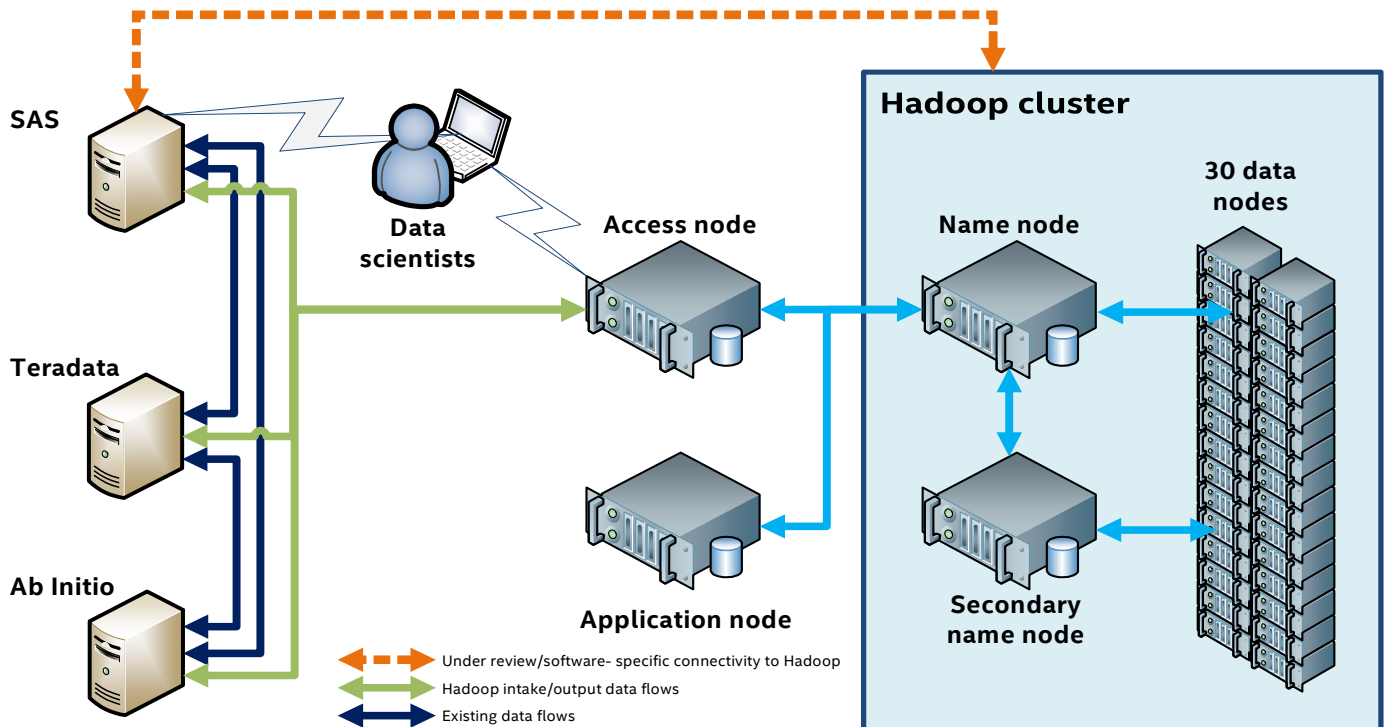
Figure 1 shows the Customer’s high-level Hadoop configuration. Intel architecture leverages hardware and system efficiencies to deliver higher performance and efficiency. For example, our recently released Intel® Xeon® Processor E5-2600 v3 product family, with enhanced Intel® Advanced Encryption Standard New Instructions (Intel AES-NI) embedded, reduces performance latency for encryption/decryption dramatically. In our internal tests with Cloudera, we encrypted an entire HDFS database with a collective increase in encryption/decryption processing time of

less than 0.2% over unencrypted data.<sup>1</sup> That’s virtually zero performance penalty to transparently encrypt an entire Hadoop database.

Fraud networks identify common intermediates (people, businesses, payout recipients, agents) across claims and look for large outliers, using all prior submissions and loss history. All fraud outcomes, referrals, and suspects are stored for pattern analysis and loop back to the underwriting system (Figure 2).

Previous search tools required users to submit static search terms for monthly batch execution, and search results required days of processing time, even when run on a subset of the data. The new system creates more effective models based on all of the data, and users gets results within minutes.

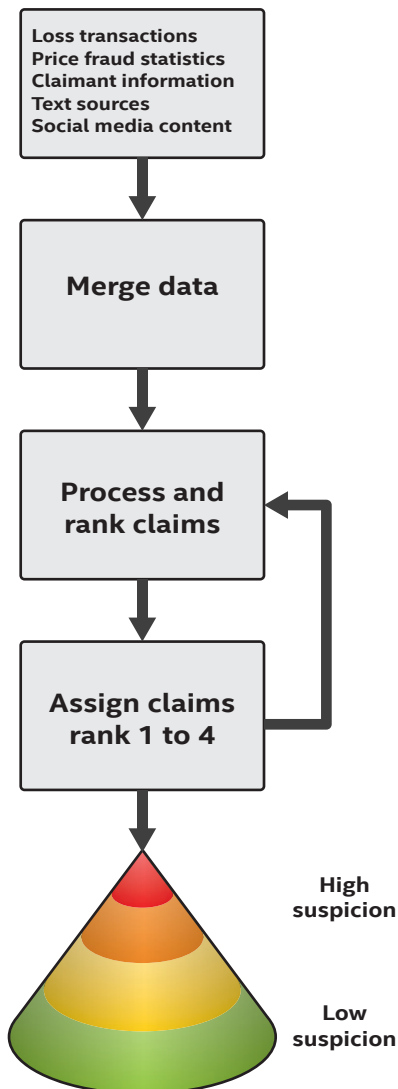
**Figure 1 High-level Hadoop configuration.** Data scientists leverage existing applications along with Hadoop to increase the accuracy of their modeling in near-real-time. The name nodes keep an index of all data in the Hadoop file system; they know the locations of data stored on the data nodes. Direct connection between data scientists and the Hadoop environment requires a Pig scripting language (such as Pig Latin), Apache\* Hive (a structured query language), and R. **Hardware configuration:** Data nodes: HP DL380p with Intel® Xeon® Processor E5-2670 (2.6 GHz, 8 cores), ten 2 TB HDD, two 480 GB OS disk, and 64 GB RAM. Name nodes: The same except 128 GB RAM.



**Cloudera Enterprise**

Adopting Cloudera allows the Company to integrate existing legacy applications and data sources with a variety of powerful new tools and approaches. This enable analytics and processing that was impractical or impossible before. Being open source, Cloudera ensures that the continuing evolution of Hadoop will not “lock in” solutions or tie them to a single source.

**Figure 2 Fraud loss detection modeling.**  
Analyzing new submissions against the historical loss patterns and claimant’s demographic and behavior attributes to find a typical pattern.



Leveraging commodity hardware and open source software, Hadoop offers scalable high-performance storage capacity and processing power with a low initial investment. Hadoop lets data scientists manipulate massive data volumes of structured and unstructured data without the need for highly normalized and inflexible ER modeling and DBMS implementation.

The insurance industry, like many industries that deal with sensitive personally identifiable information (PII), must adhere to security and compliance requirements when handling PII. Cloudera’s comprehensive security solution, with enterprise security controls built into it, provides compliance-ready security out of the box.

Intel’s partnership with Cloudera and our commitment to the open source development of Hadoop ensures stable and consistent delivery of functionality, and continued growth and expansion as Hadoop evolves and matures.

**Summary**

The Company’s data scientists now have real-time access to 60-plus years of adjuster claim notes data, including free text notations, descriptions of conditions, conversations, audio, images, videos, and other claim-supporting documents. They can interactively research and analyze all of the Company’s data in real-time, and get results in a fraction of the time.

The definition of additional and more meaningful data elements will enhance the warehouse environment used by other departments in the Company.

Senior management responded effusively when we showed them this solution, with one exclaiming, “This is what we need right now!”

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](http://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](http://Intel.com/bigdata/services)

**Hadoop sizing guide**

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

**Endnotes**

1. "Cloudera Hadoop\* Single-Node Encryption Performance Case Study on Intel® Xeon® E5-2600 v3 Product Family". [https://software.intel.com/en-us/articles/cloudera-hadoop-single-node-encryption-performance-case-study-on-intel-xeon-e5-2600-v3?utm\\_content=buffer01687&utm\\_medium=social&utm\\_source=twitter.com&utm\\_campaign=buffer](https://software.intel.com/en-us/articles/cloudera-hadoop-single-node-encryption-performance-case-study-on-intel-xeon-e5-2600-v3?utm_content=buffer01687&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer).



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.