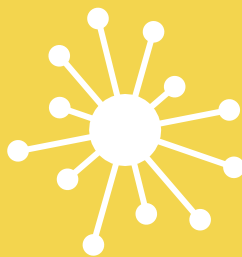intel®

Intel and Cloudera Use Telematics and Analytics to Personalize Auto Insurance Coverage

Intel and Cloudera ingest telematics data from vehicles to identify high/low risk clients and help an auto insurance company individualize customer rates.

Why Intel and Cloudera

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

A large European auto insurance company uses Cloudera and telematics to gather, store, and analyze data in real-time from black box devices installed in their clients' vehicles. The Company uses this information to adjust rates and deductibles based on each individual's driving patterns, mitigating risk while increasing both profit and market share.

Results

The Company uses telematics—a blend of GPS and wireless telecommunications technologies packaged in a "black box" that is installed in a vehicle—to create an individualized prediction of a customer's likelihood of involvement in collisions or other claims-related events. Customers "opt in" with the promise of lower rates for demonstrated safer driving. Using Cloudera-based analysis of the telematics data, the Company benefits in many ways:

- Greatly minimizes fraudulent claims and reduces the number of all claims by 30%.
- Attracts and retains low-risk (safer) drivers with lower cost incentives, increasing user base.
- Encourages high-risk drivers to change their driving habits.

- Provides customers with an online portal to study and improve their driving skills.

Business drivers

The Company and auto insurers in general face a number of issues:

- Italy has the highest frequency of accidents as well as the highest average cost of damages per claim in Europe; total benefits and claims paid by the Company have increased every year for several years.
- In Italy, 2.81% of claims involve fraud, representing 2.44% of total claims paid by insurers.
- The emergence of price-comparison websites has intensified price competition, which has driven less profitable premiums.
- New cars have higher safety standards, but they are also more expensive to insure, repair, and replace. As a result, new vehicle registrations in Italy declined 7% in 2014, which adversely affects profit-and-loss rates.
- Regulation compliance laws in Italy prevent insurers from refusing consumers, or discouraging them from seeking coverage by quoting a higher than usual rates.

Solution details

Insurers' records show trends among drivers in various demographic groups, which serve as a good indicators of a driver's likelihood of becoming a claimant based on age, gender, insured vehicle, and other factors. What's not included in these records, unless documented by a history of claims or violations, is the actual driving habits of the insured party.

Telematics has opened up new business models like pay-as-you-drive (PAYD) and pay-how-you-drive (PHYD). Whereas conventional vehicle insurance premiums are calculated based on age, gender, and driving experience, behavior-based insurance allows insurers to charge based on actual driving habits. The Company hoped a telematics-based solution would help them reduce fraud and manage the risk effectively using Big Data technologies.

A black box fitted to a customer's car would constantly record information, such as GPS location, driving speed, distance and time of drive, rapid/smooth acceleration or braking, and cornering habits, as well as information from external sources, such as weather, traffic reports, and collisions involving other vehicles.

The Company will then analyze this data to create a personalized record of each customer's driving habits. By focusing on the individual characteristics of a customer's driving tendencies, the Company will improve its ability to accurately predict the odds and cost of the customer filing a claim, and can adjust rates and deductibles accordingly. A driver who drives less responsibly, would be charged a higher premium than a driver who drives smoothly and with less calculated risk of claim propensity.

Furthermore, the Company hopes this solution will allow them to offer more products and value-added services to customers, such as road-side/accident assistance, stolen vehicle recovery, remote diagnostics, fuel management, etc. Data analysts could also configure models that would automatically detect accidents and airbag deployments, locate stolen vehicles.

Figure 1 shows an overview of the solution Intel helped design:

Data collection and ingestion. The Company subcontracts with three geographically distinct third-parties to install the black boxes in customer vehicles. These third-parties also capture the telematics data from the black boxes and upload it to the Company's edge node using secure FTP. Every hour, data files from the edge node batch-load into HDFS in raw format. Other data from external sources—such as weather, accidents information, traffic, etc.—also loads into HDFS.

ETL and persistence. Hive queries with user-defined functions, Oozie (a workflow scheduler system to manage Hadoop jobs) workflows,

Figure 1 **A telematics-driven Big Data solution.** Because of the large volume of data generated by the black boxes, the Company chose Cloudera to gather all the end-user data along with weather and road conditions and other data, then extract, cleanse, transform, and aggregate the data into Cloudera Impala and HBase tables. From there, users log on with existing authentication tools to perform analysis on the data within the Cloudera enterprise data hub.
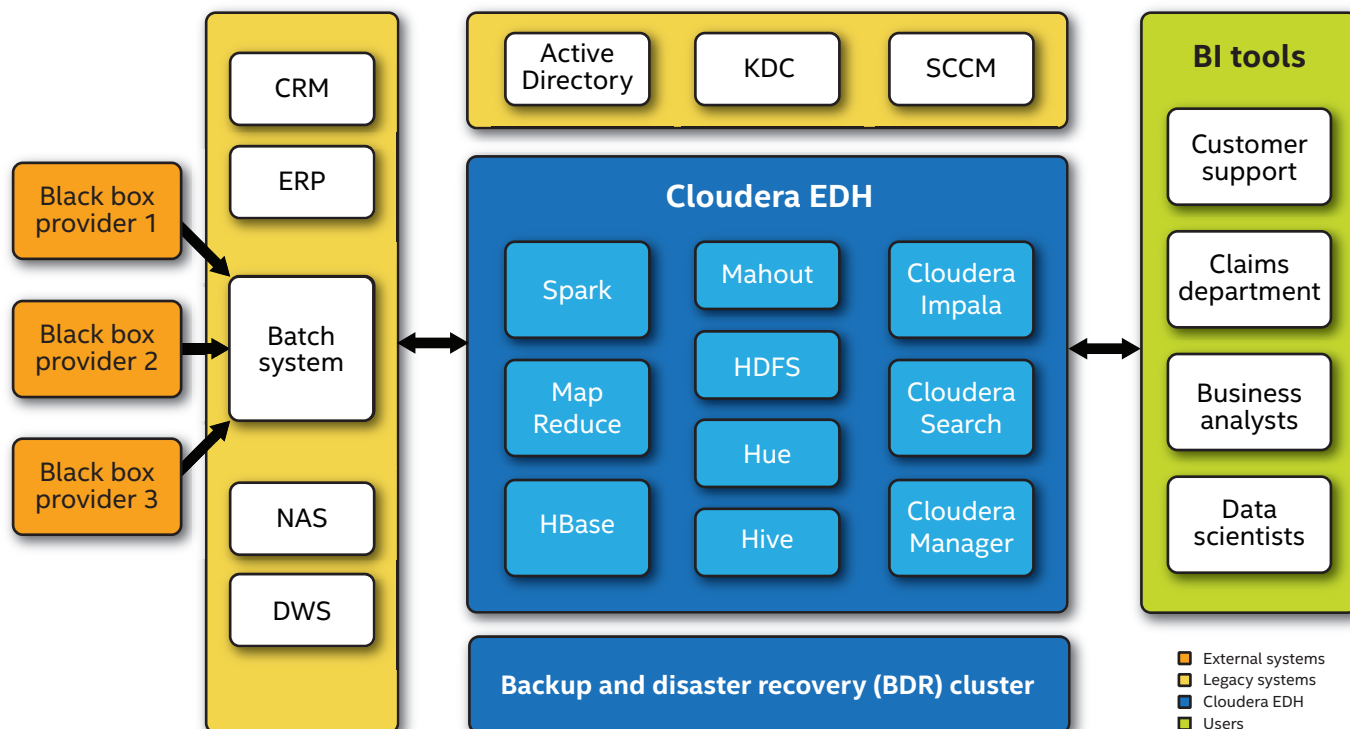
and custom applications extract, cleanse, transform, and aggregate the data. The output loads into Cloudera Impala and HBase tables.

Modeling and analytics. Mahout, R, MLlib (Apache Spark's machine learning library), and custom applications analyze customer behavior, classification, and prediction, then calculate a risk score.

Enterprise reporting and consuming. The Company's business analysts use BI tools like QlikView to connect Cloudera Impala tables. Data is also consumed through a RESTful API and exported from other enterprise applications.

Backup and disaster recovery (BDR). For regulatory compliance, automatic backups between the production and BDR clusters occur at regularly scheduled intervals.

Cloudera Enterprise

The Company chose Cloudera for its several reasons:

Volume. Each telematics device generates events—including date, time, speed, location, acceleration/deceleration, etc.—every second during a drive. Thousands of such devices would generate terabytes of data every month.

Velocity. Telematics devices generate data at high rate. Hadoop is a natural solution to capture, store, and process the telematics data that coming into an organization.

Processing. The platform must handle semistructured data natively. Hadoop makes it possible to store data in any format and convert data downstream to more sophisticated formats like Parquet or event HBase tables.

Versatility. Hadoop supports all of the workloads the Company requires to perform telematics analytics: stream processing, low-latency queries against semistructured data, machine learning algorithms running in batch mode (MapReduce), etc.

Scalability. With Hadoop, it is easy to add nodes as the number of customers or the workload increases.

Enterprise-grade features. The Company deemed these features very important:

- Encryption at rest and in motion.
- Automated backup and disaster recovery.
- SNMP support and alerts.
- AD/Kerberos integration.
- Rolling updates.
- Quota management.

Cost effectiveness. Storing and processing historical telematics data in Hadoop is cost effective compared to traditional data warehouse solutions.

Summary

By combining automobile black box telematics with Cloudera-based analytics, the Company harnesses a powerful new source of information to accurately predict risk levels among its consumer base. Using CDH, the Company now makes more consistent and accurate predictions based on drivers' real world habits and driving conditions, adjusting rates accordingly. Conversely, conventional methods rely only on previous driving records, as well as age, gender, and locale demographics to predict driver risk.

With the advanced analytics Cloudera provides, the Company is able to mitigate losses from high-risk drivers and fraud, as well as obtain and retain more customers while encouraging safer driving habits. The Company needed a robust solution that would reduce fraud and make their business more profitable in a more competitive landscape, and Intel was there to help them realize their Big Data vision.

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.