



Gaining Greater Insight into Advertising Impact

Overview

Cox Automotive is a leading provider of products and services for automotive dealers and car buyers. A subsidiary of Cox Enterprises, the company unites 20 brands, including AutoTrader.com, Kelley Blue Book, Manheim, NextGear Capital, and vAuto, to help transform the way people buy and sell cars.

Each of the company's brands addresses a different segment of the automotive lifecycle—from manufacturing and distribution, through ownership and reselling, and, ultimately, to end of life. As Cox Automotive was formed, executives sought to implement an enterprise data strategy that would enable the company to ingest, enrich, blend, and expose rich automotive data from across its brands. The goal: to deliver new customer and consumer experiences through greater insight.

By working with **Cloudera** to deploy a data discovery and analytics environment on subject-based enterprise data hubs, the company can better organize, manage, and analyze diverse data, spanning vehicles, customers, products, audiences, and users.

This work has helped Cox Automotive realize a 50 percent decrease in total cost of ownership for its data environment. However, according to executives, an even greater benefit is in the new products and services the company can offer.

For example, streaming applications will enable AutoTrader.com to deliver real-time updates, instead of hourly updates, regarding the impact of auto ads during the Super Bowl. Likewise, new inventory management insights, culled from data generated across its ecosystem, will help dealers better gauge supply and demand, so they can get the best price for their cars in the shortest amount of time.

The Challenge

“Data is fundamental to our strategy,” said Scott Salter, vice president, Enterprise Data Services, Cox Automotive. “We’re bringing data from our brands together in a strategic way to introduce new products that wouldn’t be possible otherwise.”

In 2014, Cox Enterprises established Cox Automotive to bring its wholesale and retail automotive businesses under a single operational entity. Doing so, company executives believed, could bring new levels of efficiency and innovation that would be difficult for its competitors to match.

Key Highlights

Industries

- Automotive
- Digital marketing
- eCommerce

Locations

- Headquarters: Atlanta, Georgia, USA
- 150 locations worldwide

Business Application Supported

- Enterprise data hubs integrating data from 20 brands
- Streaming applications for delivery of real-time dashboards

Impact

- Reduces total cost of ownership by 50% per terabyte
- Enables real-time vs. hourly updates on ad influence during the Super Bowl
- Helps dealers optimize inventory management

Technologies in Use

- Apache Hadoop Platform: Cloudera Enterprise, Data Hub Edition
- Apache Hadoop Components: Apache Hive, Apache Hue, Apache Kafka, Apache Spark Streaming, Cloudera Manager, Impala
- Visualization & Analytics Tools: Datameer, Platfora

Big Data Scale

- 1 PB of data and 2 million files

As the new organization coalesced, Salter and his team worked to create an enterprise data strategy that would enable the company to uncover insights across the automotive lifecycle and deliver new offerings.

The challenge, however, was that its existing data environments couldn't offer the scalability and flexibility the company needed to support an enterprise data strategy across its 20 brands.

"Our legacy environments include everything from Netezza to Oracle to SQL Server, and they each have their own challenges with scale and with ingesting new data," said Salter. "It was just not feasible to bring all of our data from all of our brands together in these traditional environments."

Added Jon Gregg, senior analytics engineer, Cox Automotive, "As we've expanded the number of companies under the Cox Automotive umbrella, we needed to lay a foundation for rapid data ingest. In the past, we had to do surgery overnight to get data ready for analysis; it was far from real time. If an analyst working at AutoTrader.com wants to incorporate data from Kelley Blue Book or VinSolutions, they need to get access to that data right away."

The Solution

"As we've moved to Cloudera's platform, powered by [Apache Hadoop](#), we have been able to eliminate the challenges we faced with our legacy environments," said Salter. "We can acquire the data quickly and bring it directly to the analysts across the company very quickly. We can also avoid the pain of building extract, transform, load (ETL) processes and structuring all the data in potentially multiple environments. Now, everyone across multiple business units can have access to the same datasets, and analyze them with the tools and skills they already use."

According to Salter, the team evaluated several options, selecting Cloudera based on the company's vision, proven experience in the marketplace, and performance and manageability. "We needed to have a partner that was invested in their product and in their direction as much as we were, and Cloudera gave us that confidence," said Salter. "Additionally, Cloudera's enterprise data hub offered the analytic performance necessary for real-time processing, and easy administration through [Cloudera Manager](#) meant we could quickly get started."

[Cloudera Professional Services](#) and [Support](#) also provided significant value in helping the organization implement Hadoop. "There's a lot involved in setting up a Hadoop cluster," said Gregg. "For example, there are a number of settings that a user can select for each [Apache] [Hive](#) and [Spark](#) job, and there are security issues, including determining which users get which permissions and how we assign users to groups. It was hugely helpful to have Cloudera Professional Services help us through these foundational issues. Now for ongoing support, Cloudera Manager makes it easy for the Support team to see our logs and quickly understand and address any issues."

Today, the company has created five enterprise data hubs (EDHs) that enable it to better organize and manage diverse data sets from its brands. Totalling more than 1 petabyte (PB) of data and incorporating 200 million rows of data daily, these EDHs support the following subject areas:

- **Vehicle:** A consolidated group of attributes that can be associated with a vehicle through a US or Canadian VIN
- **Audience:** A consolidated group of attributes and information associated with an individual consumer
- **Customer:** A consolidated group of people and companies that pay Cox Automotive for products or services
- **Product:** A catalog of products and services representing value that could be monetized or measured by customers
- **User:** A consolidated group of individual representatives within the B2B relationships with Cox Automotive companies

Cloudera provides the data discovery and analytics environment to enable common access, storage, and processing of the data for analytics. “Cloudera serves as our analytics platform and helps us better understand our customers, our consumers, and our evaluations,” said Salter.

The company performs a wide range of analytics on the data, from reporting to sophisticated statistical modeling, using Apache **Impala**, Apache **Hive**, and **Hue**. “Impala is going to be a core tool going forward,” said Gregg. “It will give us the opportunity to provide analysts with near-Netezza speeds, but on the Hadoop cluster. Analysts benefit from having access to all these datasets across Cox Automotive, using the SQL skills and tools they’re already used to, and getting the same speed as they’re used to.”

Third party products, such as **Datameer** and **Platfora**, help analysts better understand the data. “We wanted to provide visualization tools and spreadsheet-like interfaces to help analysts see what’s possible,” said Salter. “The ability to integrate Datameer and Platfora tools within our platform has been a tremendous asset to help build the momentum and the adoption around our analytics community.”

Spark Streaming will be used to create real-time dashboards for both internal and external use cases. “There are a lot of use cases for our Spark Streaming applications,” said Gregg. “We’ve created executive dashboards to display core growth metrics. We’ve also found that we can use these streaming dashboards to monitor core application and IT metrics in real-time.”

As the team builds its Spark Streaming applications, it uses **Apache Kafka** to help prevent data loss during testing.

“If there was a problem during testing, we needed a way to be able to pick up at the previous index and keep rolling,” says Gregg. “Kafka was made specifically for that and Spark enables us to link into Kafka directly.”

Cloudera Manager helps the team track how many resources each job requires. “We’ve been able to catch some cluster issues early on using Cloudera Manager that we probably would not have caught otherwise,” added Gregg

Impact: 50 Percent Reduction in Total Cost of Ownership

With Cloudera, the team reduced the total cost of ownership of its data environment considerably. “We’re seeing approximately a 50 percent decrease per terabyte in the cost between our current environments and Cloudera,” said Salter.

However, while the savings is significant, Salter said company executives are most excited by the potential to bring new offerings to market faster.

“The biggest value that we’ll get is throughput on our data from an analytics and product development standpoint,” said Salter. “Fundamentally, that’s what we’re trying to do, is bring products and analytics to market faster than anybody.”

Impact: Real-time Tracking of Super Bowl Ads

Advertisers are constantly trying to measure the impact of different ads. AutoTrader.com provides one measure for automakers, illustrating how ad campaigns impact searches and other activity on specific vehicles. For Super Bowl ads in particular, there is a lot at stake and AutoTrader.com has provided this analysis following the game. Using Apache Hive, they were able to speed this process from overnight to hourly updates throughout the day of the Super Bowl.

Now, using a Spark Streaming application, Cox Automotive will offer real-time updates in the next Super Bowl. “We built a streaming application that we tested this year with our executive team,” said Salter. “There was a lot of excitement and emails going around asking if we could show real-time data next year as opposed to hourly data. It wasn’t something we could even consider before.”

“As we’ve moved to Cloudera’s platform, powered by Apache Hadoop, we have been able to eliminate the challenges we faced with our legacy environments.”

— Scott Salter, Vice President, Enterprise Data Services,
Cox Automotive

Impact: Improved Inventory Management

The new EDHs are also helping the company provide auto dealers with new insight to help them better manage their inventory.

“On the inventory management side, we understand what cars are for sale in the market and we understand the consumer demand in that market for those cars,” said Salter. “By connecting this data, we very easily see if there’s too much supply based on the demand, and we can provide pricing and sales recommendations to help dealers get the best value from a particular car in the quickest amount of time.”

For Salter, this is just one of the many new opportunities an enterprise data strategy offers. “When we show our analysts, executives, and IT staff what’s possible, we can see the light bulbs going off,” he said. “This work has stimulated innovation and is enabling us to do things with our data that we hadn’t even thought about before.”

About 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 1,700 partners and a seasoned professional services team help deliver greater time to value. Leading organizations in every industry plus top public sector organizations globally run Cloudera in production. www.cloudera.com

cloudera.com

1-888-789-1488 or 1-650-362-0488

Cloudera, Inc. 1001 Page Mill Road, Palo Alto, CA 94304, USA

© 2015 Cloudera, Inc. All rights reserved. Cloudera and the Cloudera logo are trademarks or registered trademarks of Cloudera Inc. in the USA and other countries. All other trademarks are the property of their respective companies. Information is subject to change without notice.

Cloudera_CaseStudy_Cox_Automotive_102

