Digital Advertising Company

Delivering 300% Better Marketing Campaign Performance with Big Data Expertise

Company Overview

In the advertising space, sensory inputs help advertisers correct course in order to meet a specific goal: to drive people into a store, purchase a product from their website, or register for an event or document online. The internet acts as a giant sensor, and an increasingly common pattern in big data is that sensors provide the data, that data is used to create models, and those models are executed by actuators which initiate change.

The challenge: in marketing, sensors and actuators are disconnected. Because of this disconnect, it takes weeks to modify a campaign and that latency is no longer acceptable.

One digital advertising company has built a marketing automation system to connect sensors and actuators. In advertising, there are analytical decisions that provide strategy and there are analytical decisions that are tactical. This company's platform automates the tactical decisions by streaming digital data and running analysis on it in real time.

Business Challenges before Cloudera

As a cost-conscious start-up, the digital advertising company needed an affordable data management platform that could support petabyte-scale data processing and real-time analytics. After researching various technologies in the market, Hadoop emerged as the best fit. Because Hadoop uses industry standard hardware, the cost per terabyte (TB) of storage is, on average, 10x cheaper than a relational data warehouse system. Hadoop's flexibility allows the company to constantly think and re-think about the best way to read and calibrate data. To accomplish the same tasks in a relational database would mandate the development of user defined functions (UDFs), requiring as much work as Hadoop, Hive and MapReduce while offering less flexibility.

The company started using Apache Hadoop running MapReduce jobs from scratch. Their co-founder and chief technology officer explained, "It was fairly easy to learn in the beginning, but quickly became hard to manage at scale. Dealing with bugs in the analytics code and keeping the system running became increasingly challenging."

cloudera®

Key Highlights

Industry
- Digital Marketing

Location
- Boston, MA, USA

Business Application Supported
- Digital advertising campaign optimization

Impact
- 2–3x better campaign performance
- Hundreds of thousands of variables ingested and analyzed in real time

Technologies in Use
- Hadoop Platform: Cloudera Enterprise Core + RTD
- Hadoop Components: HBase, Hive, MapReduce
- Data Warehouse: Greenplum

Big Data Scale
- 1 PB+ on Hadoop
- 100 TB on Greenplum

Advice to New Hadoop Users
- Use a commercial Hadoop distribution to simplify production and expedite roll-out
- Don't think of Hadoop as an enterprise data warehouse (EDW) — it's different
- Leverage an open source platform to take advantage of rapid innovation and the community's expertise

These challenges led to an evaluation of commercially packaged distributions. The team engaged with Cloudera to evaluate Cloudera Enterprise Core with Cloudera Enterprise RTD. Cloudera Enterprise Core includes the market-leading Hadoop distribution (CDH) plus Cloudera Manager and Cloudera Support to ensure ongoing reliability and simplified management as the cluster continues to grow. Cloudera Enterprise RTD is an add-on package which provides management and support for HBase.

The alternative: hire full-time administrators to manage the environment in-house.

During the sales process, Cloudera engineers helped solve existing problems the team was facing, and they valued the number of core Hadoop Committers that Cloudera has on staff. The fact that CDH is 100% open source and Apache licensed was also an important factor, allowing the company to take advantage of the rapid innovation within the open source community while having the peace of mind that their platform would be reliable and supported over the long haul. Overall, Cloudera offered strong performance at a reasonable price, which led the advertising company to sign up for a Cloudera Enterprise Core subscription.

Use Case

The company collects data and feedback from all digital events pertaining to online advertising campaigns and streams it into a Hadoop-based learning system that was built from scratch to perform several functions:

- Processes real-time data extract, transform, and load (ETL)
- Sorts data by user
- Builds media buying algorithms

The system captures information on how consumers are engaging with different ads, including where and how long they hover, what they're clicking on, and how they're navigating across the internet. This information is combined with data from advertisers, the company's own customer data, feeds from third party providers, and more. They use HBase to facilitate real-time read/write data access in Hadoop, and run analytics on Hive-based on custom-built machine learning algorithms using MapReduce code. The petabyte (PB) sized Hadoop system is integrated with a 100-TB Greenplum reporting data warehouse that offers analytics and business intelligence to clients via TIBCO Spotfire. Data is currently migrated from Hadoop to Greenplum on an hourly or daily basis, depending on user requirements.

The results of this learning system are sent to a media buying system which is connected to 30 different media suppliers around the world.

Impact

This Hadoop-powered platform has allowed clients to achieve 2–3x better performance from their marketing campaigns. In the traditional world, even the best analysts can only comprehend 100 variables affecting campaign performance per day. In comparison, this platform ingests and analyzes hundreds of thousands of variables in real time.

By increasing their processing power and storage capabilities, a small team is equipped to roll out more products at a faster rate than ever before. Their developers can focus on competencies that are unique to the digital advertising business and provide value to customers rather than spending their time managing the Hadoop cluster. Further, the company is better equipped to serve large enterprise brand clients with the stability, support and real-time capabilities enabled by Cloudera Enterprise Core + RTD.

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, process 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. Only Cloudera offers everything needed on a journey to an enterprise data hub, including software for business critical data challenges such as storage, access, management, analysis, security and search. As the leading educator of Hadoop professionals, Cloudera has trained over 40,000 individuals worldwide. Over 1200 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. www.cloudera.com.

cloudera-casestudy-Anonymous_DX-102