



## Helping the Fortune 100 Access a 360-degree Customer View Through a Self-serve Data Hub

### Overview

**Magnify Analytic Solutions** — a wholly-owned division of Detroit, Michigan-based **Marketing Associates** — uses technology-driven data analysis to offer clients a range of informed business services that increase profitability through its four lines of service: business intelligence, digital intelligence, credit risk management, and marketing analytics. Magnify works with Marketing Associates' blue-chip portfolio of clients — including BMW Financial Services, Comcast, Ford Motor Company, Jaguar Cars, TD Bank, and Whirlpool.

To meet the growing demands of its Fortune-100 client base, Magnify deployed an **enterprise data hub** (EDH) on **Cloudera**, delivering flexibility, agility, and fast time to value, at a fraction of the cost of alternative solutions.

### The Challenge

Since its inception, Magnify has built broad expertise around its legacy RDBMS environment, processing large datasets for customers to support deliverables such as business-to-consumer (B2C) online marketing contests and product giveaways. Magnify has executed about 150 of these campaigns in the last year; each one drives hundreds of thousands of conversions and offers lots of insight about customer sentiment within seconds of launching.

**Andy Frey**, Magnify's CIO, noted, "It's common to get as many as 200,000 hits in the first minute for some of the more popular brands' campaigns. For instance, in one household brand's giveaway, we issued one half million products in less than 12 hours."

Running campaigns like the aforementioned one present a big data challenge: the huge spikes in web event data resulting from each campaign should be efficiently collected, processed, and analyzed to derive consumer insights that make the campaigns worthwhile. From a cost standpoint, the company must control how much processing power will be required and for how long, taking into consideration network limitations and other factors. To support this campaign, for example, Magnify needed 80 web servers for a very short period of time. In another instance, "We once launched two high-profile campaigns simultaneously, whose combined response literally shut down the data center that we were operating in," said Frey.

## KEY HIGHLIGHTS

### Industry

Business Services

### Location

Detroit, Michigan, US

### Business Applications Supported

Enterprise data hub empowering digital marketing and analytics services

### Impact

- All-inclusive Cloudera EDH costs less than relational database license alone
- Fast time-to-value and easy administration via Cloudera Manager
- Scalability and flexibility to accommodate diverse and growing client demands

### Technologies in Use

- Hadoop Platform: Cloudera Enterprise, Data Hub Edition
- Hadoop Components: Apache Flume, Apache Hive, Apache Pig, Apache Sqoop, Cloudera Impala, Cloudera Manager, Hue, MapReduce
- BI & Analytic Tools: D3.js, SAS, Tableau
- Security Tools: Apache Sentry, Kerberos

### Big Data Scale

- 10+ years' historical data on Cloudera
- 20 TB+

Magnify also saw the need to be able to predict the precise time when each giveaway would reach its limit, monitoring conversions as they happen, to shut down registrations at the right time based on the supply allocated for that campaign.

Meanwhile, Magnify was offered the opportunity to manage an automotive client's customer relationship management (CRM) system and evolve it into a centralized data hub delivering a 360-degree view of customers, encompassing more data types from more sources, and including vehicle information from local motor vehicle departments, dealer distribution statistics, and parts data from services organizations. "We initially ran the system with four relational databases, but the magnitude of data—coming from every single customer touch point and every dealer—quickly exposed the shortcomings of this approach," recalled Frey. "We also knew that the second phase of the CRM project would include multi-structured data, and many new sources, including Google Analytics."

It became imperative that Magnify employ a new, Big Data infrastructure with:

- Elastic flexibility and scalability to handle massive, short-term spikes in demand, coupled with the ability to manage and manipulate massive data volumes from multiple sources.
- Agility and performance in the ways data can be organized and restructured over time. "As the client's CRM program developed," noted Frey, "our original design proved to be far too rigid. Query times began to slow because the data wasn't optimized to accommodate an evolving use case. We started looking for a solution that didn't place constraints on how we leveraged the data for our clients."

## Solution

Frey and his team were aware of the growing popularity of Apache Hadoop. Kerry Shaffer, IT director at Magnify, recalled, "We really liked what Hadoop had to offer and started researching the different commercial implementation options."

Magnify ultimately shortlisted two Hadoop vendors. "We had to figure out which distribution and which vendor we liked working with best," said Shaffer. **Cloudera Manager** became a major differentiator: "In comparison to other Hadoop management tools, Cloudera Manager made it so easy to manage our clusters, was extremely user-friendly, and dramatically reduced our learning curve."

As a key part of their due diligence, the Magnify team modeled costs, including licensing expenses, for a full-scale implementation of the client's production environment. "When accounting for the future growth of the business, we evaluated our existing environment and realized the opportunity cost of not trying out emerging technologies like Hadoop was too great," explained Shaffer.

Frey added, "We could purchase **Cloudera Enterprise, Data Hub Edition** and everything else we needed for our development, production, and QA environments, and even after paying for a month of development expenses, were still well below the cost we would have incurred on our traditional systems. This was a huge differentiator in Cloudera's favor."

Magnify decided to move forward with an EDH powered by Cloudera. Today, data moves between the EDH and Magnify's incumbent relational data warehouse using **Apache Sqoop** or **Apache Flume**. Magnify users have the option to interact with the data hub using a variety of tools, depending on their preferences and skill sets, including **Apache Hive**, **Apache Pig**, **Cloudera Search**, **Impala**, **SAS** (which leverages the SAS/ACCESS Interface to Impala), and **Tableau**.

With most of its clients being in the Fortune 100, security was a key requirement for Magnify's environment. The data hub is authenticated using Kerberos, and will implement **Apache Sentry** (incubating) for role-based security soon. Magnify's infrastructure is SSAE 16 certified and HIPAA compliant.

### Impact: Time-to-Value

Magnify already had installed Cloudera in a development environment, but engaged **Cloudera Professional Services** to accelerate the rollout and training. "We wanted to get it right the first time, and Cloudera's expertise enabled us to achieve this. The Cloudera team was super helpful and the knowledge we've accumulated in-house has had a positive impact on employee retention. Everyone is excited to work with such an innovative technology," noted Shaffer.

### Impact: Assimilation & Expansion

Magnify's continually evolving production environment frequently has to cope with the addition of new data sources. As Frey described the process, "We can add a whole new system that couples existing data with new: We can bring up a second Hadoop instance on the same cluster and build another interface with absolutely no need to replicate data. This is just great from so many perspectives." For instance:

- It reduces security risk by not having important customer data in multiple places.
- It ensures that the business is always leveraging the cleanest data.
- It enables Magnify to track data without any breaks in the lineage path.

Also, with Hadoop, if transaction volumes are projected to create a potential performance bottleneck, the remedy is very straightforward: "If we see the possibility for throughput degradation we just snap in a couple of extra nodes. Everyone is impressed with how quickly it can be done and with the immediate impact it has on performance. It's a huge win for us and our customers," noted Shaffer.

### Impact: Self-Service Data Discovery

Magnify also worked with **Cloudera Support** to optimize query performance. By deploying Impala and allowing for self-service data discovery capabilities, Magnify can offer clients a web-based solution through which they interact directly with Hadoop. Frey recalled, "We used to distribute Microsoft Excel reports to customers every one or two days but now dealers can search on their own by customer, sales deal, or even service type. Impala is used to query millions of rows to identify specific records that match the dealers' criteria. We've even given clients a 'Query Hadoop' option that allows them to create simple SQL statements and query Hadoop directly via Impala. We're able to offer a faster, richer, and more accurate selection of services without the labor or latency concerns that we used to have."

"Hadoop really lives up to expectations," Shaffer summarized. "It gives us exactly the flexibility and scalability that we need, as well as being impressively cost-effective. Our partnership with Cloudera has enabled us to create an environment that optimally leverages what Hadoop has to offer, and permits us to maximize the value we deliver to our own clients."

"In comparison to other Hadoop management tools, Cloudera Manager made it so easy to manage our clusters, was extremely user-friendly, and dramatically reduced our learning curve."

Kerry Shaffer, IT Director, Magnify

## 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 22,000 individuals worldwide. Over 1,200 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.

---

[cloudera.com](http://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-magnify-casestudy-Q115-101

