**Case Study**

**Extracting Insight and Value from a Lake of Data**

EMC champions internal use of its data analytics and storage solutions based on Intel® technologies to promote smarter, insights-driven marketing.

All industries today—from retail and healthcare to telecommunications and manufacturing—are witnessing the impact of the data explosion driven by growth in mobile devices, connected “things,” and cloud-enabled services. Extracting value from growing data stores, however, remains challenging for many businesses. Early and compelling success stories around the use of analytics are continuing to drive strong interest among business decision makers to use data-driven insights for competitive advantage, ongoing innovation, and operational efficiency. Traditional BI is rapidly evolving to accommodate more advanced analytics that go beyond rear-view reporting to predictive capabilities based on broad data sources—from structured databases such as CRM, SCM, and ERP as well as unstructured data from sensors, mobile apps, social networks, and IoT devices.

As an industry leader in storage and analytics technologies, EMC is promoting data-driven business transformation for its customers by helping them store, manage, protect, and analyze their most valuable asset—data. Understanding those customers and their behaviors is important for EMC to stay ahead of market preferences. The company was interested in developing more sophisticated and targeted marketing outreach that could intercept new business opportunities in a timely fashion. But the key challenge at EMC—as with most businesses—wasn’t lack of data to deliver that insight but rather an absence of a holistic, strategic approach to data management that could drive business insights and results quickly and with impact.

**Business Challenge**

With the acquisition of over 80 companies over the past decade, EMC had grown very quickly, leading to data management issues. IT controlled multiple siloed data stores, which led to a controlled reporting environment with no coordinated master data management approach. This resulted in extensive “shadow IT” with limited governance and EMC’s IT department serving as a gatekeeper, and controlling the data and reports required to run the business. Business units relied on IT to run reports, creating unnecessary bottlenecks. Lack of an agile mechanism to deliver on-demand analytics to business leaders severely limited the potential to use data to drive impact. While data reporting helped with the “rear-view mirror” insight, the business was unable to fully exploit data’s predictive potential for targeted marketing and lead generation campaigns. Infrastructure limitations also kept EMC from easily and cost-effectively scaling to accommodate the variety and the volume of data that came with increasing demands on query and load performance.

**Swimming in a Lake of Data**

In 2014, the company created the Marketing Science Lab to begin addressing the problem. As part of this change, EMC consolidated multiple siloed islands of data into a data lake that ingests all of EMC’s structured and unstructured data sources—from customers (such as past purchases), contact demographics, interests and marketing history, to unstructured data from social networks, sensors, mobile apps, and IoT devices.

**Faster, Real-Time Customer Insights for EMC Marketing Using a Data Lake**

**Business Need:** Drive more targeted, data-driven marketing and lead generation programs based on customer insights harnessed from broad data sources (structured and unstructured).

**Challenge:** Company-wide data silos, lack of strategic approach to data management and governance, heavy dependence on IT for analytics reporting.

**Solution Approach:** Consolidate “islands” of data into a data lake and implement governance for seamless data sharing across business groups.

**Technology Platform for Data Lake:** EMC Greenplum® Data Computing Appliance and EMC Isilon® Scale-Out NAS. Based on Intel® Xeon® processors, Intel® Solid State Drives (SSDs), and Intel® non-volatile memory. Supported by custom algorithms developed by EMC team.

**Key Benefits:** Data query timelines reduced from 4 hours to less than 1 minute. 80% predictive accuracy in analytics models for customer behavior, helping influence near real-time decision making for marketing programs. Analytics “sandboxes” for business units to enable self-serve access to rich and comprehensive data sources for analysis, minimizing dependence on IT.
the web, and sensors. The company also implemented governance to enable business groups to share and collaborate on analytics initiatives.

Michael Foley, the head of EMC’s Marketing Science Lab, the division tasked with turning EMC into a marketing-driven organization, says, “With this information, EMC is able to micro-segment customers for more relevant and targeted marketing communications.” Breaking down the data silos meant that the marketing organization could build a more comprehensive understanding of EMC’s customers and their purchasing behaviors.

“Using a data lake approach that enables us to look at all data holistically to track predictable patterns, we’re able to optimize our media mix for cost efficiency and better lead quality, and intersect more meaningfully with our customers’ purchase journeys,” added Foley.

The data lake used by the EMC Marketing Science Lab was built on EMC federation technology, including the EMC Greenplum* Data Computing Appliance (DCA) and Isilon* Scale-Out NAS. The EMC Greenplum DCA family is a group of purpose-built, highly scalable, parallel data warehousing appliances that architecturally integrate database, compute, storage, and network into a single, easy-to-manage enterprise-class system. EMC Isilon Scale-Out NAS is a flexible, multiprotocol, high-performance storage platform. Both solutions are powered by Intel® technologies, including Intel® Xeon® processors, Intel® Solid State Drives (SSDs), and Intel® non-volatile memory.

“Intel computing power allows you to run demanding analytics workloads, which would have been too computationally complex before. Some of these techniques are very iterative and use thousands of variables; other systems simply cannot keep up,” says Foley.

Deeper Insights with Advanced Analytics

Foley’s team at the Marketing Science Lab developed all the analytics algorithms themselves. “The reality is, if you outsource your analytics, or use an off-the-shelf solution, you may not get the deep insight your business needs. Because we approach each problem as unique, and in a consultative way—with better hardware and technology—we are able to gain a deeper knowledge about our business and achieve a sustainable marketing advantage through data and analytics,” says Foley.

Because the algorithms are tuned and the data lake holds all of EMC’s data, it enables near real-time decisions about which data points to choose. In fact, the time it takes for some data querying has gone down from four hours per quarter to less than a minute per year. Using this approach, Foley explains, “You don’t have to commit at the beginning which data and technique you are going to use. Say you come in with one thousand variables and some might be redundant. Typically, you want the most simple, elegant model and the machine can help the data scientist select the most powerful variables for prediction, resulting in the most efficient and accurate model.”

With high-performance Intel technology-based hardware and custom algorithms, EMC is able to run analytics across all data sources in the data lake simultaneously, and in near real-time. In fact, the time it takes for some data querying
has gone down from four hours per quarter to less than a minute per year. “The ability to rapidly develop and run many different modeling approaches lets you test a lot of potential solutions to a given problem. As a result, we are able to choose the model with the highest predictive accuracy, explanatory power, and marketing effectiveness,” says Foley.

With the wide-ranging and diverse data set EMC has to work from, their prediction models produce a high rate of accuracy. “We are able to correctly predict what and when a customer is going to purchase 80 percent of the time.”

The analytics engine creates personas that are updated each month, and then EMC looks at who is predicted to make a purchase in the next 3 to 6 months. The analytics engines continually ingest new data and hence “learn” more each time. “Every time they make a mistake, the model learns and gets smarter,” says Foley.

Letting Everyone Play in the Sandbox
EMC has secured the data lake and has given lines of business their own “sandboxes,” otherwise known as analytics environments, where they can run analytic scenarios in a self-service capacity. This allows EMC IT to transition from being a gatekeeper to a position of managing and supporting the data lake platform and providing the technical and data expertise. Meanwhile, business groups can now be more self-sufficient with their reporting and analytics initiatives without heavy dependence on IT.

Foley says, “Data insight is currency everyone can use and at EMC. Everyone is now lined up out the door to get into the data sandbox so those insights can help shape daily business decisions and drive impact.”

The deep technology alignment between EMC and Intel helps enable high-performance, cost-efficient platforms such as the EMC Greenplum DCA and Isilon Scale-Out NAS that are optimized to handle analytics workloads of all kinds. EMC’s successful use of data lakes in its Marketing Science Lab is just one example of how organizations can derive value from data to address a specific business need—in this case, more fine-tuned insights on EMC customers to help tailor marketing programs. Having a robust Intel technology-based infrastructure across their storage and compute environments delivered the trusted analytics performance for EMC’s data lake to solve a real business problem. “At EMC, we’re all about dogfooding our analytics solutions internally because the data challenges we wrangle with as a business are not too unlike those of our customers. We could not do this without the horsepower of Intel technology within our solutions,” Foley says.

For more information on Intel® software-defined infrastructure products, visit: www.intel.com/analytics