EMC ACCELERATES JOURNEY TO BIG DATA WITH BUSINESS ANALYTICS-AS-A-SERVICE

An account of EMC IT’s transformation to empower business and IT users with streamlined access to Big Data Analytics

ABSTRACT
This white paper examines how EMC is exploiting the Big Data opportunity with analytics as a service, a new agile model for analytics and reporting based on shared infrastructure. Business-Analytics-as-a-Service is providing EMC with significantly reduced total cost of ownership, predictive analytics proficiency, and increased business agility. The paper details EMC’s Business Analytics-as-a-Service architecture, deployment, results, best practices and early adopter use cases.

January 2013
# TABLE OF CONTENTS

INTRODUCTION .................................................................................5

LEGACY BI ENVIRONMENT.................................................................6

NEW DELIVERY PARADIGM:
BUSINESS ANALYTICS-AS-A-SERVICE...........................................7

PRIORITIZED ROLLOUT .....................................................................8

EARLY TANGIBLE RESULTS.............................................................9

USE CASE: TOTAL CUSTOMER EXPERIENCE ................................9

USE CASE: SALES AND MARKETING..............................................10

JOURNEY TO BIG DATA ANALYTICS.............................................10

LESSONS LEARNED .........................................................................11

CONCLUSION..................................................................................12

REFERENCES...................................................................................12
EXECUTIVE SUMMARY

Enterprises are being flooded with information and are more challenged than ever to harness insights that bring value to the business. As islands of information across enterprises grow without centralized management, companies face increasing challenges with ensuring security, compliance and data quality. In addition, some business units, frustrated with long lead times for IT to deliver business intelligence (BI) platforms, are pursuing their own solutions, creating “shadow IT” environments that increase risk exposure and expense for enterprises.

This white paper examines how EMC solved these challenges by transforming IT and the business with analytics as a service, a new approach that unlocks the value of Big Data.

Business Analytics-as-a-Service (BAaaS) introduces a new agile model for reporting and analytics, enabling IT and business users to focus on what they do best. BAaaS empowers business analysts and data scientists across the enterprise with secure access to EMC’s global data warehouse and advanced tools to generate their own analytics and reports. The paper details EMC’s Business Analytics-as-a-Service architecture, deployment, results, best practices and two early adopter use cases.

While EMC is still rolling out BAaaS, the savings and value are already clear. Reports that once took hours are generated in minutes. By using a shared BAaaS infrastructure, EMC IT expects to reduce total cost of ownership to the business. With a few hundred BI applications, savings are projected to be in the tens of millions of dollars based on infrastructure, licensing and support costs.

And that’s not all. With greater agility, increased analytics proficiency and standardized access, BAaaS is proving to be a critical success factor in EMC’s transformational journey to Big Data, generating growth and increasing competitive advantage.
INTRODUCTION

Big Data is a growing reality among enterprises worldwide. In a report titled Big Data: The Next Frontier for Innovation, Competition, and Productivity, management consulting firm McKinsey & Company cites that 15 out of 17 industry sectors in the U.S. already have more data per company than the entire U.S. Library of Congress. In IDC’s white paper, Extracting Value from Chaos, researchers also reported that information in the enterprise will grow more than 50 times before 2021.

EMC Corporation is no exception when it comes to Big Data. With 55,000 employees and 400,000 customers and partners in 80 countries, the company depends on a vast IT infrastructure. EMC’s IT organization oversees 2,500 databases, 500 business applications and tools, 5 data centers and 12 petabytes of information.

Steady revenue growth and the acquisition of 70 companies have driven rapid accumulation of Big Data at EMC. For example, EMC’s global data warehouse growth has increased from 300 gigabytes quarterly to nearly 400 gigabytes monthly over the last 4 years.

With Business Analytics-as-a-Service, EMC IT is transforming how Big Data is delivered and consumed (Figure 1). Business analysts can leverage EMC enterprise data and still develop their own Big Data analytics in their own workspace. EMC IT and EMC’s business units are embracing new strategies and tools that are improving data loading, performance and functionality of Big Data. With these advancements, EMC is creating opportunities for fostering innovation, collaboration, growth, productivity and competitive advantage.

Figure 1: Journey to Big Data Analytics

1. Consolidation of Islands of Information
   - Master Data
   - Common BA Tools
   - BA Governance

2. BA-As-A-Service
   - Data Scientist Services
   - Collaborative Analytic Tools
   - Unified Analytical Platform With Chorus

3. Advanced Predictive Analytics With Business Units
   - Deep Data Center Analytics
   - Advanced Security Analytics

BIG DATA INFRASTRUCTURE TECHNOLOGY FOCUS

AGILE ANALYTICS PEOPLE & PRODUCTIVITY FOCUS

PREDICTIVE ENTERPRISE APPLICATION FOCUS

2010 2011 2012 2013 2014
LEGACY BI ENVIRONMENT

The center of EMC’s legacy BI infrastructure (Figure 2) has been a data warehouse that leverages Oracle 11g databases running on Red Hat Linux supported by EMC VMAX tiered storage. With EMC’s fast-growing data sets and demand for more complex analytics, existing BI databases and tools, such as Microsoft Excel and Minitab, were not delivering the speed, collaboration and sophistication the business units needed.

For example, EMC IT required 31 hours to load and prep data sets and an additional five days to run batches of hundreds of static reports. If data sets were corrupted, it could take six to eight weeks to rebuild everything.

In addition, EMC’s existing infrastructure lacked the performance and scalability to support the increased demand and workflows generated by predictive modeling and data mining. EMC’s mobile workforce also was demanding access from any location.

While EMC was outgrowing the capabilities of standard reporting tools and platforms, solutions to manage fast-growing unstructured data had not been rolled out. To complicate the situation further, there was a shortage of skilled IT resources and data architects who were needed to navigate the world of Big Data analytics.

Figure 2: EMC BI Infrastructure – Before
Faced with long waits, some business units pursued data analysis on their own, using a mix of BI and infrastructure solutions acquired and maintained outside of IT. These “shadow IT” environments were draining business units’ resources. While increasing EMC’s risk exposure and expense, shadow IT also was limiting EMC’s ability to protect critical corporate data.

In addition, an EMC study found that staff involved with gathering data and generating reports spent most of their time using labor-intensive manual processes and tools such as Microsoft Excel and Access, with only a fraction spent deriving business insights.

**NEW DELIVERY PARADIGM: BUSINESS ANALYTICS-AS-A-SERVICE**

To tap into its growing volumes of Big Data, EMC IT launched its analytics as a service initiative to provide business units with self-service capabilities while freeing them from managing their own IT environments. The aim was also to reduce duplication of efforts for data lifecycle activities, such as data acquisition and preparation. For EMC, corporate data governance, data protection and security relating to storage and distribution of Big Data are assured.

EMC’s Big Data infrastructure (Figure 3) is based on a EMC Greenplum Unified Analytics Platform (UAP), which includes the Greenplum Database for structured data; Greenplum HD, an enterprise Hadoop offering for unstructured data; and Greenplum Chorus, which serves as the collaboration layer for EMC’s data scientists. EMC’s business units now have their choice of industry-leading analytics tools integrated with Greenplum for analysis and data visualization.

The Greenplum solutions are hosted on a production EMC Greenplum Data Computing Appliance (DCA), a unified and modular Big Data solution. EMC’s full-rack Greenplum DCA comprises two master servers and 16 segment servers running on 192 processors, as well as 768 gigabytes of memory and 36 terabytes usable storage. A similarly configured Greenplum DCA is used for testing and development in a secondary EMC data center.

The back end of the Greenplum is directly connected to EMC’s enterprise extract, transform and load (ETL) platform, which is hosted on a completely virtualized infrastructure. The ETL platform loads data at blazing speeds from various sources, including SAP, Salesforce.com, PeopleSoft, and Oracle, into the Greenplum DCA.

For high availability, EMC Data Domain deduplication storage is directly connected to the Greenplum DCA’s back-end switches to back up critical data sets. EMC also leverages Data Domain’s deduplication technology to significantly reduce backup data volumes and speed backup operations. To enable disaster recovery, Data Domain is replicated to another Data Domain system in the secondary EMC data center located 600 miles away.

Configured and managed by EMC’s IT staff, Greenplum DCA’s well designed ‘resource queue’ technique helps manage often unpredictable and high-volume analytics workloads effectively. The resource queue establishes a threshold for simultaneous user queries to ensure the system maintains high levels of performance and utilization.

In addition, EMC IT depends on a variety of tools to administer the Greenplum platform, including Greenplum’s pgAdmin’s GUI interface, Aginity Workbench, as well as internally developed tools and scripts. Primary administrative tasks include creating sandboxes for new users, managing user passwords and enabling browsing and access levels. EMC IT also uses the tools to ensure adequate compute resources and performance are available for long-running queries or high query volumes.
PRIORITIZED ROLLOUT

With the complexity of the existing BI environment, which included tens of thousands of reports and thousands of ETL jobs, EMC IT needed to prioritize which solutions would initially be migrated to analytics as a service. EMC IT conducted extensive surveys and benchmark studies to justify its decisions.

EMC’s Customer Quality organization was placed first in line. The benchmarks showed that Customer Quality would produce BI analyses and reports 2,000 percent faster using analytics as a service compared to its previous mix of IT managed platforms, business supported tools and manual processes.

Additional business unit use cases were selected based on their business requirements. For example, Marketing needed predictive analytics, which is an ideal fit for Greenplum. Other early adopters were Licensing & Compliance, Sales, IT, Global Services, Manufacturing and Engineering.

After prioritizing the order of business units, EMC IT created a standard process for migrating data from its Oracle data warehouse to Greenplum UAP. The goal was rapid data migration while ensuring data security and reliability. EMC’s Oracle to Greenplum migration best practices are described in the EMC Greenplum Data Computing Appliance Enhances EMC IT’s Global Data Warehouse white paper.

With the Greenplum DCA installation taking just 16 hours, the migration was performed in a single bulk data transfer that pre-seeded Greenplum with 17 terabytes of enterprise data from Oracle. EMC IT then focused on adding specific data sets required by the business unit, ensuring that all incremental updates and day-to-day support processes were in place. This process will continue until data sets for all business units are migrated.
EARLY TANGIBLE RESULTS

The migration itself to the new Business Analytics-as-a-Service platform turned out to signal other improvements. For example, Greenplum ingested data from Oracle much faster than expected. During the initial migration, EMC IT transferred the entire data warehouse in hours instead of days, enabling the Business Analytics-as-a-Service project to go live two months ahead of schedule.

As BAaaS is rolled out across EMC, business units are reporting tangible results. Some nightly data loads that took eight hours in Oracle now require less than two in Greenplum. Instead of waiting five days for batch reports from IT, business units can independently run ad hoc reports in minutes.

In addition, BAaaS frees business units from the time and expense of maintaining their own reporting infrastructures to focusing more on data analytics while still having the flexibility to customize their Big Data tools and reports.

By using a shared BAaaS infrastructure, EMC IT expects to reduce total cost of ownership to the business. With a few hundred BI applications, savings are projected to be tens of millions of dollars based on infrastructure, licensing and support costs.

Sean Brown, EMC Director of Enterprise Business Intelligence, explains, "EMC has a growing appetite for advanced analytics. With Business Analytics-as-a-Service, we want to empower the business users to harness Big Data and unlock its full value."

Because IT centrally manages the vendors and contracts, business units can leverage the best negotiated rates on software and resources. In addition, IT is better able to enforce corporate and data security standards while making more efficient use of EMC assets.

USE CASE: TOTAL CUSTOMER EXPERIENCE

EMC’s Corporate Quality organization is responsible for driving continuous improvement in EMC’s Total Customer Experience (TCE) programs by recording and tracking customer experience and product quality metrics, including performance of components and systems across EMC’s entire installed customer base.

With a database approaching one billion records and growing by roughly one million records daily, the organization was struggling to track hundreds of field requests each quarter. These field requests, along with other data elements, generate system reliability reports used to improve product quality and serviceability. Formerly relying on archaic hardware, SQL database tables and manual data feeds, the TCE organization often required five days to produce a single customer product quality report.

Since adopting analytics as a service, TCE can produce reports in as little as eight minutes compared to hours or days. The organization also compressed time to process field reported problems from 11 months to less than four weeks. In addition, the TCE organization was able to transform the roles of spreadsheet managers to data analysts who leveraged the analytics as a service platform to perform analysis and predict component reliability.

Jim Bampos, Vice President, EMC Corporate Quality, summarizes, “The impact of analytics as a service to our customers is enormous because with more real-time intelligence, faster reporting, and predictive analytics, we’ll be able to dig deeper and potentially fix problems before they occur. And it lowers our cost of service, improves product quality, and enriches the customer experience.”
USE CASE: SALES AND MARKETING

Prior to analytics as a service, EMC Marketing had a fragmented view of customer interactions across Sales, websites, online forums, social media, transactional support and service. The organization needed tools for capturing and consolidating data about these interactions and understanding buying behavior.

With BAaaS, data scientists can build predictive models to identify customers with a high propensity to buy. Marketing also linked BAaaS to its cloud-based Aprimo integrated marketing automation solution, which is used to design and execute targeted marketing campaigns.

Todd Forsythe, Vice President, EMC Direct Marketing, says, “With analytics as a service, we’ll have a far superior analytic workbench. An enterprise-wide view of all customer interactions will turbo-charge our marketing efforts. It will also enable us to collaborate with EMC Sales and other organizations and leverage our respective analyses to create a force multiplier on the same target customers.”

EMC’s Global Business Operations credits analytics as a service and recruitment of data scientists as enabling significant revenue and go-to-market opportunities for the sales and marketing organizations.

John Smits, Senior Director, Sales Analytics and Territory Management, EMC Global Business Operations, explains, “Thanks to Business Analytics-as-a-Service, we’re now armed with a Greenplum sandbox environment and SAS statistical modeling solution. There are countless examples of how Greenplum has reduced OBIEE reporting activities from 20 days or more to minutes.

“We’ve also been able to move third-party analytical models from external vendors into our own data rich production-ready SAS environments,” Smits continues. “It’s remarkable what we’ve achieved in such relatively short period of time.”

JOURNEY TO BIG DATA ANALYTICS

Business Analytics-as-a-Service has facilitated the intersection of EMC’s journeys to Cloud and Big Data. With 90 percent virtualization, EMC’s cloud environment has been the ideal infrastructure for managing vast quantities of data because of its extraordinary flexibility, scalability and performance. The journey to Big Data has involved consolidating islands of information onto a global data warehouse, rolling out an IT-as-a-Service model that includes Business Analytics-as-a-Service, and deploying analytics engines (see Figure 3).

Ramesh Razdan, EMC Senior Director and EMC Distinguished Engineer, comments, “With Business Analytics-as-a-Service, we put data from millions of records and dozens of sources at the fingertips of our data analysts and business users, while improving overall data quality. This new window on analytics promises to expand our ability to mine our growing reservoir of Big Data and bring new value to the enterprise.”

The next step focuses on productivity by placing powerful collaboration tools such as Greenplum Chorus in the hands of business data analysts. Then the focus will be on building Big Data-enabled applications to create a predictive enterprise, with the ultimate goal to enable real-time decisions and data monetization.
LESSONS LEARNED

EMC IT is already applying lessons learned from the earliest stages of its Business Analytics-as-a-Service implementation. As a new best practice, EMC IT is rolling out BAaaS-focused marketing communications early in the process for each business unit. By fully understanding the objectives and benefits of BAaaS, the business units are more likely adopt the new platform and processes more quickly and successfully.

Another important realization was the need to provide additional upfront Greenplum training to data analysts to ensure a smooth transition. The massively parallel processing (MPP) architecture of Greenplum is substantially different from the Oracle database environment.

EMC IT was pleasantly surprised that its own workloads increased since business units started using tools and resources enabled by analytics as a service to create their own solutions. Furthermore, the business units are consulting with EMC IT on how to leverage predictive modeling, tuning, solution optimization and measurement.

With the advent of Business Analytics-as-a-Service, EMC IT is retraining its BI team and expanding their skill sets. The staff has evolved from writing reports and dashboards to higher value-add activities that require consultative and communication skills. EMC IT also has made a strong effort to develop and attract the data science expertise that is necessary to build and support BAaaS infrastructures. For additional perspective on the burgeoning data science field, see McKinsey & Company’s report: Big Data: The Next Frontier for Innovation, Competition, and Productivity.
CONCLUSION
As illustrated by the Customer Quality and Marketing use cases, Business Analytics-as-a-Service offers compelling benefits that will continue to grow as additional use cases and data sources emerge, technologies advance and creative analytical techniques proliferate across EMC.

BAaaS is saving potentially tens of millions of dollars, speeding time to business value and improving quality—all while ensuring data security and IT governance.

As more business units adopt Business Analytics-as-a-Service, opportunities for collaborating and sharing expertise will grow. EMC is continuing to add tools for more advanced analytics. And as EMC's Big Data journey continues, the company expects analytics as a service to enable a truly predictive enterprise, leveraging Big Data across every dimension for real-time decision-making that drives business growth and value.

REFERENCES
For more information, please visit:

www.EMC.com/EMCITProven

EMC IT’s IT Transformation blog at http://itblog.emc.com/

EMC Greenplum Data Computing Appliance Enhances EMC IT’s Global Data Warehouse