Introduction

Companies are capturing, storing, and analyzing terabytes to petabytes of data in hopes of uncovering insights that can improve customer acquisition, cross-sell and up-sell, drive product and service strategies, optimize business operations, and create competitive advantages in a dynamic global marketplace. The promise of new business drivers such as social networking and smart networks are driving business users’ expectations for more refined customer insights, timely business monitoring, and optimized business execution. But existing technologies, platforms, tools, and “know-how” are hindering their ability to monetize their data. Teams responsible for implementing these initiatives are often challenged by a lack of time, knowledge about new tools, and big data experience to properly develop and execute a plan that puts them in a position to be a first mover in the big data revolution.

Enter Big Data

The recent introduction of the term “big data” refers to more than significantly larger volumes of data. It is also about the velocity at which the data is being captured, the variety of data types that now need to be managed, and the complexity of the different data types and formats. Existing data integration and data quality processes conducted with current tools and methodologies become bottlenecks when existing challenges are exacerbated as the volume, variety, and complexity of targeted data sets increase. Reporting and analyzing these big data sets requires new techniques, architectures, and approaches that exploit high performance, massively parallel processing (MPP) data platforms, and advanced analytics on both structured and unstructured data sets. And all of this needs to be considered in light of how the cloud can provide a more agile, less costly operational platform. Only when all of these moving parts can be orchestrated by an experienced team can organizations realize the business potential and impact of big data.

What Defines a Big Data Analytics Project?

Any time the business has a complex analytics challenge that involves massive volumes of data that is processed and analyzed in real-time with advanced analytic tools while leveraging new types of structured and unstructured data, you have a big data analytics project. Like any other analytics endeavor, big data projects require a goal, a measurable ROI, key performance indicators against which success will be measured, experienced people, and a solid, repeatable process. This is no different from any data analytics project—except that the massive volume, scale, and complexities that can follow a big data project makes it all the more crucial that these key components are aligned throughout the process. The business side of the house expects actionable insights that produce meaningful and measurable business results. Often, stakeholders need to be educated on the “where” and “how” of big data’s business value, which creates a need for skilled and experienced resources that know both the business problem and technology capabilities. Let’s explore a couple of examples where companies have embraced big data to deliver differentiated business value

Health Care

A large health care provider wants to integrate medical, lifestyle, and heredity data to improve patient care and treatment effectiveness. Called Comparative Effectiveness Research (CER), the Institute of Medicine defines CER as “the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care. The purpose of CER is to assist consumers, clinicians, purchasers and policy makers to make informed decisions that will improve health care at both the individual and population levels.” This requires the integration and detailed analysis of massive volumes of structured and unstructured data across a multitude of disparate internal and external systems. This data will be used to determine treatment attribution effectiveness, cohorts analysis and recommendations, reduce patient over-treatment, and monitor the health of the overall population.
Retail

A large department store retailer is looking to minimize in-store theft and fraudulent activities by analyzing point-of-sale (POS) and inventory data in real-time. The fraud application requires real-time data feeds from the POS system that must be merged with additional data feeds from inventory management, labor scheduling, human resources, and physical security systems.

In both cases, having the data is just one small aspect of the challenge. The data integration flows must be designed to ensure quality, the reporting and analytics need to be modeled to deliver actionable insights, and the analytic results must be delivered wherever the user might be to whatever platform they may be using. And all this must be done while considering how new sources of big data may affect each step.

Coming up with project measures, a solution design, and a plan for execution requires experience and creativity that may not always be available in existing data analytics teams. Organically building the know-how for big data is possible, but it can be a complicated and drawn-out process. The technologies and techniques used for big data are different from traditional business intelligence and data warehousing dynamics and may be perceived as far more complicated than current approaches and technologies can support. For example, the proper application of statistical analysis, predictive analytics, data mining, neural networks, and other analytical techniques require experts in statistical sciences who also have a strong background in business fundamentals. The “know-how” to “code” advanced analytics is not enough anymore—you must also understand how to build and manage analytic models that can actually run on billions of records, and then be able to operationalize the analytic results back into the product systems and dashboards that run the business. Once internal teams learn how to master these skills, the potential for exploiting big data value is limitless. But how do companies get there before their competitors?

A Shortage of Key Skills

According to ESG’s latest data management survey, 28% of respondent organizations identified a lack of skilled resources as a challenge they encounter as they attempt to manage their current database and supporting infrastructure. When asked about the specific skilled resources they lacked, almost half identified database administrators and data architects (see Figure 1).

Data architects are the key to a successful data analytics project. These resources bridge the gap between the business problem and the physical storage of data. Interestingly, the role of “data scientist” did not exist even a few short years ago, but today 17% of respondents felt that this was a key and lacking skill. This illustrates the increase in importance of data-based skills in today’s analytics-focused business landscape.

1 Source: ESG Research Report, Big Data Analytics & Integration, to be published in September 2011.
ESG research also indicated that new application deployments, upgrades, and a renewed focus on more real-time business intelligence and data warehousing initiatives will continue to remain IT priorities in the next 12-18 months. With already-strained resources and a shortage in these key areas, organizations will need to prioritize projects and hiring plans to meet business demand. This is especially true for projects with a big data component.

Organizations would certainly benefit from consulting services that feature a repeatable process and framework for discovering, analyzing, and modeling data—big or small.

**Teach a Man to Fish**

Organizations that believe their information and analytics are the foundation of their intellectual property understand the importance of keeping knowledge close to the vest. But tribal knowledge simply does not scale. Consultants that build analytics rather than show their clients how to implement big data analytics processes on their own is limiting. What is needed is a focus on skills transfer and a framework for continued success. This outlook is akin to the proverb, “Give a man a fish and you feed him for a day. Teach a man to fish and you feed him for a lifetime.”

In the fast-paced world of analytics, the quicker teams can grasp new approaches and concepts, the better able they will be to feed their companies a buffet of value from their biggest data assets. ROI and people will make a big data project successful, but what about continued, sustainable success? Anyone who has been in an analytics project knows that business requirements can change, often before the previous project requirement has even...
been implemented. Combining services with training—whether it is through formal classroom curricula or transfer of knowledge—will have a significant impact on sustained success. In addition to training, adding process makes success repeatable. Process can help bring new people into the team, onto a project, and make them more productive faster than they otherwise could have done without a process. Yet careful attention needs to be paid to the balance between too much process and impacting a company’s ability to be agile.

Considering the emerging nature of big data analytics, third-party consulting services can add tremendous value to a project. Because there is a lack of general knowledge and skills in the industry related to how to approach a big data analytics initiative, a third-party consulting team that is specialized in this area can accelerate the learning curve. And what many may not know is that, in addition to its entrance into the big data analytics market with the Greenplum acquisition, EMC has an experienced consulting team dedicated to helping organizations accelerate time to big data value.

EMC Consulting’s Big Data Advisory Service

Embarking on a big data journey can be daunting without a solid, reasonable plan in place: a plan that takes a known path with a guide to point out danger or peril along the way. EMC Consulting has developed an offering that can help organizations define, plan, and execute a big data project using a practical methodology that is based on experience, and takes a “teach a man to fish” approach. This ensures continued success long after the consulting engagement has concluded. Focused on a set of deliverables that are not only impactful, but also business-relevant and can be operationalized, these include:

- Business Opportunity Matrix
- BI & Analytics Use Cases
- Data Quality/Data Governance Assessment
- Big Data Reference Architecture
- Vision for the Cloud
- Transformational Roadmap

Business Opportunity Matrix: A List of Prioritized Initiatives Based on the Value Big Data can Bring

Many organizations struggle with simply finding a starting point with a big data analytics project. They have seen the hype, read the marketing materials, and now want to know how their business could realize the benefits big data promises to bring. For organizations with a lack of experience and knowledge in big data techniques, giving a data analytics team insight into how to find and prioritize the best business opportunities for big data analytics is the critical first step.

BI and Analytics Use Cases: Documented Big Data Use Cases Tied to Prioritized Business Initiatives

Once the prioritized business opportunity has been identified, building the storyboards and use cases for a project involving big data is where the rubber meets the road. People learn best through example. The use cases bring together the business problem, the technology, and big data capabilities, leading the team from concept to execution and focusing the project scope.

Data Quality/Data Governance Assessment: A Readiness Review

Big data will put a strain on current data quality and data governance initiatives. As part of the engagement process, the EMC team will work with their clients identifying gaps between current capabilities and the ideal state with respect to data quality and data governance in a big data world.
Big Data Reference Architecture: High-level Reference Architecture

In order to operationalize big data findings and recommendations, a reference architecture is developed based on a complete Enterprise Information Management stack that includes the presentation layer (e.g., BI), analytics, data warehousing, ETL, data quality, data governance and, if applicable, private- or public-based cloud data management.

Vision for the Cloud: Identifies Where and How the Cloud Can Help the Big Data Architecture

Cloud computing transforms how compute power, storage, and network capacity is delivered. In the era of big data, the cloud offers a potential self-service consumption model for data analytics so that business users can quickly collaborate to address immediate business opportunities. The cloud can provide an agile data platform to support the business at a fraction of the cost of traditional data platforms.

Transformational Roadmap: Recommended Next Steps in the Big Data Transformational Journey

Having a plan tailored to applying big data technologies, capabilities, and approaches to the company’s key business initiatives is the best way to accelerate big data adoption. The plan needs to address an organization’s ability to incorporate change and change management can make the transition from traditional to big data analytics much smoother. Big data analytics requires a different way of looking at the problem and a different set of techniques. The element of change can be an organization’s biggest challenge.

These are just a subset of the deliverables one should expect from working with experienced, consulting teams such as EMC’s Big Data Advisory Service. In addition, EMC Education Services now offers the EMC Proven Professional Data Science Training & Certification program, which is designed to enhance the skills of anyone with the appropriate background who is looking to start a career in data science or big data analytics. Marketed as “open” or technology-agnostic foundational level training and certification, topics include an introduction to data science and big data, the analytics lifecycle and methods, and advanced data science theory, methods, technology, and tools plus practical hands-on labs and an exam for certification. Leveraging experts from EMC Education Services and the Greenplum resources in its data compute division, the combination of experience and an intensive curriculum sets up organizations for success.

The Bigger Truth

Big data analytics requires a different set of skills, technologies, and techniques not commonly found in many existing business and data analytics teams. This is not surprising with an emerging approach for analyzing massive amounts of information with management’s expectation that the results can be found within minutes rather than weeks. There is no doubt that big data analytics is taking on a life of its own and many companies are realizing significant benefits as a result. Yet many are still trying to figure out how to apply these approaches to their own businesses.

ESG believes that a major roadblock for market adoption is the lack of skills, knowledge, and experience necessary to deploy a big data analytics program successfully. EMC recognizes this challenge and has developed services that can help accelerate adoption and time to value for companies that are on the cusp. EMC’s practical methodology incorporates operationalizing big data analytics to take organizations down a path that focuses on the process of maximizing a team’s ability to implement big data analytics not just once, but repeatedly.

Providing outsourced big data analytics provides value to the business for a day. Teaching users to conduct big data analytics delivers value to the business for a lifetime. EMC Consulting’s approach for big data advisory consulting services offers that transfer of knowledge element that is necessary to accelerate market adoption.