BeyeNETWORK
Global coverage of the business intelligence ecosystem

Research Report Excerpt: Designing a framework for BI and analytics success

BY WAYNE ECKERSON
Director of Research, Business Applications and Architecture Group, TechTarget, September 2011

Why Big Data?

THERE HAS BEEN a lot of talk about "big data" in the past year, which I find a bit puzzling. I've been in the data warehousing field for more than 15 years, and data warehousing has always been about big data.

Back in the late 1990s, I attended a ceremony honoring the Terabyte Club, a handful of companies that were storing more than a terabyte of raw data in their data warehouses. Fast-forward more than 10 years and I could now be attending a ceremony for the Petabyte Club. The trajectory of data acquisition and storage for reporting and analytical applications has been steadily expanding for the past 15 years.

So what's new in 2011? Why are we are talking about big data today? There are several reasons:

The growth in data is fueled by largely unstructured data from websites and machine-generated data from an exploding number of sensors.

1. Changing data types. Organizations are capturing different types of data today. Until about five years ago, most data was transactional in nature, consisting of numeric data that fit easily into rows and columns of relational databases. Today, the growth in data is fueled by largely unstructured data from websites (e.g, Web traffic data and social media content) as well as machine-generated data from an exploding number of sensors. Most of the new data is actually semi-structured in format, because it consists of headers followed by text strings. Pure unstructured data, such as audio and video data, has limited textual content and is more difficult to parse and analyze, but it is also growing (see **Figure 5**, page 7).

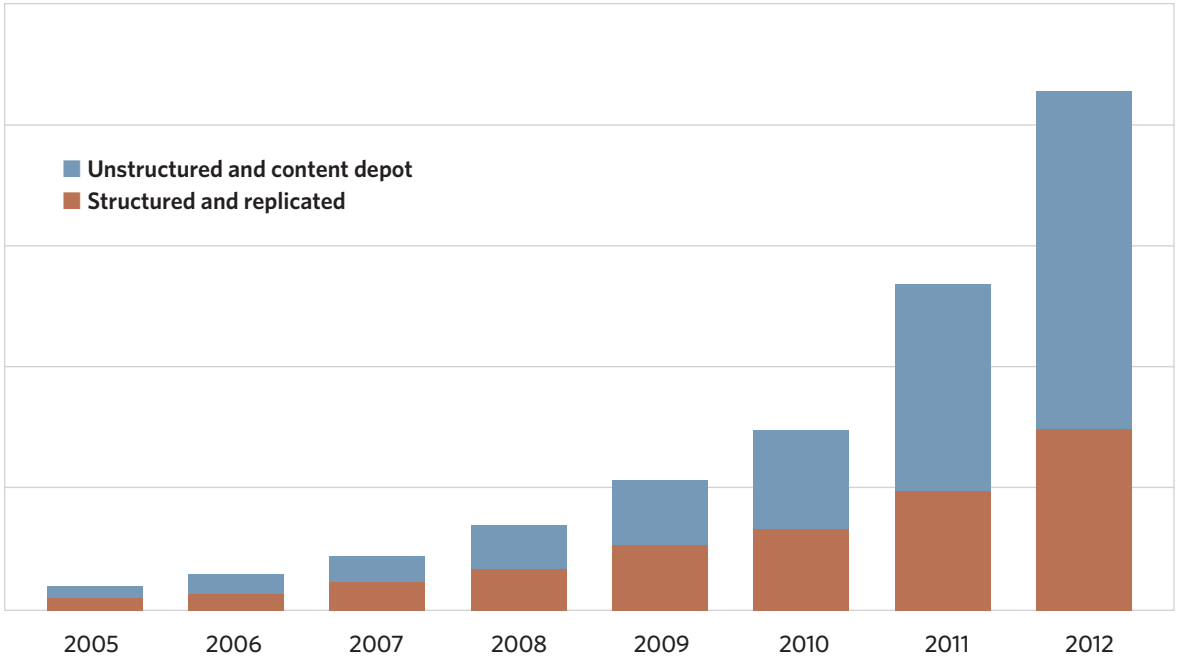
2. Technology advances. Hardware has finally caught up with software. The exponential gains in price-performance exhibited by computer processors, memory and disk storage have finally made it possible to store and analyze

large volumes of data at an affordable price. Database vendors have exploited these advances by developing new high-speed analytical platforms designed to accelerate query performance against large volumes of data, while the open source community has developed Hadoop, a distributed file management system designed to capture, store and analyze large volumes of Web log data, among other things. In other words, organizations are storing and analyzing more data because they can.

Organizations are storing and analyzing more data because they can.

3. Insourcing and outsourcing. Because of the complexity and cost of storing and analyzing Web traffic data, most organizations traditionally outsourced these functions to third-party service bureaus like Omniture. But as the size and importance of corporate e-commerce channels have increased, many are now eager to insource this data to gain greater insights about customers. For example,

FIGURE 5: ***Data growth***



SOURCE: IDC DIGITAL UNIVERSE 2009: WHITE PAPER, SPONSORED BY EMC, 2009.

automobile valuation company Kelley Blue Book is now collecting and storing Web traffic data in-house so it can combine that information with sales and other corporate data to better understand customer behavior, according to Dan Ingle, vice president of analytical insights and technology at the company. At the same time, virtualization technology is beginning to make it attractive for organizations to consider moving large-scale data processing outside their data center walls to private hosted networks or public clouds.

> ***"We are the beginning of an amazing world of data driven applications. It's up to us to shape the world."***
>
> ***—TIM O'REILLY,***
> *founder, O'Reilly Media*

4. Developers discover data. The biggest reason for the popularity of the term *big data* is that Web and application developers have discovered the value of building new data-intensive applications. To application developers, big data is new and exciting. Tim O'Reilly, founder of O'Reilly Media, a longtime high-tech luminary and open source proponent, speaking at Hadoop World in New York in November 2010, said: "We are the beginning of an amazing world of data-driven applications. It's up to us to shape the world." Of course, for those of us who have made their careers in the data world, the new era of "big data" is simply another step in the evolution of data management systems that support reporting and analysis applications. ■

Big Data Analytics: Deriving Value from Big Data

BIG DATA BY itself, regardless of the type, is worthless unless business users do something with it that delivers value to their organizations. That's where analytics comes in. Although organizations have always run reports against data warehouses, most haven't opened these repositories to ad hoc exploration. This is partly because analysis tools are too complex for the average user but also because the repositories often don't contain all the data needed by the power user. But this is changing.

- **BIG VS. SMALL DATA.** A valuable characteristic of "big" data is that it contains more patterns and interesting anomalies than "small" data. Thus, organizations can gain greater value by mining large data volumes than small ones. While users can detect the patterns in small data sets using simple statistical methods, ad hoc query and analysis tools or by eyeballing the data, they need sophisticated techniques to mine big data. Fortunately, these techniques and tools already exist thanks to companies such as SAS Institute and SPSS (now part of IBM) that ship analytical workbenches (i.e., data mining tools). These tools incorporate all kinds of analytical algorithms that have been developed and refined by academic and commercial researchers over the past 40 years.

- **REAL-TIME DATA.** Organizations that accumulate big data recognize quickly that they need to change the way they capture, transform and move data from a nightly batch process to a continuous process using micro batch loads or event-driven updates. This technical constraint pays big business dividends because it makes it possible to deliver critical information to users in near real time. In other words, big data fosters operational analytics by supporting just-in-time information delivery. The market today is witnessing a perfect storm with the convergence of big data, deep analytics and real-time information delivery.

■ **Complex analytics.** In addition, during the past 15 years, the "analytical IQ" of many organizations has evolved from reporting and dashboarding to lightweight analysis conducted with query and online analytical processing (OLAP) tools. Many organizations are now on the verge of upping their analytical IQ by implementing complex analytics against both structured and unstructured data. Complex analytics spans a vast array of techniques and applications. Traditional analytical workbenches from SAS and SPSS create mathematical models of historical data that can be used to predict future behavior. This type of predictive analytics can be used to do everything from delivering highly tailored cross-sell recommendations to predicting failure rates of aircraft engines. In addition, organizations are now applying a variety of complex analytics to Web, social media and other forms of complex structured data that are hard to do with traditional SQL-based tools, including path analysis, graph analysis, link analysis, fuzzy matching and so on.

Analytics increases corporate intelligence. ... and is the only true source of sustainable advantage.

Organizations are now recruiting analysts who know how to wield these analytical tools to unearth the hidden value in big data. They are hiring analytical modelers who know how to use data mining workbenches, as well as data scientists, application developers with process and data knowledge who write programming code to run against large Hadoop clusters.

■ **Sustainable advantage.** At the same time, executives have recognized the power of analytics to deliver a competitive advantage, thanks to the pioneering work of thought leaders such as Tom Davenport, who co-wrote the book *Competing on Analytics: The New Science of Winning*. In fact, forward-thinking executives recognize that analytics may be the only true source of sustainable advantage since it empowers employees at all levels of an organization with information to help them make smarter decisions. In essence, analytics increases corporate intelligence, which is something you can never package or systematize and competitors can't duplicate. In short, many organizations have laid the groundwork to reap the benefits of "big data analytics."

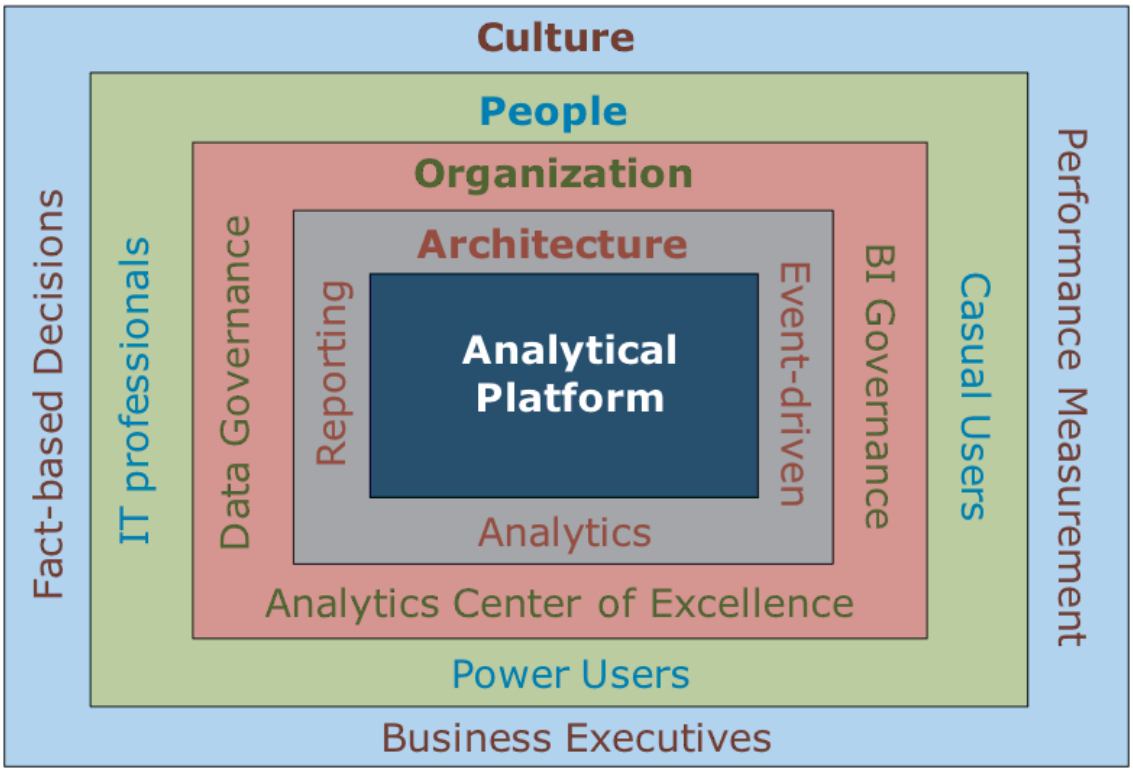
A FRAMEWORK FOR SUCCESS

However, the road to big data analytics is not easy and success is not guaranteed. Analytical champions are still rare. That's because succeeding with big data analytics requires the right culture, people, organization, architecture and technology (see **FIGURE 6**).

1. The right culture. Analytical organizations are championed by executives who believe in making fact-based decisions or validating intuition with data. These executives create a culture of performance measurement in which individuals and groups are held accountable for the outcomes of predefined metrics aligned with strategic objectives. These leaders recruit other executives who believe in the power of data and are willing to invest money and their own time to create a learning organization that runs by the numbers and uses analytical techniques to exploit big data.

2. The right people. You can't do big data analytics without power users, or

FIGURE 6: ***Big data analytics framework***



more specifically, business analysts, analytical modelers and data scientists. These folks possess a rare combination of skills and knowledge: They have a deep understanding of business processes and the data that sits behind those processes and are skillful in the use of various analytical tools, including Excel, SQL, analytical workbenches and coding languages. They are highly motivated, critical thinkers who command an above-average salary and exhibit a passion for success and deliver outsized value to the organization.

3. The right organization. Historically, analysts with the aforementioned skills were pooled in pockets of an organization hired by department heads. But analytical champions create a shared service organization (i.e., an analytical center of excellence) that makes analytics a pervasive competence. Analysts are still assigned to specific departments and processes, but they are also part of a central organization that provides collaboration, camaraderie and a career path for analysts. At the same time, the director maintains a close relationship with the data warehousing team (if he doesn't own the function outright) to ensure that business analysts have open access to the data they need to do their jobs. Data is fuel for a business analyst or data scientist.

4. The right architecture. The data warehousing team plays a critical role in delivering deep analytics. It needs to establish an architecture that ensures the delivery of high-quality, secure, consistent information while providing open access to those who need it. Threading this needle takes wisdom, a good deal of political astuteness and a BI-savvy data architecture team. The architecture itself must be able to consume large volumes of structured and unstructured data and make it available to different classes of users via a variety of tools (see "Architecture for Big Data Analytics" below).

5. Analytical platform. At the heart of an analytical infrastructure is an analytical platform, the underlying data management system that consumes, integrates and provides user access to information for reporting and analysis activities. Today, many vendors, including most of the sponsors of this report, provide specialized analytical platforms that provide dramatically better query performance than existing systems. There are many types of analytical platforms sold by dozens of vendors (see "Types of Analytical Platforms" below). ■

’ ! " & &" "

’ " !! ! $ &" !

’ " &" ! #"" &" ! " % " &!" !

#"
! $ # "
" ! ! " " ! " #
! "
" " ! $ " !
! # # #%