MARK LENKE, GLOBAL DIRECTOR,
BIG DATA AND ANALYTICS, CSC

Mark Lenke is a director with CSC's Big Data and Analytics business unit and is responsible for its global business consulting offerings. He has more than 30 years of diverse experience spanning several industries, including Food Service, Telecommunications, Insurance, Advertising, Manufacturing, Forest Products and Pharmaceuticals. He has spent the past 15 years designing, implementing and managing big data and business intelligence solutions in multivendor, domestic and international environments for Fortune 500 companies.

SHAWN NELSON, SENIOR SOLUTION ARCHITECT, BIG DATA AND ANALYTICS, CSC

Shawn Nelson is a senior solution architect with CSC's Big Data and Analytics business unit and provides business consulting and big data strategy services to CSC's clients. He has 20 years of experience supporting several industries, including Petroleum/Natural Resources, Supply Chain, Manufacturing, Insurance and Pharmaceuticals. He has spent the past 17 years developing, managing and implementing customer relationship management and business intelligence/analytics systems for CSC clients across the globe.

REINVENTION OF ENTERPRISE ANALYTICS

There's no doubt that companies have benefited tremendously from business intelligence (BI) applications. Enterprise business intelligence (EBI) has enabled companies to spot emerging trends, identify new markets, serve customers more effectively and improve operational efficiencies.

Recently, though, EBI solutions have had a hard time adapting to the information explosion that companies have experienced. Attempting to stuff massive volumes of data into the structure required by traditional BI systems is inefficient, expensive and time consuming.

What's more, systems have become more complex and difficult to use, limiting the types of insights that can be generated in a reasonable time frame. Big data solutions, which can efficiently handle large volumes of data, can also require people with a specific and hard-to-find skill set in order to get results. Typically, business process experts feel shut out from advances because new systems are too hard to use.

As a result, many companies have avoided implementing more advanced BI or next-generation big data solutions because there is a perception — real or imagined — that **they take too much time and are too hard to use to justify the expense involved in implementation.**

But avoiding the change means that companies are missing out on opportunities to gain new insights that can radically transform their business. The next generation of BI systems, commonly referred to as big data, offers a huge leap forward in capabilities and features. Big data and analytics can help companies ask sophisticated, forward-looking questions that make new connections between seemingly unrelated trends. Big data and analytics can power new types of applications that provide real-time feedback, putting insights directly into the hands of people who can use that information.

This paper examines advances in big data infrastructure and applications that can help companies overcome the challenges associated with bringing these new systems to life.

LEVERAGING BIG DATA

In addition to the "volume, variety, velocity" benefits, big data also provides lower cost and greater agility when it comes to implementing analytical capabilities and initiatives in the organization. These benefits are often overlooked or misunderstood because the new generation of enterprise analytics, powered by big data, is presumed to suffer from the same problems as its predecessor — that it too takes a long time to set up and is too expensive.

That might be true if a company uses the same approach for building big data capabilities that was used for the earlier generation of BI systems. Three key factors in the implementation of big data solutions can make a world of difference:

1. Cost savings by using commodity hardware and cloud environments

2. Cost savings by implementing open source software

3. Less time required between project inception and benefits realization

FROM BI TO BIG DATA

The enterprise class of BI solutions was originally designed to support organizations with a series of siloed, disparate sources of data. The typical BI process was: Data from the silos is cleaned and transformed through an extract, transform and load (ETL) process into a data warehouse that becomes the system of record, or one version of the truth. The data is then sliced, diced and delivered via various analytical applications to users across the enterprise.

The problem with EBI is the initial length of time and high cost of implementation, coupled with the complexities of implementing changes and enterprise-wide BI application upgrades.

The most popular and communicated benefit of big data has been its ability to ingest and process terabytes and petabytes of unstructured and semi-structured data from sources including machine, mobile and social. But EBI can't efficiently, technically or cost-effectively accomplish this.

For example, open source technologies such as Hadoop, HDFS, MapReduce, Storm and a multitude of NoSQL solutions have helped alleviate some of these issues by creating a distributed framework to efficiently ingest, transform and store structured, semi-structured and unstructured data for analysis.

These technologies ensure that all data sources are available for exploratory analytics, but at a cost. They are complicated, and many of the functions provided by traditional ETL tools are performed using custom MapReduce code. It is virtually impossible to find an analyst who is familiar with an organization's data and can run custom analytics and develop custom MapReduce code to produce the desired data set. This typically requires the use of very experienced developers intimately familiar with the technologies.

ORGANIZATIONAL IMPACT

The evolution that is occurring in the big data analytics space is having a profound effect on the business process expert as well. The impact of organizational change on entire departments and users — end users, super users, business analysts and subject matter experts (SMEs) — cannot be overlooked or trivialized.

Data analysts in finance or operations today who use typical BI applications perform standard, highly normalized (relational) methods for querying data and don't need a significant technical background.

This can lead to a significant shortcoming for today's BI applications. To explore new information relationships, analysts often have to work with IT personnel to develop custom queries that will yield the correct data. This data may, or may not, be what the analyst was originally looking for, leading to a frustrating situation where the analysts understand the data, but only the developers know how to access it.

On the other hand, the role of data scientist is growing quickly in the big data space as the role of the analyst becomes more challenging. Data scientists have traditionally been found in underwriting, actuarial and scientific positions across many industries, but these roles are not necessarily found in traditional business processes. But the data scientist can take on challenging tasks that previously seemed impossible.

To date, data scientists have been working with platforms that provide access to data sources faster and in native format, with tools that provide loosely coupled data relationships. These platforms and tools provide deep analysis of data that leads to a richer, deeper and more comprehensive business view. These methods and tools are still nascent, but some very impressive applications can be found.

But a gap exists between the relatively codified, nontechnical role of an analyst in a traditional BI system and that of the highly specialized data scientist handling a sophisticated big data application.

For ordinary users, the exposure to data and new tools, for analytics, creates a big leap in understanding and development of basic skills in data science. The giant leap, as it appears to most users, becomes a constraint on methods and adoption of analytics tools. Therefore, finding a middle path that bridges the transition is best.

CSC SOLUTION

CSC helps companies realize the potential of big data applications with solutions that improve performance, agility and flexibility while creating an environment that allows a wider range of users to use more data sources and ask more questions to produce bigger answers.

IMPROVING ON OPEN STANDARDS

CSC recently acquired a big data company called Infochimps, a provider of big data Platform as a Service for enterprise businesses and an open source innovator based in Austin, Texas. Infochimps' team of highly skilled big data and analytics professionals and engineers deliver big data systems with unprecedented speed, scale and flexibility to enterprise companies.

At the core of Infochimps' DNA is our unique, open source-based big data and cloud expertise. Infochimps was founded by data scientists, cloud computing and open source experts, who have built three critical analytic services required by virtually all next-generation enterprise applications: real-time data processing and analytics, batch analytics and ad hoc analytics — all for actionable insights and all powered by open standards.

A SIMPLER, SELF-SERVICE USER EXPERIENCE

CSC can provide a solution to enable richer, more efficient self-service for users by taking a portion of the burden off the system. 42six Solutions, a CSC company, has developed a solution called Amino to address this issue. Amino takes a fundamentally different approach to analytics. Developers use MapReduce to create microanalytics to perform common queries and transformations. These analytics are made available to analysts so they can explore data in an iterative fashion, testing hypotheses in real time without development support. This also allows organizations to effectively crowdsource analytical tasks. As analysts piece together microanalytics to solve specific problems, the results can be shared across the organization to allow analysts to contribute to and build on one another's work.

Amino was first applied to the intelligence community to provide analysts efficient access to multiple sources of information (structured and unstructured) and translate them to actionable insights. The following use case shows how microanalytics can be used to quickly compile complicated queries with no custom development:

Each query parameter (for instance, calls at a certain time of day, SMS sent at a certain time) is simple to produce in a vacuum, but once they are paired together, they represent an extremely powerful and complicated query. Under the traditional BI toolset with a relational database, analysts would have to know exactly what they are looking for, identify data sources, and prepare requirements for developers to code an application that could extract the correct data and present it to users. This process could take days or weeks, and might not even yield the desired results. At this point, an analyst would start the process again with a revised hypothesis.

Using Amino, an analyst has access to a graphical user interface (GUI) with data feeds pre-positioned in an open source NoSQL database. Using the GUI, an analyst can select microanalytics and layer them to test their hypotheses in minutes.

Amino has also been used to track social networks, predict patient outcomes in health-care and predict the probability of extreme weather events. Amino's true strength is its open source framework, which can be used to get the right data, in the right place, at the right time — empowering analysts and making the development team more efficient and the development cycle more agile.

The ultimate goal is to provide the enterprise with the ability to obtain significant insights from big data and get them into the hands of business users rather than developers, while at the same time reducing the costs.

CONCLUSION

Enabling more data sources and more data content constitutes a transformation in IT culture. But that path forward isn't always obvious. BI applications have become a staple of enterprise strategy and competitive advantage, but they remain constrained by an outdated understanding of the benefits and capabilities of the last generation.

Big data infrastructure and the latest version of enterprise analytics are doing more than breaking data out of silos. They're also tearing down barriers of time and cost. While BI applications of the last generation were highly static, today's systems are designed specifically to be much more flexible in the types of data they can handle and the types of infrastructure they can be built upon.

"It takes too long and it's too expensive" may have been a valid argument a few years ago when considering the business case for expanding an older BI system, but our work with clients is showing on a daily basis that those obstacles simply no longer exist.

Worldwide CSC Headquarters

The Americas
3170 Fairview Park Drive
Falls Church, Virginia 22042
United States
+1.703.876.1000

Asia, Middle East, Africa
Level 9, UE BizHub East
6 Changi Business Park Avenue 1
Singapore 468017
Republic of Singapore
+65.68099.000

Australia
26 Talavera Road
Macquarie Park, NSW 2113
Australia
+61(2)9034.3000

Central and Eastern Europe
Abraham-Lincoln-Park 1
65189 Wiesbaden
Germany
+49.611.1420

Nordic and Baltic Region
Retortvej 8
DK-2500 Valby
Denmark
+45.36.14.4000

South and West Europe
Immeuble Balzac
10 place des Vosges
92072 Paris la Défense Cedex
France
+33.1.55.707070

UK and Ireland Region
Royal Pavilion
Wellesley Road
Aldershot, Hampshire
GU11 1PZ
United Kingdom
+44(0)1252.534000

About CSC
CSC is a global leader in next-generation IT services and solutions. The company's mission is to enable superior returns on our clients' technology investments through best-in-class industry solutions, domain expertise and global scale. For more information, visit us at www.csc.com.