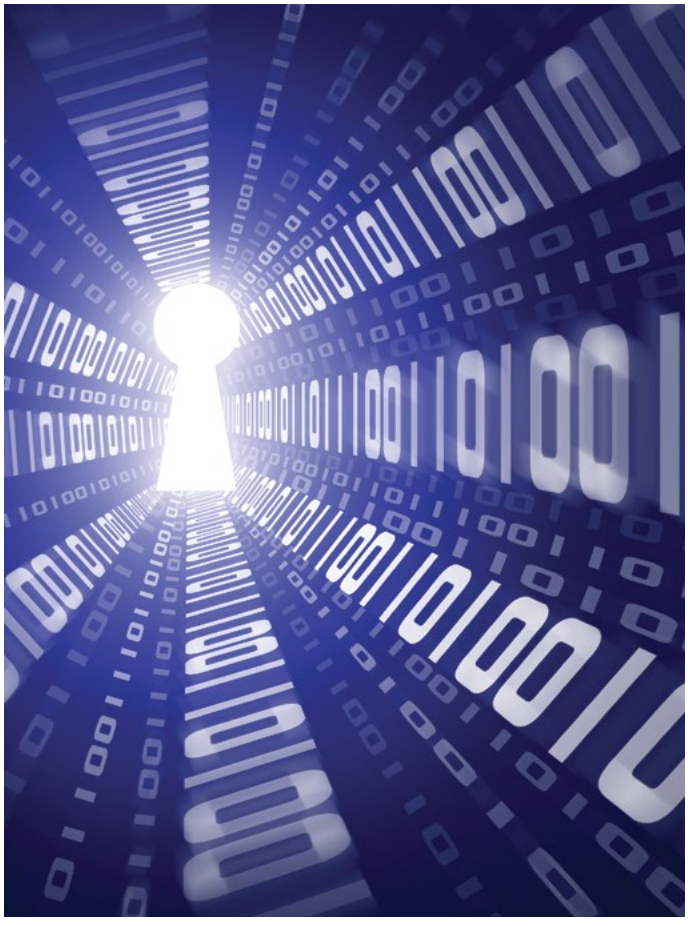
EXCLUSIVELY FOR TDWI PREMIUM MEMBERS

EXCERPT

VOLUME 19 • NUMBER 2

Business Intelligence Journal

THE LEADING PUBLICATION FOR BUSINESS INTELLIGENCE AND DATA WAREHOUSING PROFESSIONALS



tdwi
THE DATA WAREHOUSING INSTITUTE

VOLUME 19 • NUMBER 2

Business Intelligence JOURNAL

From the Editor

Put the customer first. The customer is king. Keep the customer satisfied. Those familiar adages ring especially true in this issue of the *Business Intelligence Journal*.

David Loshin discusses what a customer-centric approach to data entails, the benefits it can provide your business, and how this is reflected in a new approach to master data management.

Customers are increasingly concerned about not only what data organizations keep but also how safe that data remains. Senior editor Hugh J. Watson examines how the data breach Target suffered recently can serve as a wake-up call for enterprises to keep data private and secure. He also describes a new study on privacy focusing on three different ways to characterize privacy and online invasions of privacy. For enterprises looking to get a handle on their big data security risks, Oren Hamami continues the data protection theme by examining five key steps to take now.

Keeping data safe is just one challenge. Nenshad Bardoliwalla examines the impact of the big data revolution on business analysts and why we need a new approach beyond what current data preparation tools offer. Of course, what good is having all that data if your database technology can't keep up? Saibal Samaddar describes two column-oriented DBMS designs to help you manage your growing data volumes without hampering performance.

Finding the resources to satisfy an organization's ever-growing need for BI skills is the focus of two articles in this issue. Ryan Hart and Troy Hiltbrand explore how creative crowdsourcing can help enterprises find analytic talent. The authors show how friendly competition at three organizations brought out the best in their employees. In our BI Experts' Perspective column, Dave Schrader, Ron Swift, and Coy Yonce discuss what advice they would offer college students about BI career opportunities, how students should prepare for (and start) a BI career, and what technical and "soft" skills a graduate needs to be successful.

Looking for advice yourself? Jean-Pierre Dijcks and Martin Gubar examine the best way to implement a big data strategy and what tools you should consider. Coy Yonce discusses how an enterprise can use analytics to evaluate the effectiveness of its own analytics and create a more useful analytics solution. He calls it BI on BI.

As always, we welcome your comments. Please send them to jpowell@tdwi.org.

James E. Powell

Bridging the Analytics Skill Gap with Crowdsourcing

Ryan Hart and Troy Hiltbrand

Troy Hiltbrand is an associate editor of the Business Intelligence Journal as well as deputy chief information officer at the Idaho National Laboratory and an adjunct instructor at Idaho State University.
troy.hiltbrand@inl.gov

Ryan Hart is the business intelligence lead at the Idaho National Laboratory.
ryan.hart@inl.gov

Abstract

Few areas in the economy have generated as much attention recently as big data and advanced analytics, which have the potential to revolutionize the way businesses function in the coming years. However, organizations seeking to implement the advanced analytics that can provide a competitive advantage face a major challenge: Finding the elusive data scientists needed to execute on big data strategy. This article looks at how some businesses are bridging the gap between vision and reality.

Introduction

In 2013, McKinsey Global Institute identified that in the period of slow growth following the Great Recession of 2007–09, only a few areas continue to shine and provide the most potential as game changers for the economy. McKinsey selected areas that may spur hundreds of billions of dollars in gross domestic product (GDP) growth along with a significant number of jobs by 2020 (Lund, et al, 2013). Among the leaders were big data and advanced analytics. The firm's report explained that "Big data and advanced analytics have the potential to raise efficiency and create value in large swaths of the economy" (Manyika, et al, 2011).

Although this is great news for the economy over the next decade, it comes coupled with a great challenge. Another report by the same research organization noted that based on universities' output of graduates with deep analytic skills, by 2018 the economy may face a shortage of 140,000 to 190,000 analysts and 1.5 million managers and analysts. These are the people who can ask the right questions and consume the findings in a way that will drive the economy forward.

Over the past couple of years, the title "data scientist" has become popular across many industries. Tom Davenport and D.J. Patil labeled the position as "The Sexiest Job of the 21st Century" in the *Harvard Business Review* (Davenport and Patil, October 2012). Companies have been clamoring to find individuals who possess the elusive skills that can turn data into money. The hype has parallels to the alchemists of times past, who were purported to be able to alter the chemical composition of lead to make gold. Similarly, today's "data alchemists" have the power to transform, reorganize, and visualize data in a way that allows businesses to make critical decisions, or in some cases, to turn data directly into marketable products, monetizing the value of the data itself.

Companies have been clamoring to find individuals who possess the elusive skills that can turn data into money.

Skills in Three Domains

One of the greatest challenges is that the skills needed to perform data science are not frequently found in a single individual. Instead, the skills span three distinct domains: business, analytics, and technical.

The Business Domain

The business domain is arguably the most important because it differentiates the analytics of today from the backroom, actuarial science work of the past. This domain includes specific skills such as marketing, sales, finance, human resources, supply chain management, and operations. Generally speaking, individuals with refined skills in the business domain are usually running the business and not performing data analytics.

In addition, the business domain is the arena of extroverts who have the soft skills associated with crafting an eloquent message and communicating it in an understandable and consumable manner to key stakeholders. Those with deep analytic skills are generally introverts who love numbers but not necessarily public speaking and stakeholder engagement.

Dan Pink, referring to a psychological study by Adam Grant of the University of Pennsylvania's Wharton School of Management, indicated that the key to success as a data scientist is being an ambivert—"a term coined by social scientists in the 1920s, [ambiverts] are people who are neither extremely introverted nor extremely extroverted" (Pink, 2013). Individuals who can balance their introverted ways with the external focus needed to communicate effectively are best positioned to play the role of a data scientist.

One of the key business skills associated with deep analytics is not only unearthing valuable insight but also transforming the findings into a story that can communicate effectively to key stakeholders. The skill of "data storytelling" will become increasingly desirable as the profession matures and as businesses begin to understand how they can maximize the value of their data when it is paired with the correct contextual narrative.

The business domain is the proverbial eyes, ears, and mouth of the data scientist. It includes seeing business trends, listening to business leaders, and verbally providing direction, vision, and common sense to direct the analytic heavy lifting. Without a sense of business drivers and what business success looks like, a data scientist, no matter how intelligent, will ultimately fail to deliver value.

The Deep Analytics Domain

Traditionally, the deep analytics domain has required a Ph.D. in mathematics and statistics. These individuals can manipulate and massage data, coercing it to give up its secrets. Significant headway has been made by the leading analytic software providers to encode the complex math into simplistic interfaces and functions, putting these capabilities into the hands of the layperson. Although these software packages have improved significantly in the past decade, there is still an art associated with both making sense of the data entering the software and understanding the calculated results...

Big Data Security: Understanding the Risks

Oren Hamami is the director of security strategy for SunGard Availability Services.
oren.hamami@sungard.com

Oren Hamami

Abstract

Big data. It's this year's cloud computing—a transformative technology that is exploding into the mainstream of enterprise IT. Enterprises are wading into the big data pool without fully understanding the associated dangers. Big data introduces new tools, computing models, and classes of information assets to protect, as well as a diverse group of new technical and nontechnical users. As a consequence, traditional approaches to data security and resiliency simply no longer apply.

The potential business value of big data can't be ignored. This article explores big data security issues, including protecting a new kind of (big) information asset, understanding the risks, protecting big data, and where to begin.

Introduction

Big data is a paradigm-shifting, potentially transformative technology. Long used on the fringes of the technology world, it is now exploding into the mainstream of enterprise IT. As with cloud computing, it has become easy for enterprises to start big data projects without a full understanding of the risks.

Big data will require new tools, computing models, and classes of information assets that an enterprise must protect. It also adds a diverse group of users. Traditional approaches to data security and resiliency don't apply to managing big data. A McKinsey study proclaimed that although big data would likely deliver productivity and profit gains of five to six percent, leaders in every sector (not just a few data-oriented IT managers) "will have to grapple with the implications of big data" (Manyika, et al, 2011). Prudent enterprises must understand the risks and tackle them directly if they wish to protect the value of their big data investments.

What—and Where—Is My Data?

As is so often the case with technology trends, the term big data became widely used long before there was a consensus about its meaning. Confusion over what constitutes big data is so great that researchers at the University of St. Andrews in Scotland recently published a paper surveying the most widely used definitions in an effort to synthesize a single, cohesive meaning. This article relies on their description of big data as "the storage and analysis of large and or [sic] complex data sets using a series of techniques including, but not limited to: NoSQL, MapReduce and machine learning" (Ward and Barker, 2013).

Prudent enterprises must understand the risks and tackle them directly if they wish to protect the value of their big data investments.

Even with this specific definition, identifying the use of big data across a company is not a trivial matter. Low storage and computing costs, coupled with easily available, open source tools, allow teams throughout the enterprise to accumulate and analyze significant amounts of data outside the view of IT. Adoption is further accelerated by the availability of cloud-based services for every aspect of big data, from storage to reporting. The democratization of data provides business users with unprecedented access to information, but the lack of visibility can cause significant risk exposure to go unnoticed until problems arise. When assessing big data risk, companies must identify all their big data assets, rather than limit their analysis to those controlled by IT.

Valuing a New Kind of (Big) Information Asset

Once big data assets have been identified, an enterprise typically begins to assess the associated risk by determining the assets' value. Here we encounter our next challenge: how do you value big data? Mature models exist for valuing many traditional types of data. Theft of intellectual property can be linked to associated losses in revenue and competitive advantage, and extensive research has been conducted on the cost of a breach of personally identifiable information (Ponemon Institute, 2013). It is far more difficult to assign a value, even a qualitative one, to a data set whose value, if it has any at all, must be extracted through data mining.

One approach to valuing such an asset is to look to other extractive endeavors, such as oil and gas exploration. Oil and gas companies must be able to value a well without knowing exactly how much oil or gas it will produce during its lifetime. To do this, a well's reserves are classified as proved, probable, or possible based on the likelihood of successful extraction, with further sub-classification based on the difficulty and cost of extraction (SPE Board, 2007).

This approach can be applied to any resource that must be mined, including data. For example, a company that has been mining a data set for years in support of specific business processes with measurable outputs might classify this data as proved and could assign it a fairly accurate value based on the business process it supports. Conversely, a new, unknown data set that is purely exploratory might be classified as probable or possible based on the likelihood it will produce information of value. These are not hard and fast rules, but rather tools to help businesses assign proportional weights to more accurately reflect the expected value of a big data asset over time. Once big data assets are identified and valued, the enterprise can then examine its big data security risks.

New Tools, Old Risks

Much attention has been given to the security deficiencies in the tools commonly used with big data in general, specifically in NoSQL databases. The Cloud Security Alliance (CSA) lists the security of non-relational database tools among its top 10 big data security and privacy challenges (Cloud Security Alliance, 2013), and NoSQL security has been called out by security researchers in both industry (SpiderLabs Radio, 2013)...

BI Experts' Perspective

Business Intelligence as a Career Choice

Dave Schrader, Ph.D., just retired as director of big data marketing for Teradata. His popular "BSI: Teradata" video series on YouTube shows 12 different cases in which data scientists "solve" hard business problems using a mix of big data and traditional BI technologies. He is a board member of the Teradata University Network, which provides teaching tools and cases about BI for faculty and students (www.teradatauniversitynetwork.com). His Ph.D. in computer science is from Purdue and he worked at Teradata for 22 years.
drdaveschrader@gmail.com

Ronald S. Swift is managing partner at International Management Consulting (IMC) and "executive-in-residence" at the Terry College of Business at the University of Georgia.
ronswift@uga.edu

Coy Yonce is the product owner for software solutions from EVtechnologies, a services and software partner with SAP. Coy also advises organizations on implementing and utilizing analytics to help them understand the effectiveness of their analytics solutions.
coy@evtechnologies.com

Dave Schrader, Ron Swift, and Coy Yonce

Q **Susan Stephenson is the BI director for a national auto parts company. She has 10 years of BI experience, with the last three as director. Susan keeps up to date by reading, watching online seminars, and attending BI conferences.**

After church last week, Susan was chatting with Craig Mercer, who teaches computers and information systems at the local college. At the end of their conversation, Craig asked Susan if she would be willing to be the speaker at the school's fall management information systems (MIS) banquet. About 100 students and faculty will attend, and she will speak after the social hour and dinner and before scholarships are awarded. Her talk should be about 15-20 minutes long.

Craig further explained that the current curriculum requires students to complete course work in programming languages, databases, systems analysis and design, telecommunications and networking, project management, and computing architecture, but there is little on BI. That is why he would really like Susan to speak.

When she asked what Craig would like her to cover, he responded, "What BI is, why it is important, what the career opportunities are, how to prepare for and start a BI career, and what technical and 'soft' skills a graduate needs to be successful." He also said that students want know what's currently "hot" in BI and what is likely to be important in the future.

If you were Susan, what would you tell the students?

A **DAVE SCHRADER**

Susan should begin by explaining the differences between traditional BI and the hot new areas. Traditional BI is about the use of traditional data by back-office business analysts—with the help of IT—to track product sales, inventory, costs, customers, promotions, and campaigns. This often includes dashboards and scoreboards as well as forecasting of demand and markets.

Some of this work can be done in Excel, but much of it uses tools such as MicroStrategy, Cognos, Business Objects, or perhaps Tableau

Software to access a data warehouse provided by Teradata, IBM, Oracle, or Microsoft. Current state-of-the-art projects using traditional technologies include putting dashboards and alerts on mobile devices (so people can work more hours!) as well as placing data and BI tools in the cloud. Of course, there are ever-increasing amounts of data to collect each year, and data is arriving at a higher velocity (sometimes "near real time," meaning seconds, or even "real time," meaning milliseconds, between data creation and ingestion).

What's hotter than hot at the moment is the entire area of "big data," which is a bit of a misnomer. Although big data partly includes new data types such as voice, text, and sensors ("the Internet of things"), it's more about the ability to use algorithms to transform raw, non-traditional data types into useful additional insights—finding the signal in the noise, because many of these data inputs contain low-density information. For example, these systems might transform a raw voice file into "what did the person at the call center say?" (voice to text) or "how irritated was the person?" (sentiment score). It might mean linking the symptoms a doctor types into patient notes (text) with other traditional data to find out if the doctor is over-prescribing or allowing off-label uses.

The most interesting BI trend is the need for new and different discovery tools, as well as the people who are able to drive them. The new paradigm for discovery is "grab and go" for fast insights, often using dirty data that's not completely modeled and may not even have known schemas. This is exploratory work, often fast-fail. Only after you find interesting insights worth saving do you go back and perform the rigorous work associated with capturing, modeling, and storing information in a data warehouse for wider use within the enterprise.

The most interesting BI trend is the need for new and different discovery tools, as well as the people who are able to drive them.

A new kind of data analyst—the data scientist—is in extremely high demand to do this discovery work. In fact, data scientists can command starting salaries of $90,000 with stock options—if they have the right skills. The role requires a blend of curiosity, statistics expertise, computer science, business domain knowledge, and communication skills. Almost no one has all five of these characteristics, but students in MIS who want to increase their marketability can acquire some of these skills, partly at school but also by watching online Webinars, reading books and articles, and finding faculty who understand this trend.

Specific tips include:

1. If you are an MIS major, choose a minor that provides domain expertise and take any classes that have "Computational" in the title. For example, "Computational Chemistry" or "Computational Biology" would provide background for a data scientist role with a pharmaceutical company doing genome discovery work.

2. Make sure you are well grounded in some of the hot areas within computer science such as Web analytics, visualization, or wearable devices/sensors.

3. Read any of the new books on big data and big data analytics. My personal favorite is *Taming the Big Data Tidal Wave* (Wiley, 2012) by my friend Bill Franks, Teradata's chief analytics officer.

4. Read anything—anything at all—by Tom Davenport. His landmark book *Competing on Analytics* (Harvard Business Review Press, 2007) is a classic. See also his white paper "The Rise of Data Discovery" (available on www.asterdata.com).

5. To see some examples of big data in action, take a look at my "BSI: Teradata" series, in which business scenario investigators (BSI) use data and BI to solve cases such as "The Case of the Dropped Mobile Calls" or "The Case of the Tainted Lasagna."...

BI StatShots

TDWI Technology Survey: Cloud BI

The Las Vegas 2014 Technology Survey asked attendees to answer a few questions about cloud BI. Although cloud BI has been slow in gaining adoption, we are starting to see more interest in it, so we wanted to understand what companies that attended the conference were thinking in terms of adoption. We ran a similar survey last fall at our Boston conference and wanted to compare results between the two groups. We had 113 respondents, with varying numbers answering different questions. This should therefore be considered simply a "quick pulse" survey. Here are some findings:

Companies are making the cloud part of their information architecture. We asked attendees if they see the public cloud becoming part of their information architecture (Figure 1). Although only 18% reported that it already is, 21% said it would be in the future and 36% said it was possible that it would be part of their information ecosystem. Only 25% responded either "don't think so" or "never."

Companies are considering public cloud BI. We also asked attendees if they make use of the public cloud for BI and analytics (Figure 2). Although 26% claimed they would never use the cloud for BI, 37% responded that while they weren't using it now, they were thinking about it. These numbers are similar to those we saw in the Boston survey. Additionally, 18% were already using the public cloud for BI and analytics, many in a hybrid kind of deployment where they utilize cloud BI services as well as on-premises BI and analytics solutions.

Top cloud BI challenges still include security. For those attendees who were using or thinking about using cloud BI, we asked them to rate a set of challenges (Figure 3). "Perceived security concerns" was rated the highest with a score of 4.3. This has always been a top concern (real or imagined) and continues to be one. Interestingly, security was followed by extended governance issues (3.9), data integration issues (3.7), and internal politics (3.7). The fact that governance and integration issues ranked highly (and were also ranked high in the fall survey) suggests that respondents are becoming more informed about cloud BI. This is a good thing.

Do you see the public cloud becoming part of your information architecture?

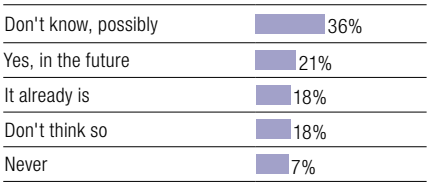


Figure 1.

Do you make use of the public cloud for BI and analytics? Please select the statement that best applies to your company.

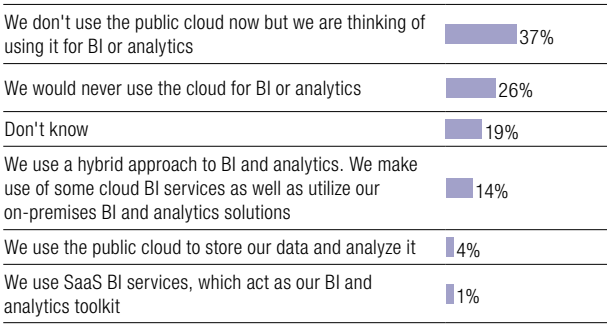


Figure 2.

What is or was your biggest challenge with getting your organization to move to the cloud for BI and analytics? Please rate each on a scale from 1–5, where 1 is not at all a challenge and 5 is a major challenge.

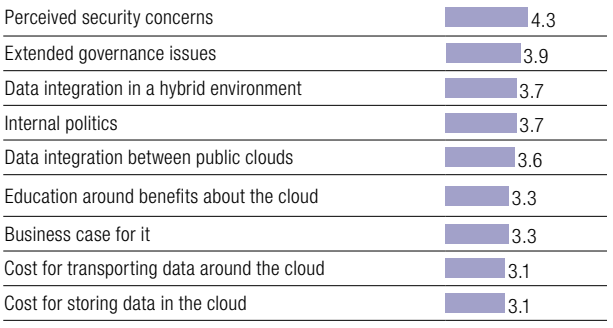


Figure 3.