STRATEGIC GUIDE TO

Big Data Analytics

Five Analytics Trends to Exploit

Big Data 101: What You Should Know

Key Questions to Get You Started

Don't Forget About Data Security

The Top 5 Big Data Challenges

Coping with the Talent Shortage

FROM THE EDITORS OF

CIO

BUSINESS TECHNOLOGY LEADERSHIP

Table of Contents

Adapted from articles published at CIO.com

EDITOR'S NOTE

Competing on Analytics

IT leaders have been on a quest to obtain competitive advantage through technology for 25 years. And that quest is about to enter a new chapter: the era of big data analytics.

One survey found that 70% of respondents can envision a "killer application" for big data that would be "very useful" or "spectacular" for their business. The catch: Most chose not to disclose what that application would be because it would provide a competitive advantage.

To get started, you'll first need to ask blue-sky questions of your business execs— such as "if only we knew..." or "if we could predict..."— so you'll know what information or answers would qualify as "spectacular."

Then you'll have to figure out what internal and external data could help. Third, you'll need to find "data scientists" who can help you make sense of that data, which won't be easy.

This report is intended to help guide you along the way.

Mitch Betts

Editorial

Editor
Mitch Betts
mbetts@cio.com

Contributors
David F. Carr
Joab Jackson
Thor Olavsrud
Bob Violino

Editorial Management
Brian Carlson
Maryfran Johnson
Dan Muse

Image Credits
AllyB208/Fotolia [cover]
AGSAndrew/Fotolia [page 3]
Andres Rodriguez/Fotolia [page 16]

Five Business Analytics Trends – And How To Exploit Them

Advances in analytic technologies and business intelligence are allowing CIOs to go big, go fast, go deep, go cheap and go mobile with business data.

Current trends center as much on tackling analytics challenges as they do on taking advantage of opportunities for new business insights. For example, technologies for managing and analyzing large, diverse data sets are arriving just as many organizations are drowning in data and struggling to make sense of it. Still, many of the cost and performance trends in advanced analytics mean companies can ask more complicated questions than ever before and deliver more useful information to help run their businesses.

In interviews, CIOs consistently identified five IT trends that are having an impact on how they deliver analytics: the rise of big data, technologies for faster processing, declining costs for IT commodities, proliferating mobile devices and social media.

1. Big Data

Big data refers to very large data sets, particularly those not neatly organized to fit into a traditional data warehouse. Web crawler data, social media feeds and server logs, as well as data from supply chain, industrial, environmental and surveillance sensors all make corporate data more complex than it used to be.

Although not every company needs techniques and technologies for handling large, unstructured data sets, Verisk Analytics CIO Perry Rotella thinks all CIOs should be looking at big data analytics tools. Verisk, which helps financial firms assess risk and works with insurance companies to identify fraud in claims data, had revenues of more than $1 billion in 2010.

Technology leaders should adopt the attitude that more data is better and embrace overwhelming quantities of it, says Rotella, whose business involves "looking for patterns and correlations between things that you don't know up front."

70%

of respondents can envisage a **"killer application" for big data** that would be **"very useful" or "spectacular"** for their business. The majority chose **not to disclose** what that application would be because it would provide a **competitive advantage**.

Source: AIIM survey of 345 information professionals, 2012

Big data is an "explosive" trend, according to Cynthia Nustad, CIO of HMS, a firm that helps contain healthcare costs for Medicare and Medicaid programs, as well as private businesses. Its clients include health and human services programs in more than 40 states and more than 130 Medicaid managed care plans. HMS helped its clients recover $1.8 billion in costs in 2010 and save billions more by preventing erroneous payments. "We're getting and tracking so much material, both structured and unstructured data, and you don't always know what you're looking for in it," Nustad says.

One of the most talked about big data technologies is Hadoop, an open-source distributed data processing platform originally created for tasks such as compiling web search indexes. It's one of several so-called "NoSQL" technologies (others include CouchDB and MongoDB) that have emerged to organize web-scale data in novel ways.

Hadoop is capable of processing petabytes of data by assigning subsets of that data to hundreds or thousands of servers, each of which reports back its results to be collated by a master job scheduler. Hadoop can either be used to prepare data for analysis or as an analytic tool in its own right. Organizations that don't have thousands of spare servers to play with can also purchase on-demand access to Hadoop instances from cloud vendors such as Amazon.

Nustad says HMS is exploring the use of NoSQL technologies, although not for its massive Medicare and Medicaid claims databases. These contain structured data and can be handled with traditional data warehousing techniques, and it makes little sense to depart from traditional relational database management when tackling problems for which relational technology is the tried and true solution, she says. However, Nustad can see Hadoop playing a role in fraud and waste analytics, perhaps analyzing records of patient visits that might be reported in a variety of formats.

Horizontal & Vertical Applications

Big Data technology can be deployed for business processes such as the following:

- Customer relationship management (sales, marketing, customer service)
- Supply chain and operations
- Administration (finance and accounting, human resources, legal)
- Research and development
- Information technology management
- Risk management

In addition, big data technology can be used for industry-specific applications such as the following:

- Logistics optimization in the transportation industry
- Price optimization in the retail industry
- Intellectual property management in the media and entertainment industry
- Natural resource exploration in the oil and gas industry
- Warranty management in the manufacturing industry
- Crime prevention and investigation in local law enforcement
- Predictive damage assessments in the insurance industry
- Fraud detection in the banking industry
- Patient treatment and fraud detection in the healthcare industry

Source: IDC, 2012

Among the CIOs interviewed for this story, those who had practical experience with Hadoop, including Rotella and Shopzilla

CIO Jody Mulkey, are at companies that provide data services as part of their business.

"We're using Hadoop for what we used to use the data warehouse for," Mulkey says, and, more importantly, to pursue "really interesting analytics that we could never do before." For example, as a comparison shopping site, Shopzilla accumulates terabytes of data every day. "Before, we would have to sample data and partition data- it was so much work just to deal with the volume of data," he says. With Hadoop, Shopzilla is able to analyze the raw data and skip the in-between steps.

Good Samaritan Hospital, a community hospital in Southwest Indiana, is at the other end of the spectrum. "We don't have what I would classify as big data," says CIO Chuck Christian. Nevertheless, regulatory requirements are causing him to store whole new categories of data such as electronic medical records in great quantities. Doubtless there is great potential to glean healthcare quality information from the data, he says, but that will probably happen through regional or national healthcare associations rather than his individual hospital. It's unlikely he'll invest in exotic new technologies himself.

VERBATIM

"Many organizations will struggle to deploy big data applications until they improve their current levels of information management and reduce content chaos."

Source: "Big Data: Extracting Value from Digital Landfills," study by AIIM, 2012

John Ternent, CIO at Island One Resorts, says that whether his analytic challenges are driven by big data "depends on how capital your B and D are." But he's seriously considering using Hadoop instances in the cloud as an economical way of running complex mortgage portfolio analytics for the company, which manages eight timeshare resort properties across Florida. "That's a potential solution to a very real problem we have now," he says.

2. Business Analytics Get Faster

Big data technologies are one element of a larger trend toward faster analytics, says University of Kentucky CIO Vince Kellen. "What we really want is advanced analytics on a hell of a lot of data," Kellen says. How much data one has is less critical than how efficiently it can be analyzed, "because you want it fast."

The capacity of today's computers to process much more data in memory allows for faster results than when searching through data on disk—even if you're crunching only gigabytes of it.

Although databases have, for decades, improved performance with caching of frequently

With Big Data, Semantics Really Matter

Getting ready to enter the much-hyped world of big data analytics? Here's some advice from **David Saul, senior vice president and chief scientist at State Street**, a major financial services firm.

One of the keys to **extracting useful business insights from unstructured data**–audio, video, images, text, tweets, wikis, forums and blogs–is to **create a semantic data model as a layer that sits on top of your data** and helps you make sense of everything, Saul says.

The traditional approach is to pull data from disparate sources into a single repository for analysis, but Saul says that's too time-consuming for big data sets. To make the process more efficient, State Street established a semantic layer that allows data to stay where it is but provides descriptive information about it.

For example, if State Street needs "a **risk profile** for all the exposures we have to a particular entity or geography," a semantic description of the various information sources "means we can **quickly pull together** a consolidated risk profile or an ad hoc request," Saul says.

He adds that using the semantic layer means State Street doesn't have to "go back and redo all of our legacy systems and database definitions. It lays on top of that, so it's **much less disruptive** than another type of technology that would require us to go to a clean slate."

State Street built a set of tools to help end users–generally a **business person who understands the data well**, rather than a programmer or database administrator–write the semantic data description. "For years we've talked about being able to blur the line that exists between IT and the business and having business be able to have tools where they can **more clearly express requirements**. This is a step in that direction," Saul says.

Thor Olavsrud, CIO.com, March 2012

accessed data, now it's become more practical to load entire large datasets into the memory of a server or cluster of servers, with disks used only as a backup. Because retrieving data from spinning magnetic disks is partly a mechanical process, it is orders of magnitude slower than processing in memory.

Rotella says he can now "run analytics in seconds that would take us overnight five years ago." His firm does predictive analytics on large data sets, which often involves running a query, looking for patterns, and making adjustments before running the next query. Query execution time makes a big difference in how quickly an analysis progresses. "Before, the run times would take longer than the model building, but now it takes longer to build the model than to run it," he says.

Columnar database servers, which invert the traditional row-and-column organization of relational databases, address another category of performance

requirements. Instead of reading entire records and pulling out selected columns, a query can access only the columns of interest—dramatically improving performance for applications that group or measure a few key columns.

Ternent warns that the performance benefits of a columnar database come only with the right application and query design. "You have to ask it the right question the right way for it to make a difference," he says. Meanwhile, he says, columnar databases only really make sense for applications that must handle over 500 gigabytes of data. "You have to get a certain scale of data before columnar makes sense because it relies on a certain level of repetition" to achieve efficiencies."

To improve analytics performance, hardware matters, too. Allan Hackney, CIO at the insurance and financial services giant John Hancock, is adding GPU chips-the same graphical processors found in gaming systems-to his arsenal. "The math that goes into visualizations is very similar to the math that goes into statistical analysis," he says, and graphics processors can perform calculations hundreds of times faster than conventional PC and server processors. "Our analytic people love this stuff."

VERBATIM

"Some organizations will discover that big data solutions create new opportunities to launch additional lines of business that are focused on selling information as well as analytic services based on the data."

Source: IDC, 2012

3. Technology Costs Less

Along with increases in computing capacity, analytics are benefitting from falling prices for memory and storage, along with open-source software that provides an alternative to commercial products and puts competitive pressure on pricing.

Ternent is an open-source evangelist. Prior to joining Island One, he was vice president of engineering for Pentaho, an open-source business intelligence company, and worked as a consultant focusing on BI and open source. "To me, open source levels the playing field," he says, because a mid-sized company such as Island One can use R, an open-source application, instead of SAS for statistical analysis.

Once, open-source tools were available only for basic reporting, he says, but now they offer the most advanced predictive analytics. "There is now an open-source player across just about the entire continuum, which means there's tooling available to whoever has the gumption to go and get it."

HMS' Nustad sees the changing economics of computing

altering some basic architectural choices. For example, one of the traditional reasons for building data warehouses was to bring the data together on servers with the computing horsepower to process it. When computing power was scarcer than it is today, it was important to offload analytic workloads from operational systems to avoid degrading the performance of everyday workloads. Now, that's not always the right choice, Nustad says.

"With hardware and storage so cheap today, you can afford to juice up those operational systems to handle a BI layer," she says. By factoring out all the steps of moving, reformatting and loading data into the warehouse, analytics built directly on an operational application can often provide more immediate answers.

Hackney observes, however, that although the price/performance trends are helpful for managing costs, potential savings are often erased by increased demands for capacity. "It's like running in place," he says. While John Hancock's per unit cost for storage dropped by 2 to 3 percent this year, consumption was up 20 percent.

Big Data: Problem or Opportunity?

Only 30% of IT pros consider big data a problem. There's no doubt that big data presents technical challenges due to its volume, variety and velocity. Data volume alone is a show-stopper for some organizations.

The vast majority (70%) considers big data an opportunity. Through exploratory, detailed analyses of big data, a user organization can discover new facts about their customers, markets, partners, costs, and operations–then use that information for business advantage.

Source: The Data Warehousing Institute survey of 325 IT professionals, 2011

4. Everyone's Mobile

Like nearly every other application, BI is going mobile. For Nustad, mobile BI is a priority "because everybody wants Nustad herself wants access to reports on whether her organization is meeting its service level agreements "served up on my iPad when I'm very mobile and not at my desk." She also wants to deliver mobile access to data for her firm's customers, to help them monitor and manage healthcare expenses. It's "a customer delight feature that was not demanded five years ago, but is demanded today," she says.

For CIOs, addressing this trend has more to do with creating user interfaces for smartphones, tablets and touch screens than it is about sophisticated analytic capabilities. Maybe for that reason, Kellen dismisses it as fairly easy to address. "To me, that's kind of trivial," he says.

Rotella doesn't think it's that simple. "Mobile computing affects everyone," he says. "The number

of people doing work off of iPads and other mobile devices is exploding. That trend will accelerate and change how we interact with our computing resources in an enterprise." For example, Verisk has developed products to give claims adjusters access to analytics in the field, so they can run replacement cost estimates. That's a way to "leverage our analytics and put it at the fingertips of the people that need it," he says.

What makes this challenging is how much more quickly technology changes, Rotella says. With multiple device operating systems in play, "we're trying to understand how to best leverage our development so we're not writing these things three, four, five times over," he says.

On the other hand, the requirement to create native applications for each mobile platform may be fading now that the browsers in phones and tablets are more capable, says Island One's Ternent. "I'm not sure I'd invest in a customized mobile device application if I can just skin a Web-based application for a mobile device."

5. Social Media in the Mix

With the explosion of Facebook, Twitter and other social media, more companies want to analyze the data these sites generate. New analytics applications have

Big Data: How to Get Started

Ask blue-sky questions of your business such as "if only we knew..." or "if we could predict..." or "if we could measure..." Consider how useful that might be to the business before thinking about how it can be done or at what cost.

Play those questions off against the **data you already have**, data you could collect, or data that you could get elsewhere.

Include in your thinking **structured** transactional data, **semi-structured** logs and files, and **text**-based or rich media content.

Incoming **communications from your customers**, outbound communications to your customers, and what customers (or employees) are saying about you on social sites can all be useful for monitoring sentiment, heading off issues and analyzing trends.

Consider **high-volume streams** such as telemetry, geolocation, voice, video, news feeds, transactions, Web clicks, or any combination of these.

If your content is currently **"digital landfill"** spread across disparate file shares and content systems, consider **how this could be rationalized prior to any big data projects**.

Source: "Big Data: Extracting Value from Digital Landfills," study by AIIM, 2012

emerged to support statistical techniques such as natural language processing, sentiment analysis, and network analysis that aren't part of the typical BI toolkit.

Because they're new, many social media analytics tools are available as services. One prominent example is Radian6, a software-as-a-service product recently purchased by Salesforce.com. Radian6 presents a dashboard of brand mentions-tagged positive, negative, or neutral-based on Twitter feeds, public Facebook posts, posts and comments on blogs and discussion board conversations. When purchased by the marketing and customer service departments who use them, such tools may not require heavy IT involvement. Still, University of Kentucky's Kellen believes he needs to pay attention to them. "My job is to identify these technologies, see what the match is for the organization in terms of competitiveness, and start educating the right people," he says.

For example, monitoring student posts on social media could help faculty and administrators learn earlier when students are having academic trouble. IT developers should also build alerts generated by social media analytics into applications

Don't Forget About Securing Big Data

Collecting all this data and making it more accessible also means organizations need to be serious about securing it. And that requires **thinking about security architecture from the beginning**, says David Saul, chief scientist at State Street, a financial services provider that serves global institutional investors.

"I believe the biggest mistake that most people make with security is they leave thinking about it until the very end, until they've done everything else: architecture, design and, in some cases, development," Saul says. "That is always a mistake."

Saul says that State Street has implemented an enterprise security framework in which **every piece of data in its stores includes with it the kind of credentials required to access that data**.

"By doing that, we get better security," he says. "We get much finer control. We have the ability to do reporting to **satisfy audit requirements**. Every piece of data is considered an asset. Part of that asset is **who's entitled to look at it, who's entitled to change it, who's entitled to delete it**, etc. Combine that with **encryption**, and if someone does break in and has free reign throughout the organization, once they get to the data, there's still another protection that keeps them from getting access to the data and the context."

Thor Olavsrud, CIO.com, March 2012

for responding to those events, he says.

While John Hancock's efforts in this area are "nascent," according to Hackney, he envisions a role for IT in correlating the data provided by a social analytics service with corporate data. For example, if the social media data shows comments about the company in the Midwest are becoming more negative, he would want to see if the company has made price or policy changes in that region that might explain the trend.

Finding such correlations could make a big difference in getting company leaders to believe in the return on investment of social media, Hackney says.

"In my industry, everybody's an actuary, everyone's looking for the numbers—they don't take anything on belief."

David F. Carr, CIO.com, March 2012

Does Big Data = Big Spender?

Big data is a powerful lure, but that lure can lead you into **an expensive trap** if you don't plan carefully.

"Big data has big spending risks," says Jeff Muscarella, IT spend management consultant with NPI Financial. Muscarella warns that **big data projects can easily ring up seven-figure price tags** after you finish paying for the hardware, software and services, and sometimes the glowing business cases presented by vendors lose their luster when you look closely. "A lot of times, when you pull them apart, they're not as rosy as they seem," he says.

That's not to say that harnessing the power of big data is a mistake, Muscarella explains. But it does mean that organizations need to **start by gathering real data** on how a big data project will **benefit the business**.

"It's new technology solving a business problem that we often haven't proved," Muscarella says. "The business is going to be coming to them with all sorts of **half-baked ideas** for what they can do with big data. They have to ask: Will it really drive revenue? How, and for how long? What will it take to build it? They need to make sure they **have a crisp focus on the mission**; that it is going to have a return on investment."

Muscarella recommends starting with open-source tools like Apache Hadoop and build a test case. "Pick something that's manageable. **Start on a small scale to prove your hypothesis.**"

"Don't get trapped into building the infrastructure yet," he adds. "Prove it first and then go back and architect your solution. Assume that however you solve the problem, you're probably going to throw it away and start over. That's OK because at least you proved the business need before you spent a lot of money."

Thor Olavsrud, CIO.com, March 2012

Big Data 101: What CIOs Should Know

1. You will need to think about big data.

Big data analysis got its start from the large Web service providers such as Google, Yahoo and Twitter, which all needed to make the most of their user generated data. But enterprises will big data analysis to stay competitive, and relevant, as well.

You could be a really small company and have a lot of data. A small hedge fund may have terabytes of data, says Jo Maitland, GigaOm research director for big data. In the next couple of years, a wide number of industries—including health care, public sector, retail, and manufacturing—will all financially benefit by analyzing more of their data, consulting firm McKinsey and Company anticipated in a recent report.

There is an air of inevitability with Hadoop and big data implementations, says Eric Baldeschwieler, chief technology officer of Hortonworks, a Yahoo spinoff company that offers a Hadoop distribution. It's applicable to a huge variety of customers. Collecting and analyzing transactional data will give organizations more insight into their customers' preferences. It can be used to better inform the creation of new products and services, and allow organizations to remedy emerging problems more quickly.

2. Useful data can come from anywhere (and everywhere).

You may not think you have petabytes of data worth analyzing, but you will, if you don't already. Big data is collected data that used to be "dropped on the floor," Baldeschwieler says.

DEFINITIONS

"Big data analytics is the application of advanced analytic techniques to very big data sets."

Source: "Big Data Analytics" study by The Data Warehousing Institute, 2011

"Big data is a new generation of technologies and architectures designed to extract value economically from very large volumes of a wide variety of data by enabling high-velocity capture, discovery and/or analysis."

Source: IDC, 2012

Big data could be your server's log files, for instance. A server keeps track of everyone who checks into a site, and what pages they visit when they are there. Tracking this data can offer insights into what your customers are looking for. While log data analysis is nothing new, it can be done don to dizzying new levels of granularity.

Another source of data will be sensor data. For years now, analysts have been speaking of the Internet of Things, in which cheap sensors are connected to Internet, offering continual streams of data about their usage. They could come from cars, or bridges, or soda machines."The real value around the devices is their ability to capture the data, analyze that information and drive business efficiencies," says Microsoft Windows Embedded General Manager Kevin Dallas.

Big Data Drivers

Analysis of....

1. Operational data
2. Online customer data
3. Sales transactions data
4. Machine or device data

And...

5. Service innovation

Source: IDC survey of 2,699 data professionals, 2012

3. You will need new expertise for big data.

When setting up a big data analysis system, your biggest hurdle will be finding the right talent who knows how to work the tools to analyze the data, according to former Forrester Research analyst James Kobielus.

Big data relies on solid data modeling. Organizations will have to focus on data science, Kobielus says. They have to hire statistical modelers, text mining professionals, people who specialize in sentiment analysis. This may not be the same skill set that todays analysts versed in business intelligence tools may readily know.

Such people may be in short supply. By 2018, the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts with the know-how to use the analysis of big data to make effective decisions, McKinsey and Company estimated.

Another skill you will need to have on hand is the ability to wrangle the large amounts of hardware needed to store and parse the data. Managing 100 servers is a fundamentally different problem than handle 10 servers, Maitland pointed out. You may need to hire a few supercomputer administrators from the local university or research lab.

4. Big data doesn't require organization beforehand.

CIOs who are used to rigorously planning out every sort of data

Top 5 Big Data Challenges

1. Deciding what data is relevant
2. Cost of technology infrastructure
3. Lack of skills to analyze the data
4. Lack of skills to manage big data projects
5. Lack of business support

Source: IDC survey of 2,699 data professionals,

that would go into an Enterprise Data Warehouse (EDW) can breathe a little easier with big data setups. Here, the rule is, collect the data first, and then worry about how you will use it later.

With a data warehouse, you have to lay out the data schema before you can start laying in the data itself. "This basically means you have to know what you are looking for beforehand," says Jack Norris, vice president of marketing for MapR. As a result, "you are flattening the data and losing some of the granularity," he says. "Later on, if you change your mind, or want to do a historical analysis, you've limited yourself." "You can use a [big data repository] as a dumping ground, and run the analysis on top of it, and discover the relationships later," Norris says. Many organizations may not know what they are looking for until after they've culled the data, so this kind of freedom "is kind of big deal," he says.

5. Big data is not only about Hadoop.

When people talk about big data, most times they are referring to the Hadoop data analysis platform. "Hadoop is a hot button initiative, with budgets and people being assigned to it" in many organizations, Kobielus pointed out. Ultimately, however, you may go with other software.

Recently legal research giant LexisNexis, no slouch at big data analysis itself, open-sourced its own platform for analysis, HPCC Systems. MarkLogic has also outfitted its own database for unstructured data, the MarkLogic Server, for big data style jobs as well. Another tool gaining favor is the Splunk search engine, which can be used to search and analysis data generated by machines, such as the log files from a server. "Whatever data you can extract from your logs, there is a good chance that Splunk can help," notes Curt Monash of Monash Research.

Joab Jackson, CIO.com, May 2012

VERBATIM

"Big data is not just big. It's also diverse data types and streaming data."

Source: "Big Data Analytics" study by The Data Warehousing Institute, 2011

Coping with the Big Shortage Of Big Data Talent

Just as historical reports alone aren't sufficient for making corporate decisions—executives want business intelligence to identify current and future trends—IT staffers need to know more about BI than how to run a data warehouse or build a dashboard. That puts CIOs in a bind, according to industry experts, who have raised alarms about a data analytics skills deficit.

For example, a report released last spring by the McKinsey Global Institute predicts that by 2018, the United States could lack 140,000 to 190,000 workers with deep analytical skills and another 1.5 million managers and analysts who know how to use analysis of large data sets to make effective decisions.

"We see our BI leader as being the catalyst to drive our organization away from pure historical reporting to true inferential analysis, says Greg Meyers, vice president of global IT at Biogen Idec, a $5 billion biotech company. "This is both a technical and change management challenge."

Yet despite continued high U.S. unemployment rates, there's a BI talent shortage says Boris Evelson, analyst at Forrester Research. "Every single client I talk to tells me they are struggling with finding and retaining BI talent."

To fill the gap, CIOs are competing for workers with strong math skills, proficiency working with massive databases and with emerging database technology as well as with expertise in search, data integration, and other areas such as business knowledge, Evelson says. In fact, he says, business

"The big deal is accumulating knowledge of best practices and lessons learned from successful and failed implementations."

knowledge, such as understanding processes, customers and products, "is at least equally as important as the tech skills."

IT leaders are thinking about how to get the needed analytics talent now as well as developing the pipeline for technologists with the right skills for the future.

Business Analytics Education Gap

Foote Partners, a research and advisory firm that tracks IT skills demand and pay levels, pins the gloomy outlook for BI talent on a low supply of young workers in roles such as architects, modelers, integrators, analysts and developers. The finding is preliminary, says David Foote, the company's co-founder, CEO and chief research officer, but he suggests that one problem is that many colleges and universities haven't yet risen to the challenge of teaching the skills that are potentially needed for analytics jobs.

He cites the need for government and industry partnerships with academia, such as the U.S. Cybersecurity Challenge, that use online competitions and incentives to attract students to possible careers in information security. "The same sort of thing needs to happen for analytics/statistics/BI," Foote says. "Fill the pipeline with students eager to enter the field and focus on careers."

Unfortunately, academic credentials, like a class or even a related degree, go only so far. Qualified workers require several years of experience to understand how to deal with "real world" BI challenges. "One can learn the technical skills needed for BI in a six-month class; that's not a big deal," Evelson says. "The big deal is accumulating knowledge of best practices and lessons learned from successful and failed implementations."

Wanted: Real World Analytics Experience

So IT executives are scouring the country to find people with the data analytics skills they need now. Douglas Menefee is CIO at Schumacher Group, a privately held company that provides emergency room management services to hospitals nationwide. He wants to hire database extract,

transform and load (ETL) developers, as well as presentation analysts "who know how to tell a story with data."

"Both need to have very strong critical thinking skills and need to be able to draw on asking the hard questions," Menefee says. The ETL developers need "strong math and logic matching skills," he says, while the presentation developers "need to be able to use right and left brain thinking"-in other words, be both logical and creative. "We want them to use creativity to tell a graphical story."

Because projects change very quickly depending on the "fire of the day," Menefee says, Schumacher looks for individuals who also are experienced with agile development and can adapt to change easily.

The hiring cycle takes three to six months, sometimes preventing the company from moving as quickly as it would like on projects, Menefee says. If necessary, consultants fill the gap, such as when the company was building a BI center of excellence and needed architecture and design expertise. Hitachi Consulting "worked with us a couple of years until we were [internally] staffed," Menefee says.

The Most Difficult Big Data Skills to Find:

1. Advanced analytics, predictive analytics
2. Complex event processing
3. Rules management
4. Business intelligence tools
5. Data integration

Source: IDC/Computerworld, 2012

Adding to Menefee's challenge is the company's Lafayette, La., location. It's hard to convince people who aren't already familiar with Lafayette to relocate. He concentrates on hiring locals, as well as people who want to move back to Louisiana. Schumacher Group has also taken advantage of job recruitment, hiring and training services provided through the Louisiana state government's FastStart workforce program.

Like Menafee, Meyers at Biogen Idec wants to hire staff with data warehousing, ETL and reporting experience. That's relatively easy, he says. But he also wants these workers who know how to elicit details from users about the metrics they use to make decisions and the unanswered questions they have that data can help them to find. Identifying them is more challenging. He thinks he can recruit them from other companies, where they've done similar work.

Planning for the Future

Meyers says he can find people to staff current projects. "These skill sets might have been fine for the

"We want them to use creativity to tell a graphical story."

past several years, but to truly use BI as a competitive advantage you have to focus decision support on predicting the future—not simply reporting the past," he says. Menefee adds: "Our next generation of skills will be blended heavily on the business side with statistical modeling and quantitative analysis." However, he adds, "these skills probably won't live within IT."

He is making an effort to develop a compatible IT workforce in part by working with the University of Louisiana, which graduates 70 students annually with computer science master's degrees. He serves on a course curriculum committee at the university's Lafayette campus, which gives "feedback on gaps between what is being taught and what our business needs are," he says. He's also working with the school to develop an internship program.

Pacific Coast Companies provides business and IT services to a dozen subsidiaries that supply building products and related services. CIO Mike O'Dell needs workers with statistical and analytical skills as well as knowledge of economics "and the wisdom to understand causal relationships in the data," he says. "The types of projects we are working on are focused on making the people we have more effective, from the executive to the salesman to the shop floor."

The need for business skills, in fact, is driving O'Dell's staffing and recruitment strategy. He taps business people internally and teaches them technology. He also recruits local college students with technology skills and teaches them about the business.

"Understanding the business is the more complex side of the equation, so the best way to develop those skills is to expose technology-oriented people to the business and put them in the field and let them learn," O'Dell says.

Bob Violino, CIO.com, March 2012

VERBATIM

"The application of big data technology will fall into two primary categories: doing more efficiently tasks that have been done for years; and doing completely new things that were never before possible, driving up long-term strategic organizational value. Identify opportunities to apply big data to both."

Source: IDC, 2012