The Path to Payoff on Big Data Analytics

The hullabaloo over the promised business benefits of big data technologies may have some companies thinking they're the only ones not invited to the party. But analytics deployments have yet to become regular fare.

A Suitable Pairing

THE COMPONENTS OF the phrase *big data analytics* seem to fit together like hand and glove. For example, in a 2013 survey conducted by The Data Warehousing Institute, analytics was far and away the most-cited response to a question about the business and technology tasks most likely to improve inside organizations that harness big data. It was chosen by 61% of the 461 respondents; the closest follower was selected by just 39%.

Another data point: Gartner predicted in February 2014 that 25% of large, global businesses will adopt big data analytics tools for at least one data security or fraud detection application by 2016—up from 8%. Companies that do so can give themselves a better chance to "stay ahead of malicious actors," said Gartner analyst Avivah Litan. In healthcare, meanwhile, big data analytics could provide the scientific means to help foster better treatments for

patients. "This is completely about outcomes, outcomes, outcomes," said Lisa Khorey of the UPMC health system in discussing its big data initiative as part of the first story in this guide.

But overall, it's still early days for big data analytics deployments: Forty-five percent of the respondents to the TDWI survey said their organizations didn't have any big data strategies in place. And all the promise is accompanied by plenty of possible pitfalls. The three articles here offer advice to help you find the former and avoid the latter. First, we look at the long-term thinking that UPMC and financial services firm CIBC are applying to their big data analytics projects. Next we provide a list of project management to-do items. We close with tips on evaluating big data technologies. ■

CRAIG STEDMAN
Executive Editor, SearchBusinessAnalytics

Large Companies Take Long View on Big Data Programs

Like many other organizations that have embarked on big data programs, healthcare services provider UPMC sees the flood of information it's generating as a blessing and a curse. "We're both drowning in big data and starving for it," said Lisa Khorey, vice president of enterprise systems and data management at the organization, based in Pittsburgh.

But UPMC is 20 months into a five-year plan to harness a wide variety of that data for analytics uses—and to bring in the big data management and analytics technologies needed to support the effort. In fact, company officials made a conscious decision not to invest in all of the required tools up front, said Khorey, who took part in a panel discussion on big data trends at the Oracle Industry Connect conference in Boston in March 2014.

"I don't think you have to buy it all on day one," Khorey said in an interview after the panel discussion. She added that UPMC's chief financial officer encouraged the phased approach by telling the big data project team "not to overload the buggy." The health system did select an initial set of hardware and software at the outset, including Hadoop and products from Oracle, IBM and Informatica. It plans to add predictive analytics tools in the summer of 2014 at the initiative's two-year mark; prescriptive analytics technologies will follow 18 to 24 months later.

Khorey noted, though, that project advocates got executive backing and organizational support for the full five-year program before beginning any deployments. That was crucial to making the three-step technology selection process work: She said the prudent approach wouldn't be feasible "if we had to re-justify everything each year, because that takes a lot of energy."

SWEATING THE TECHNOLOGY DETAILS

UPMC, a sprawling organization that operates 22 hospitals and about 400 outpatient facilities, also developed the plan for its big data systems with clinical precision, according to Khorey. "We spent a lot of time designing this architecture and then picking the [technology] elements that would fulfill each job," she said. For example, a Hadoop cluster is being used to capture and stage data on its way to a data warehouse; in addition, data discovery tools can be run against the Hadoop data to find relevant information for planned analyses.

The company didn't set up a formal committee to evaluate and select the big data technologies, but Khorey said a cross-functional group has been involved in developing the business requirements and technical specifications, as well as assessing the available options. IT is at the head of the table on that process, she said. But physicians and representatives from UPMC's life sciences operations also have a say on the technology plans and decisions.

The end goal is to enable collaborative analysis of genomics data and information on patient outcomes, physician performance, the cost and quality of care and other metrics—all in an effort to improve treatment and care delivery. "This is completely about outcomes, outcomes, outcomes," Khorey said. "We're seeking a scientific orientation so we practice [healthcare] based on measurements."

Thus far, UPMC has built the big data infrastructure, captured some initial data sets and run several proof-of-concept projects. Planned next steps include working to prove that the analytical processes can be repeated across different data sets and starting to deploy data and self-service analytics tools for use by business analysts, data scientists and other end users. Starter sets of clinical and cost data are due to be made available in June 2014, and Khorey said there will be "constant data landings" over the next few years as the program proceeds.

INITIAL BIG DATA DEPOSIT JUST THE START

Canadian Imperial Bank of Commerce (CIBC) is also in the early stages of a big data analytics program. The Toronto-based bank is testing marketing analytics, fraud detection and financial risk assessment applications; as part

of the pilot projects, it is working with various vendors and "playing around with different technologies," said Sam Dotro, CIBC's executive director of enterprise architecture. That includes Oracle's Big Data Appliance, Cloudera's Hadoop distribution and a mix of business intelligence tools, said Dotro, who works in the bank's New York offices.

Big data applications make more information available—often in real or near real time—in an effort to boost business processes.

Dotro also took part in the panel discussion at the Oracle conference. In a follow-up interview, he said the technology evaluation process is being driven by his group, but with "a lot of collaboration" and input from CIBC's business units. The company has set up a 20-member executive committee with representatives from IT, data security, corporate operations and the business units to plan out the big data

architecture. The process "is somewhat of a democracy," Dotro said. "But ultimately, it's the business that dictates this."

And demonstrating a business case for the proposed big data implementation is an important next step. In the coming months, CIBC executives will review the results of the pilot projects and decide what to move forward on. "But for sure, it's happening," said Dotro, who expects to get approvals for deployments and perhaps begin some of them during 2014.

For one thing, the bank's CEO has made the big data strategy one of his priorities, according to Dotro. In addition, competitive forces are pushing the bank to step up its analytics game. CIBC's data analysts typically look at only "small chunks of data," he said. The big data applications will make more information available—often in real or near real time—in an effort to boost functions, such as marketing and customer service. The result, Dotro said, will be a more data-driven—and hopefully more successful—company. *—Craig Stedman*

Project Managers Must Take the Big Data Helm

Big data analytics projects are at the top of the IT priority list for many organizations looking to wring business benefits out of all the data flowing into their systems. But with any initiative that offers big rewards, there are also accompanying big risks. That's certainly true of big data analytics programs, which makes planning and managing them effectively a must.

There are ways to go right—and ways to go wrong. What follows is a list of steps that big data analytics project managers should take to help set their programs on the right path, one that leads to the expected business value and a strong return on investment.

■ **Find business sponsors with solid business plans in mind.** With all of the hype surrounding big data analytics, business executives might well be lining up to sponsor a project. The key criteria for selecting sponsors should be whether they can articulate a clear set of

business objectives with a realistic timeline. By having a well-defined target of the business results you're looking to achieve, you can establish a scope for the data management and analytics systems that need to be built along with the supporting technology that needs to be installed. If a project starts without that kind of scoping, it's likely to spin out of control and try to do too much, too soon.

■ **Make learning—and mistakes—part of the project plan.** Big data analytics will introduce new technologies, techniques and methodologies in your organization, and likely require new skills. In addition, big data technologies are still evolving; a considerable amount of custom development work is often required, and there's a serious shortage of those required new skills, both for IT developers and the data scientists and other analytics professionals who will lead the data analysis work. As a result, your project

team will be learning as it goes, and business managers and users will be figuring out what big data analytics really means to them. You need to create project schedules and budgets based on a long learning curve, including the inevitable mistakes that will be made in the process of that learning.

■ **Get Agile on application development.** Because substantial training and education is likely to be required on everyone's part, and detailed business requirements might change as you proceed, Agile development methodologies are a better fit for big data analytics applications than standard waterfall approaches are. An Agile approach that delivers functionality in small, iterative chunks and accommodates quick changes in development plans works best amid all the uncertainty. It should be coupled with a visible and transparent change management process and regular communications with project sponsors and participants about progress and the changes that do occur.

■ **Time-box everything.** One of the tried-and-true project management rules, especially when it comes to software development, is that work will fill whatever available time bucket there is. As a big data analytics project manager, it's very likely that you'll be blessed with an extremely enthusiastic community of business executives and workers looking for information they can use to drive operational strategies and tactics. While learning on the fly and being open to changing requirements are part of the process, you need to leverage that enthusiasm by fitting scheduled work into tight time boxes in order to keep the project moving forward—and to keep people from getting discouraged by the initiative becoming stuck on particular tasks.

■ **Treat data scientists as artists.** Data scientists and other skilled analysts have a key role to play in pulling business insights out of big data stockpiles. Generating those insights, through applications such as predictive analytics and data mining, is an incremental and iterative process. A data scientist will devise an analytical model, test it, refine it, validate it and finally run it and publish the results internally. In doing so, he might test out dozens or hundreds

of variables using a variety of statistical methods. The term *data science* is somewhat misleading: Creating analytical insights is equal parts science and art. Treat data scientists as talented artists rather than common laborers and you'll encourage better productivity—and get better results.

■ **Set realistic expectations and manage them proactively.** In organizations that are new to big data initiatives, lofty expectations can be set by technology vendors that claim big data tools are easy to use and point to other enterprises that have gained significant business value from projects. It's important to keep in context that many of the early adopters of big data systems were large Internet companies that have

significant expertise and, in many cases, played leading roles in developing Hadoop and other big data technologies. If you let expectations get out of hand and then can't meet them, your project could be viewed as a failure regardless of the business value it does produce. Constrain expectations to realistic levels at the outset—and continue to do so throughout the project.

Clearly, there are both big risks and big rewards in undertaking a big data analytics project. But with proper attention to sound project management practices, project managers and their teams can minimize the downsides and make initiatives a big business opportunity. —*Rick Sherman*

In Evaluating Big Data Tools, Look at the Bigger Picture

MIKE LURYE HAS evaluated many of the software tools that have become nearly synonymous with big data. But the Time Warner Cable executive feels that some of the big data technologies getting the most attention today have relatively little business value, at least for his organization's current requirements.

Lurye, senior director of business intelligence architecture at Time Warner Cable Enterprises LLC, spoke about his company's big data challenges and business opportunities at the 2014 TDWI BI Executive Summit in Las Vegas. Two years ago, the country's second-largest cable TV provider wanted to create a database of the programming watched by every subscriber so advertisers could view reports and identify the shows watched by their target audiences more easily. Given the large volume of data involved in the project, big data tools like Hadoop or a NoSQL database might seem like the right choice. But Lurye said those options each had significant drawbacks from Time Warner Cable's perspective.

On the whole, Lurye is skeptical of the concept of big data. He sees the term as being too nebulous and would prefer to focus his attention on specific technologies that can help solve specific business problems. "I never met the person who invented the term *big data*, but that person is brilliant," he said. "It's a brilliant marketing term. It's not a technical term."

CHANGING THE CHANNEL ON HADOOP

Hadoop is the most-hyped big data technology now, and a new Hadoop 2 release that became available in October 2013 broadens its potential uses beyond MapReduce batch-processing applications. But Lurye said that when he evaluated Hadoop, he didn't see the business value for his company. At the time, he felt it was an immature technology that likely would

present new technical challenges. Integrating a Hadoop cluster with existing data sources and getting reports out of it seemed like difficult tasks given the skills that Time Warner Cable had in-house.

Lurye also looked at NoSQL databases for the project, but in the end he decided that they also weren't a good fit. He said most of the NoSQL technologies he reviewed required special programming skills that are hard to find. While NoSQL databases offer some interesting capabilities, the fact that Time Warner Cable would have had to hire new programmers specifically to operate one limited the technology's value.

"It would bring us back to the days where, to retrieve any data, someone would have to write code," Lurye said. "Why would we want to do that?"

Ultimately, Time Warner Cable went a more traditional route for the BI and analytics project. Viewing data gets loaded into a standard relational database. The company then uses an in-memory BI system from MicroStrategy to get reports out of the database and deliver them to advertisers.

DOOR STILL OPEN TO BIG DATA SOFTWARE
But even though the big data technologies weren't a good fit in this case, that doesn't mean Time Warner Cable—which in February 2014 agreed to be acquired by Comcast—has written them off for good. Lurye said he thinks Hadoop has matured in the time since he last evaluated it and now looks like it could be a cost-effective option for data integration.

Additionally, the cable provider is looking to start incorporating more real-time data, which might help predict who will watch a particular show, enabling vendors to perform ad bidding in real or near real time as advertisers try to target specific types of viewers. That application might be a good fit for NoSQL software, according to Lurye.

But even as the company considers possible deployments of new technologies, it isn't likely to get rid of the systems it has recently put in place any time soon.

"We see [big data technology] as a complement," Lurye said. "It's not going to replace a warehouse with 15 months of data. But it could make data available much sooner than what we have today." —*Ed Burns*

CRAIG STEDMAN *is executive editor of SearchBusiness Analytics and SearchDataManagement. He has been an IT journalist for more than 25 years and was an editor and reporter at* Computerworld *before joining TechTarget in 2009. Email him at* cstedman@techtarget.com *and follow him on Twitter:* @craigstedman*.*

RICK SHERMAN *is the founder of Athena IT Solutions, a consulting and training services company that focuses on business intelligence and data warehousing. Sherman is also an adjunct faculty member at Northeastern University's Graduate School of Engineering, and he blogs at The Data Doghouse. Email him at* rsherman@ athena-solutions.com*.*

ED BURNS *is site editor of SearchBusinessAnalytics; in that position, he covers business intelligence, analytics and data visualization topics. Before joining the site, he was a news reporter for SearchHealthIT. Email him at* eburns@techtarget.com *and follow him on Twitter:* @EdBurnsTT*.*

About TechTarget: TechTarget publishes media for information technology professionals. More than 100 focused websites enable quick access to a deep store of news, advice and analysis about the technologies, products and processes crucial to your job. Our live and virtual events give you direct access to independent expert commentary and advice. At IT Knowledge Exchange, our social community, you can get advice and share solutions with peers and experts.

COVER PHOTOGRAPH: THINKSTOCK