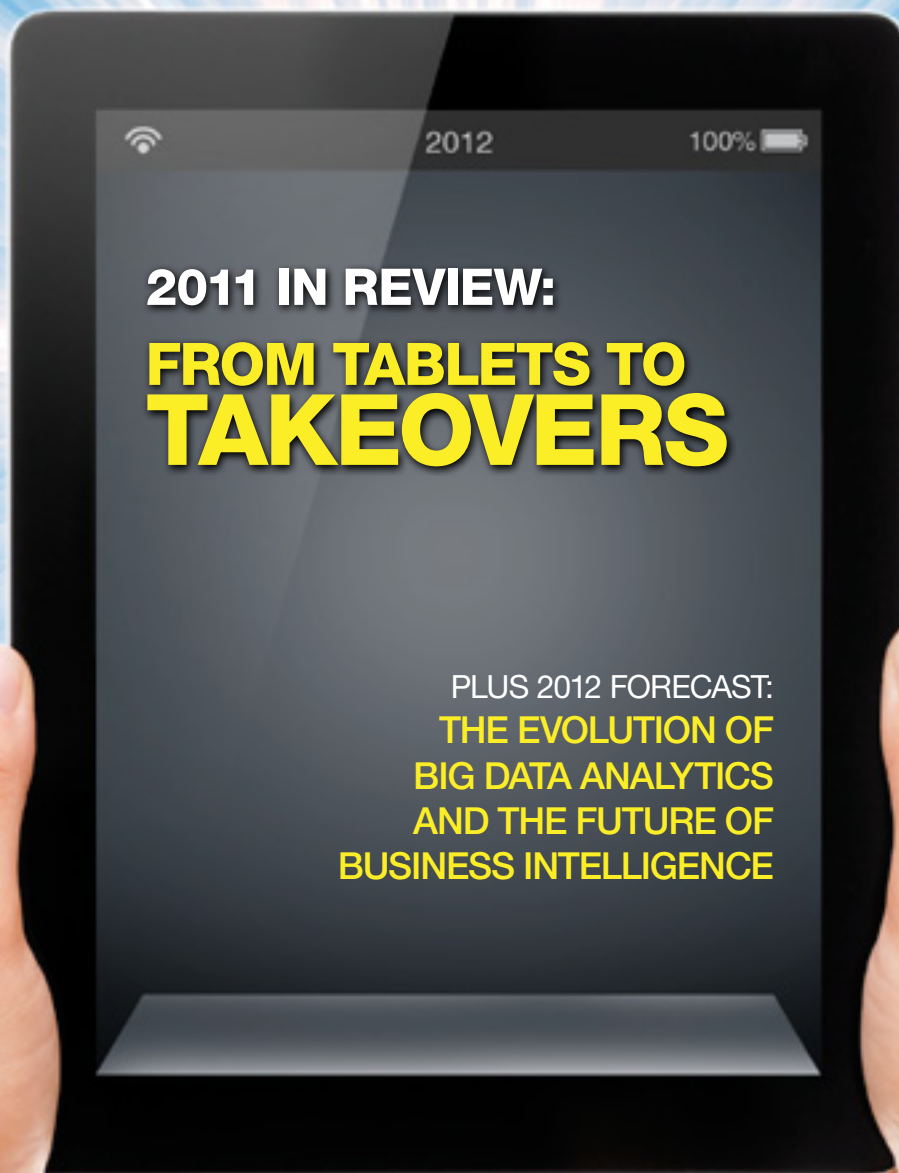


# tdwi's Best of Business Intelligence

VOLUME 9

*A Year in Review*

THE VERY BEST OF  
TDWI'S BI ARTICLES,  
RESEARCH, AND  
NEWSLETTERS



## RESEARCH EXCERPTS

The State of Operational Data Warehousing  
Introduction to Next Generation Data Integration  
Data Visualization Technology  
Introduction to Self-Service Business Intelligence

## INSIGHTFUL ARTICLES

Cloud Business Intelligence  
Tablets and Mobile Business Intelligence  
Structured and Unstructured Data  
Agile Requirements Gathering



# Explore. Collaborate. Discover.

From interactive dashboards and data visualizations to predictive analytics, Spotfire Analytics is an enterprise analytics platform that helps business and technical users quickly get to actionable insight.

## **Fastest to Actionable Insight**

Act on answers to relevant questions at the speed of thought.

## **Visibility into the Unknown**

Uncover trends, patterns, & unexpected insights hidden in data.

## **Self-service Discovery**

Increase agility and productivity by reducing reliance on IT.

## **Universal Adaptability**

Empower a broad spectrum of business & technical users with Spotfire's highly-scalable analytics platform.

» [Learn more at spotfire.tibco.com](http://spotfire.tibco.com)



*First to Insight. First to Action.*

## SPONSOR INDEX

Birst

DataFlux

Information Builders

Lavastorm Analytics

Morgan Kaufmann

SAND

Sybase, An SAP Company

TIBCO Spotfire

Vertica, An HP Company

## TABLE OF CONTENTS

### FEATURES

- 5 2011 in Review: From Tablets to Takeovers  
STEPHEN SWOYER
- 9 2012 Forecast: The Evolution of Big Data Analytics  
and the Future of Business Intelligence  
PHILIP RUSSOM, DAVID STODDER

### 2011 BEST PRACTICES REPORTS

- 19 The State of Operational Data Warehousing  
PHILIP RUSSOM
- 23 Introduction to Next Generation Data Integration  
PHILIP RUSSOM
- 26 Data Visualization Technology  
WAYNE ECKERSON, MARK HAMMOND
- 30 Introduction to Self-Service Business Intelligence  
CLAUDIA IMHOFF, COLIN WHITE

### 34 TEN MISTAKES TO AVOID SERIES

Ten Mistakes to Avoid When Setting Your  
Cloud Business Intelligence Strategy  
SHAWN ROGERS

### TDWI FLASHPOINT

- 39 An Introduction to Business Intelligence Life  
Cycle Management  
STEVE DINE
- 42 Business Requirements for Business Intelligence  
DAVE WELLS

### BUSINESS INTELLIGENCE JOURNAL

- 47 How Gamification Will Change Business Intelligence  
MARSHA BURKE, TROY HILTBRAND
- 55 BI Experts' Perspective: Integrating Structured  
and Unstructured Data  
BARRY DEVLIN, HARRIET FRYMAN, MICHAEL GONZALES,  
DAVID SCHRADER, DAVID STODDER

### BI THIS WEEK

- 65 How the iPad Experience Will Change the Face of BI  
STEPHEN SWOYER
- 67 7 Tips for Working with Data Analysts  
TED CUZZILLO

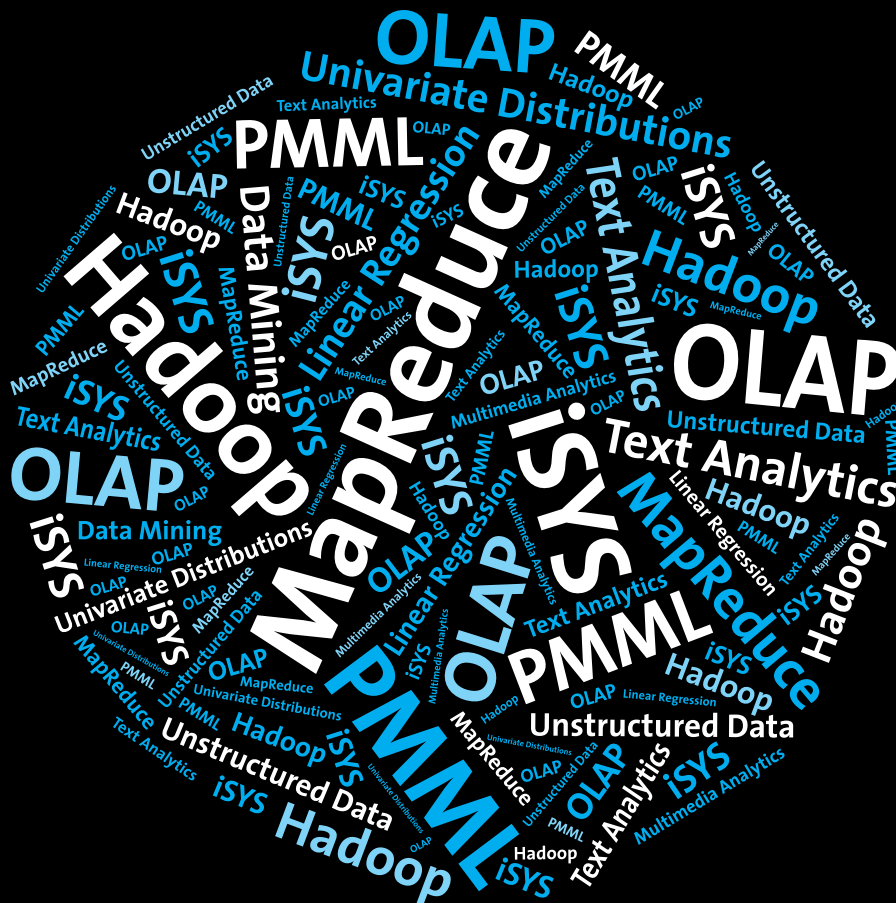
### TDWI EXPERTS

- 69 How Agile Requirements Gathering Can Conquer  
Analysis Paralysis  
MICHAEL A. SCHIFF

### 72 BEST PRACTICES AWARDS 2011

### 78 BI SOLUTIONS

### 83 ABOUT TDWI



## WELCOME TO THE WORLD OF **BIG DATA.**

**Sybase IQ delivers a Big Data solution that scales with your enterprise needs.**

With Sybase IQ, your data warehouse can answer thousands of simultaneous ad hoc and complex queries against terabytes or even petabytes of data. No other solution can take your Big Data efforts from one-time projects to enterprise-wide analytics of structured and unstructured data.

**Handle Big Data demands with Sybase IQ.**

[www.sybase.com/bigdata](http://www.sybase.com/bigdata)

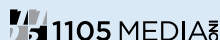


tdwi.org

Editorial Director **Denelle Hanlon**  
Senior Production Editor **Roxanne Cooke**  
Graphic Designer **Rod Gosser**



President **Richard Zbylut**  
Director of Education **Paul Kautza**  
Director, Online  
Products & Marketing **Melissa Parrish**



President &  
Chief Executive Officer **Neal Vitale**  
Senior Vice President &  
Chief Financial Officer **Richard Vitale**  
Executive Vice President **Michael J. Valenti**  
  
Vice President, Finance  
& Administration **Christopher M. Coates**  
Vice President, Information  
Technology & Application  
Development **Erik A. Lindgren**  
Vice President,  
Event Operations **David F. Meyers**  
  
Chairman of the Board **Jeffrey S. Klein**

#### REACHING THE STAFF

Staff may be reached via e-mail, telephone, fax, or mail.

**E-MAIL:** To e-mail any member of the staff, please use the following form: FirstInitialLastname@1105media.com

**RENTON OFFICE** (weekdays, 8:30 a.m.–5:00 p.m. PT)  
Telephone 425.277.9126; Fax 425.687.2842  
1201 Monster Road SW, Suite 250, Renton, WA 98057

**CORPORATE OFFICE** (weekdays, 8:30 a.m.–5:30 p.m. PT)  
Telephone 818.814.5200; Fax 818.734.1522  
9201 Oakdale Avenue, Suite 101, Chatsworth, CA 91311

#### ADVERTISING OPPORTUNITIES

Scott Geissler, [sgeissler@tdwi.org](mailto:sgeissler@tdwi.org), 248.658.6365

**REPRINTS AND E-PRINTS:** For single article reprints (in minimum quantities of 250–500), e-prints, plaques, and posters, contact PARS International.

Phone 212.221.9595; E-mail [1105reprints@parsintl.com](mailto:1105reprints@parsintl.com);  
Web [www.magreprints.com/QuickQuote.asp](http://www.magreprints.com/QuickQuote.asp)

© Copyright 2012 by TDWI (The Data Warehousing Institute™), a division of 1105 Media, Inc. All rights reserved. Reproductions in whole or in part are prohibited except by written permission. Mail requests to "Permissions Editor," c/o Best of BI 2012, 1201 Monster Road SW, Ste. 250, Renton, WA 98057. The information in this magazine has not undergone any formal testing by 1105 Media and is distributed without any warranty expressed or implied. Implementation or use of any information contained herein is the reader's sole responsibility. While the information has been reviewed for accuracy, there is no guarantee that the same or similar results may be achieved in all environments. Technical inaccuracies may result from printing errors, new developments in the industry, and/or changes or enhancements to either hardware or software components. Produced in the USA.

TDWI is a trademark of 1105 Media, Inc. Other product and company names mentioned herein may be trademarks and/or registered trademarks of their respective companies.

## EDITORIAL DIRECTOR'S NOTE

Welcome to the ninth annual *TDWI's Best of Business Intelligence: A Year in Review*. Each year we select a few of TDWI's best, most well-received, hard-hitting articles, research, and information, and present them to you in this publication.

Stephen Swoyer kicks off this issue with a review of major business intelligence (BI) developments. In "2011 in Review: From Tablets to Takeovers," he names social BI and the success of tablets and mobile devices as some of 2011's trends. Swoyer also calls 2011 "the year in which social media emerged as one of several forces ... that will fundamentally transform BI as we know it."

In "2012 Forecast: The Evolution of Big Data Analytics and the Future of BI," TDWI Research analysts Philip Russom and David Stodder share their predictions for the coming year. Russom offers insight on how the evolution of big data analytics will impact business intelligence and data warehousing professionals, and Stodder shares five trends he sees shaping the future of BI.

To further represent TDWI Research, we've provided excerpts from the past year's Best Practices Reports. Russom's "The State of Operational Data Warehousing" covers the benefits and barriers of operational data warehousing, and "Introduction to Next Generation Data Integration," also by Russom, covers data integration's impressive evolution in recent years. "Data Visualization Technology," by Wayne Eckerson and Mark Hammond, defines the two main types of data visualization. Finally, Claudia Imhoff and Colin White describe the objectives of self-service BI in "Introduction to Self-Service Business Intelligence."

This volume's *Ten Mistakes to Avoid* will help you avoid some common pitfalls when creating your cloud BI strategy. And thanks to articles from TDWI's e-newsletters, you'll learn more about BI life cycle management, agile requirements gathering, mobile BI, and working with data analysts.

In "How Gamification Will Change Business Intelligence," one of our selections from the *Business Intelligence Journal*, Marsha Burke and Troy Hiltbrand discuss ways to enhance the user experience through the effective application of game mechanics to BI. Our second *Journal* piece, "BI Experts' Perspective: Integrating Structured and Unstructured Data," with commentary from experts in the field, offers suggestions for utilizing unstructured data.

TDWI is committed to providing industry professionals with information that is educational, enlightening, and immediately applicable. Enjoy, and we look forward to your feedback on the *Best of Business Intelligence*, Volume 9.

Denelle Hanlon  
Editorial Director, *TDWI's Best of Business Intelligence*  
The Data Warehousing Institute  
[dhanlon@tdwi.org](mailto:dhanlon@tdwi.org)

# Remember when:

- Bellbottoms were the height of fashion
- Boogie fever ruled the dance floors
- It took a forklift to upgrade your data warehouse

Sand ad

**SAND can't bring back disco,  
but we can easily upgrade your  
outdated data warehouse.**

- Scalable
- Predictable
- Agile
- Faster than sound

**[www.sand.com](http://www.sand.com)**





2011 in Review

# From Tablets to Takeovers

BY STEPHEN SWOYER

Every winter, we pronounce a verdict on The BI Year That Was. This time around, we have an especially full slate, and not much room in which to recap it. So, in the words of the great jazz pianist Horace Silver, let's get down to the nitty gritty, because this has been one heck of a year in BI.

## Interacting with BI Apps and the Impact of Mobile Computing

At some point in the next 8 to 18 months, Microsoft will ship Windows 8, the successor to its Windows 7 operating system. From a BI perspective, Windows 8 isn't any great shakes, but from a *user experience* perspective, it's an excellent indication of where we're headed.

Windows 8 will ship with a new user interface (UI) based on Microsoft's Metro design language, which currently powers Windows Mobile, Microsoft Zune, and other technologies. The point isn't that Metro is new and/or different—because it isn't, necessarily—it's that the Metro UI that Microsoft has showcased with



its Windows 8 beta incorporates non-PC-like concepts and methods. It borrows, to a degree, from the mobile or tablet user experience, in which touch—and not the gesturing of the mouse—is the primary means of interacting with an application.

## FOR NOW, LET'S JUST SAY THAT THE WAY IN WHICH WE INTERACT WITH BI IS GOING TO CHANGE. DRASTICALLY.

In 2011, the evolving BI usage paradigm officially moved beyond the desktop LCD: smartphones and tablets aren't simply being *supported*, they're being *actively cultivated*. One upshot of this is that nearly every purveyor of BI client offerings has a mobile or tablet strategy that tends to mix support for mobile devices of all kinds—chiefly via HTML5-based Web applications—and device-specific support, usually (but not always) limited to Apple Inc.'s iOS devices. We'll say more about tablets and mobile devices a bit later in this article, as their success—which most people saw coming—was a big part of the year in BI.

For now, let's just say that the way in which we *interact* with BI is going to change. Drastically. After all, it already *has* changed, and what we expect from BI applications has changed, too. Even before 2011, innovation in BI had already outstripped the bread-and-butter contexts in which business intelligence had been served up for decades. These old-school interfaces—viz., spreadsheets, with their cells, rows, and columns; static charts, diagrams, and scorecards; and so on—haven't so much been replaced as been *supplemented*: enriched.

They're still in use—for the majority of users, they're still a primary interface—but today's BI tools mix a highly interactive, increasingly visual application environment with an increasingly collaborative user experience. They likewise bundle amenities—e.g., self-service, sophisticated visualization, analytic technologies—that once seemed like the stuff of pipe dreams. Furthermore, they're increasingly social, and that's key.

### Advanced Social Studies

The socialization of BI has been inexorable. Not fast. Not slow. Inexorable.

Over the last 16 months, for example, big-name vendors (e.g., IBM, MicroStrategy, Oracle, SAP AG) moved to incorporate social media–like features into their BI suite offerings. They were mostly following the lead of smaller

BI vendors, such as Lyza, a start-up that launched a fully-formed, social-ready BI platform almost two years ago.

The trend this year might be called "creeping socialization." It was a phenomenon in which at least a dozen BI vendors kicked off efforts to retrofit their offerings with social media (or social media–like) capabilities. This didn't just include prominent names (such as MicroStrategy and SAP), but scrappy players like QlikTech, which radically fleshed out its social media feature set with its QlikView 11 release.

Elsewhere, vendors such as data integration (DI) specialist iWay Software—along with its parent company, Information Builders—and complex event processing (CEP) giant TIBCO touted their own achievements in social BI. Even social BI trailblazer Lyza announced a new—and (not surprisingly) even more sociable—version of its flagship BI offering.

However, 2011 wasn't the year in which BI first became social. Nor was it the year in which BI completed its social makeover. It was, instead, the first full year in which the industry gravitated, inexorably and irrevocably, toward social BI.

It was likewise the year in which social media emerged as one of several forces—all of which have to do with how we use, interact with, and consume BI—that will fundamentally transform BI as we know it.

### No Room for Squares

BI usage models aren't the only things changing. This year brought a reconfiguration of the industry pecking order, with the emergence of several small or upstart competitors and the reinvention—or reorientation—of a number of established players, too.

Quiterian, Metric Insights, and JackBe were just a few of the BI vendors to make splashes this year. Both Quiterian and JackBe have been around for some time: the former, was—until this year—an EU-only player; the latter had its start (back in 2006) focusing on Ajax and SOA application development. Metric Insights, on the other hand, is a new entry, the brainchild of BI consultant Marius Moscovici. The emergence of these and other vendors illustrates one of the chief ways in which the BI industry contrives to remain competitive, innovative, and, yes, *interesting*—even in the midst of significant consolidation.

They focus on the gaps; on the missing pieces; on the areas of user or IT frustration; on all the ways existing BI just isn't getting it done.

Barcelona-based Quiterian, for example, positions its DDWeb offering as a complement to the established data warehousing and analytic practices that exist in most organizations today. “If you want to test ideas with incomplete data or with less-than-perfect data to see if it makes sense to do something down the line using traditional BI, this is what to do,” said Alberto Saavedra, vice president of U.S. sales at Quiterian, when we interviewed him in August.

JackBe, for its part, released version 3.1 of its Presto product back in March. Presto serves up analytics as reusable mashups of different service-enabled data sources, which (in the JackBe lexicon) are dubbed “mashables.” Add in a wizard-driven self-service component and you have an approach that “make[s] any kind of data source—regardless of its location—equally accessible. We fuse that together with our other theme: as much self service as possible. We try to make it easy for nontechnical users to mash together as many of the mashables as they want,” said JackBe CTO John Crupi, in a September interview.

Metric Insights says its flagship product, Instant Insight, is an analytics-for-the-rest-of-us kind of tool. “There [are many] companies that do a very good job at the deep analytics, [but these are] tools analysts would normally use,” said founder Marius Moscovici, who positions Instant Insight as a lightweight analytic offering that eschews the monolithic architectures of heavy-duty BI tools. “Rather than taking ... the approach ... where you have to define a metadata layer first and then build the reports or the analysis, here you build the reports first. The metadata is the reports and the metrics.”

These and other vendors say they’re exploiting gaps in the existing BI marketplace: i.e., segments, pain points, or use cases that they say are ill-served by existing offerings—or by existing usage paradigms. It’s the same kind of approach that a previous generation of upstart or insurgent competitors—e.g., QlikTech, Lyza, WhereScape—exploited, too. It’ll likely function as an on-ramp for BI innovation, and BI differentiation, in 2012 and beyond.

## Changes in Attitude

Even as Quiterian, JackBe, and Metric Insights trumpeted a new spin on traditional BI, established players repositioned themselves to exploit opportunities in the ever-evolving data integration (DI) marketplace.

Extract, transform, and load (ETL) specialist Syncsort is a case in point. For decades, ETL has been Syncsort’s bread and butter. It still is, but with this year’s release

of DMEExpress 7.0, Syncsort is itself evolving—namely, by combining its traditional focus on hot-rodded ETL with an ambitious vision, which it dubs ETL 2.0—that emphasizes self-service and collaboration and which itself augurs the development of a more platform-like DI vision. Next year will be an interesting—indeed, perhaps pivotal—year in Syncsort’s and ETL 2.0’s development.

Vendors such as Composite Software and Informatica have been talking up data virtualization (DV) for a few years. In 2011, they seemed to be pitching to an *especially* receptive audience. Composite, for example, convened its second-annual “Data Virtualization Day” event in midtown Manhattan, drawing several hundred attendees.

Informatica, for its part, trumpeted its DV bona fides, even as other DI players—from giants such as IBM, iWay Software, and Oracle to specialists including Denodo Technologies and SnapLogic—articulated DV visions, or used DV-like language, of their own. This list of would-be DV-ers includes DataFlux, the face of data quality for SAS from time immemorial (or at least 2000). Two years ago, DataFlux became the face of SAS DI. As of this year, DataFlux added DV to its list of talking points.

Everybody, it seemed, had DV on the brain. Well, *almost* everybody. Insurgent DI specialist WhereScape isn’t backing away from its bread-and-butter focus on ETL. This year, however, it unveiled WhereScape 3D, a kind of free test laboratory for DI. (Yes, WhereScape 3D is free. It doesn’t even require special glasses!) Although 3D *can* be used to generate code for Red (WhereScape’s flagship ETL tool), it can also double as a feasibility testing tool: i.e., is such a project possible, and—more important—at what time and what cost?

## Maximum Consumption

Few industries have produced as much acquisition activity as the BI space over the last decade. This year, consolidation in the analytic database arena—long-anticipated, and, starting in 2010, partially realized—continued apace, with HP gobbling up columnar database powerhouse Vertica. Not to be outdone, Teradata acquired analytic database highflier Aster Data. Along with the former Greenplum Software, which was acquired last year by EMC, Aster was one of the first two analytic database entrants to bring a native implementation of MapReduce to market. There are still plenty of extant analytic database competitors, however. Veterans Infobright, Kognitio, and ParAccel, along with newer entrants including Algebraix Data and Actian (the former Ingres), seem determined to keep things interesting.

Also this year, Oracle ponied up \$1 billion to acquire Endeca, an established vendor that specialized in faceted search. Although search was indisputably Endeca's bread and butter—its technology is used to power at least half of the top 100 e-commerce sites—the company believed it could make a big splash in the BI market, too.

## 2011 WAS THE YEAR IN WHICH THE TABLET VOUCHSAFED ITS ROLE IN THE ENTERPRISE.

"Within the last year and a half, we've had a big BI push, but even before that we started working with some early adopter organizations ... [such as] Raytheon about solving business intelligence problems, taking our search technology, adding analytics to it, adding visualizations to it, and bringing it within the enterprise," said John Joseph, Endeca's director of product marketing, in an August interview, two months before Oracle acquired Endeca. (This wasn't Oracle's only BI-related acquisition in 2011, by the way: it nabbed data profiling specialist Datanomic in April.)

Eight years ago, for the record, the former Business Objects SA ponied up \$820 million for the former Crystal in an acquisition that—at the time—was pronounced one of the biggest in BI history. How times change!

### A Torrent of Tablets

You knew this was coming, because although 2011 might not have been the Year of the Tablet, it *was* the year in which the tablet—as embodied in the category-defining form-factor of the iPad—vouchsafed, seemingly in perpetuity, its role in the enterprise.

Enterprise IT organizations are proceeding apace with tablet computing adoption efforts: many are already using iPads (or, less frequently, competitive tablet offerings) in production, while most are at least *mulling* tablet strategies. In Apple's Q4 2011 earnings conference call, for example, CFO Peter Oppenheimer claimed that the iPad enjoys almost complete penetration across the *Fortune* 500.

"Today, 92 percent of the *Fortune* 500 are testing or deploying iPads within their enterprises, up from 86 percent last year," Oppenheimer told investors. Yes, there's a sense in which this claim stretches credulity. After all, if 86 percent of the *Fortune* 500 was in fact "testing or deploying" the iPad by the end of its first year on the market, you can bet that many were more testing than

deploying, and you can likewise question what it means to "test" in this context.

That said, there's an expectation among market watchers that shops will increasingly tap tablets to complement—and in many cases to displace—desktop or laptop computers. For example, tablet shipments will increase by as much as 300 percent—from 18 million in 2010 to 62.5 million in 2011, according to projections from IDC. In 2010, tablet unit sales came in at about 5 percent of PC unit sales (per IDC's figures); for 2011, IDC puts that number at 17.5 percent. That's a simply staggering growth rate.

Even though Apple controlled almost 70 percent of the tablet market at the end of 2011, few, if any, BI industry execs are willing to call it a one-tablet race. What's more, BI vendors are putting their development dollars where their non-provocative/partner-pleasing public statements are: they're developing versions of their software offerings for platforms other than the iPhone and the iPad.

"Last year was the year of the iPad as an interactive client device. We would hear from our customers, 'We want a native app just for the iPad. We're only going to be iPad.' But that tune has changed a bit over the last four or five months. What we're also seeing is that people are bringing their own devices into work, and while many of these are iPads, some of them aren't," said Jake Freivald, product marketing manager with Information Builders, back in August.

That's why IBI and other vendors are developing HTML5-based supplements to any of the native or platform-specific applications they offer. (If a BI vendor offers mobile or client support, there's a good chance they'll likewise offer a native app for the iPhone, if not the iPad.) In fact, a study from software development researcher Evans Data found that a majority of mobile developers are focusing on Web apps instead of native apps—in spite of the dominance of the iPad or (in the small-form-factor mobile arena) the iOS or Android platforms.

Check back next year on your tablet of choice—by the way, market watcher IDC predicts that Apple's share of the tablet market will drop to about 50 percent by the end of 2011—as we recap the The Year that Will Be: 2012.

---

*Stephen Swoyer is a technology writer based in Nashville, TN. Contact him at [stephen.swoyer@spinkle.net](mailto:stephen.swoyer@spinkle.net).*





## 2012 FORECAST:

# THE EVOLUTION OF BIG DATA ANALYTICS AND THE FUTURE OF BUSINESS INTELLIGENCE

BY PHILIP RUSSOM AND DAVID STODDER

## TOP 10 TRENDS IN BIG DATA ANALYTICS

BY PHILIP RUSSOM, RESEARCH DIRECTOR FOR  
DATA MANAGEMENT, TDWI

This time last year, I wrote about how advanced analytics has become such an important priority for user organizations that it's influencing most of what we do in business intelligence (BI), data warehousing (DW), and data management (DM). Analytics continues to be a high priority for 2012, as well as for the next few years. However, advanced analytics is more and more applied to very large data sets, which have come to be known as *big data*. A prominent change we witnessed in 2011 is that advanced analytics and big data are now used in tandem so commonly that they have coalesced into a new practice we call *big data analytics*.

From this recent evolution, you can see that big data analytics is really two things. First, it involves advanced analytic tools and techniques, which are usually enabled by data mining, statistical analysis, complex SQL, artificial intelligence, advanced data visualization, or natural language processing. These tools and techniques are also enabled by database technologies such as MapReduce, in-database analytics, in-memory databases, columnar data stores, and so on. Note that online analytic processing (OLAP) is not considered an advanced form of analytics, although it is the most commonly used form of analytics today. Second, the tools and techniques of advanced analytics operate on big data, defined as multi-terabyte or petabyte-scale data sets that

typically collect data and documents of multiple structures and types from many diverse sources.

Big data analytics is a relatively new practice, so its evolution is still ongoing and will be for years. To help you understand the evolution—and how its changes may impact you—allow me to discuss the top 10 trends in big data analytics.

1. **From hoarding big data to leveraging big data.** From the time user organizations first started bringing their businesses onto the Internet in the 1990s, they've been hoarding clickstream data and other Web-based big data. More recently, organizations have hoarded big data from RFID, test equipment, robotics, e-mail, social media, and sensors, as well as traditional enterprise systems. For 10 years or more, organizations collected and saved massive data sets under the assumption that they would eventually figure out multiple ways of wringing business value from big data.

Fast-forward to 2006 or 2007. User organizations of all types came to the realization that advanced analytics is a way to get business value from big data. But the realization wasn't enough; it was Moore's Law that tipped many organizations into actually practicing advanced analytics with big data. Several innovations from vendor companies made analytic operations on tens of terabytes more technically feasible—and less expensive. Innovations include multi-core CPUs; the massive memory space of 64-bit computing; the increased speed and capacity (yet lower cost) of storage, server virtualization, and clouds; and the arrival of many new

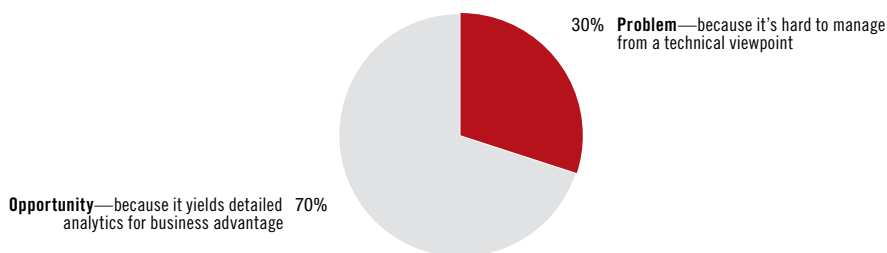
database management systems (DBMSs) and appliances built specifically for multi-terabyte data warehousing and analytics.

## BIG DATA HAS "ARRIVED" IN THAT 70 PERCENT THINK IT'S A BUSINESS OPPORTUNITY WHEN ANALYZED PROPERLY.

Hence, over a five-year period, big data has gone from being an IT cost center to a valuable source of business insight. It can also be a source for developing data-driven products via analytics. In the 2011 TDWI Research survey on big data analytics, only 30 percent of respondents considered big data a data management problem. Big data has “arrived” in that 70 percent think it's a business opportunity when analyzed properly. (See Figure 1.)<sup>1</sup>

2. **To boldly go where no analytic application has gone before.** The kind of analytics applied to big data is usually called “advanced analytics.” Better terms would be “discovery analytics” or “exploratory analytics,” because that's what users are trying to accomplish. In other words, with big data analytics, the user is typically a business analyst, data scientist, or similar user who is exploring large data sets to discover new business facts that no one in the enterprise knew before. For example, the point may be to discover the root cause of the newest form of customer churn, hidden costs

In your organization, is big data considered mostly a problem or mostly an opportunity?



**Figure 1.** Source: TDWI, June 2011. Based on 325 responses.

<sup>1</sup> For a more detailed discussion and more survey results, see the TDWI Best Practices Report *Big Data Analytics*, available on [tdwi.org](http://tdwi.org).

that reduce profitability, subtle shifts in customer-base segmentation, evolving affinities among products, relationships that might indicate fraud, and so on.

## BIG DATA ISN'T JUST BIG— IT'S ALSO DIVERSE.

Note that the practice of big data analytics depends on large volumes of data that include a lot of details. For this reason, big data for analytics may consist of detailed data straight from enterprise, Web, and other sources, with little or no cleansing or transformation. This is often data that the enterprise has not tapped for analytics, due to hoarding or the newness of the business issue being addressed. Hence, big data analytics should take analysts and analytic applications where none have gone before, because users' needs are trending toward discovery analytics.

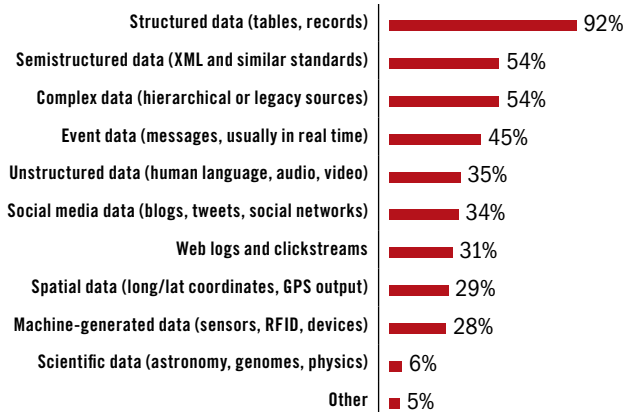
3. **From terabytes to petabytes of data for analytics.** Organizations with big data typically tap a subset of it for analytics—and the subset for analytics is growing, just as you'd expect from any collection of big data. According to the 2011 TDWI Research survey on big data analytics, one-third of organizations are already managing 10 terabytes or more of big data just for analytics. In fact, the 10- to 100-terabyte range seems to be today's norm for big data for analytics. However, one-fifth of survey respondents expect to have more than 500 terabytes of big data *just for analytics* by 2013. At this rate, we'll soon measure big data for analytics in petabytes instead of terabytes. You should plan capacity accordingly.
4. **Bigger and better analytic applications, both old and new.** Big data analytics enables new analytic applications, such as social media analytics and customer sentiment analysis. But big data can also take older analytic applications to a new level. For example, analytics for risk and fraud that are based on statistical analysis or data

mining benefit from the much larger data samples of big data. Furthermore, most 360-degree customer views include hundreds of customer attributes. Big data can bump that up to thousands of attributes, which in turn provides greater detail and precision for customer-base segmentation and other traditional customer analytics. As big data analytics becomes more mainstream, it will be as much about improving older analytic applications as it is about enabling new ones.

5. **An opportunity to finally address multi-structured data for BI.** Big data isn't just big—it's also diverse. For example, 92 percent of users responding to the 2011 TDWI Research survey on big data analytics reported managing structured data in their big data for analytics. As you'd expect, structured data still retains its hegemony over other data types. Yet, survey respondents are also managing data that's semi-structured (54 percent), hierarchical (54 percent), event-based (45 percent), unstructured text (35 percent), social media (34 percent), Web-based (31 percent), spatial (29 percent), machine generated (28 percent), and so on. (See Figure 2.) TDWI expects big data to become even more diverse. Hence, organizations hoping to get full value from big data via analytics will need to beef up their skills and platform support to embrace more multi-structured data types.
6. **Hadoop moves beyond Internet companies.** Apache Hadoop and related technologies (such as MapReduce, Hive, and Hbase) have been with us for a few years now. Hadoop's value for big data analytics is clear, at least with Web data at large, Internet-based companies. We're just now seeing mainstream user organizations that are not exclusively Internet companies use Hadoop and MapReduce for analytics with their Web data, plus traditional enterprise data and sensor data. Hadoop is known for massive scalability, but it can also handle a wide range of data, file, and document types, at a time when most BI and DW implementations have barely begun to leverage these multi-structured data types. At the current rate of adoption, we can

<sup>2</sup> For a complete appraisal of Hadoop's value for BI, see *TDWI Checklist Report: Hadoop—Revealing Its True Value for Business Intelligence*, available on [tdwi.org](http://tdwi.org).

Which of the following data types are you collecting as big data and/or using with advanced analytics today? Select all that apply.



**Figure 2.** Source: TDWI, June 2011. Based on 450 responses from 109 respondents who report practicing big data analytics.

expect Hadoop and related technologies to soon be a common complement to BI and DW implementations, to help them embrace both big data and multi-structured data.<sup>2</sup>

7. **Fast and frequent data, not just big and diverse data.** The strongest trend in BI today is the movement toward real time. For example, think of the practice of operational BI, a popular management technique that requires regularly refreshed reports, often in dashboard style. Operational BI increases the frequency of data extracts, and also provides fast queries for speedy report refresh. Just as operational BI accelerated reporting into real-time operation, some analytic applications are doing the same for big data analytics. The bleeding edge of big data analytics today is where an analytic application receives, analyzes, and responds to streaming data coming from sensors, robotics, monitoring devices, RFID, Web sites, and many types of operational applications. Although real-time big data analytics is somewhat rare today, it will soon be commonplace due to the general trend toward real-time BI and the fact that a host of sensors are coming online.
8. **Stretching data warehouse architecture to fit big data analytics.** As organizations adopt more analytic methods and deploy more analytic applications, they introduce more data-oriented workloads into their business

intelligence (BI) and data warehouse (DW) environments. When the architecture of a DW is designed and optimized for the most common deliverables—namely standard reports, dashboards, and OLAP—it makes sense to put other data-processing workloads (including those for advanced analytics) on separate platforms on the edge of the data warehouse (instead of inside the data warehouse proper). Other workloads related to big data are being relegated to the edge, too, including those for real time, massively detailed source data, and multi-structured data. In turn, the increasing number of edge workloads is driving data warehouses toward distributed architectures, except where a hefty DW platform can support multiple, diverse, and concurrent workloads. Hence, as users embrace big data analytics, they typically must also adjust DW architectures.<sup>3</sup>

9. **Take out the old and bring in the new analytics platforms.** An analytics platform can take many forms. To some users, it's the analytic tool used to create analytic models or fashion complex queries. To others, it's the DBMS where analytic big data is managed and operated on. It could be both. And everyone longs for heftier hardware for their analytics platform. Regardless of the definition, one-third of respondents to the 2011 TDWI Research survey on big data analytics say they're considering

3 See the article "Analytic Workloads Affect Data Warehouse Architecture" in TDWI's e-newsletter *TDWI Experts*, available on [tdwi.org](http://tdwi.org).

a replacement of their analytic platform by 2014. But why make such an expensive and intrusive move? Survey respondents say that their current platform(s) suffer from poor scalability, data loading, query speed, real-time operation, and so on—or it simply doesn't support the style of modeling or visualization they need. As users reevaluate their business requirements for analytics, they sometimes conclude that satisfying the requirements demands new platforms and tools for big data analytics.

10. **Aggressive growth for some big data analytics techniques and tool types.** Based on results of the 2011 TDWI Research survey on big data analytics, several techniques (whether hand coded or based on a vendor tool) will see strong growth in the next three years, including advanced analytics, predictive analytics, and text analytics. However, the strongest growth predicted by the survey is for advanced data visualization, which reminds us that most users prefer a strongly visual user interface for exploring big data and presenting analyses of it. Survey results predict good growth for the many new DBMSs and appliances built specifically for multi-terabyte data warehousing and analytics. Good growth is also projected for new techniques and platforms, such as Hadoop, MapReduce, event processing, no-SQL databases, and clouds.

---

*Philip Russom is the research director for data management at The Data Warehousing Institute (TDWI), where he oversees TDWI's research-oriented publications, services, and events. He is a well-known figure in data warehousing and business intelligence, having published over 500 research reports, magazine articles, opinion columns, speeches, Webinars, and more. Before joining TDWI in 2005, Russom was an industry analyst covering BI at Forrester Research and Giga Information Group. He also ran his own business as an independent industry analyst and BI consultant and was a contributing editor with leading IT magazines. Before that, Russom worked in technical and marketing positions for various database vendors. You can reach him at [prussom@tdwi.org](mailto:prussom@tdwi.org), @prussom on Twitter, and on LinkedIn at [linkedin.com/in/philiprussom](http://linkedin.com/in/philiprussom).*

## FIVE TRENDS SHAPING THE FUTURE OF BUSINESS INTELLIGENCE

BY DAVID STODDER, RESEARCH DIRECTOR FOR BUSINESS INTELLIGENCE, TDWI

Whether the economy rises or falls in 2012 is anyone's guess, but we can be pretty certain either way that a key competitive differentiator will be how well organizations use information. Business intelligence (BI) and analytics are central to accessing and analyzing data and sharing insights with colleagues, business partners, and customers. Using information to decipher the first indications of customer demand and capitalize on it with new products and services is critical to profitability.

Strong capabilities for sharing information insights to ensure coordination between different departments and globally dispersed divisions can be the difference between saving and losing money. BI applications, tools, and services must work in tight harmony with underlying data integration and warehousing systems. But that's not all. As information needs branch out, the technology environment must include data visualization, search, text analytics, social media, event processing, and database and content management systems for unstructured and semi-structured information. We are still in the early, innovative stages of bringing all these components into an integrated whole.

There are five big trends that I see shaping the future of BI. Certainly, 2012 may bring more of the kind of industry consolidation that we've seen in recent years, including 2011, which will impact how trends play out. Consolidation can be good, in that it makes room for innovative start-ups that propel technology implementation forward faster than is possible with the established players.

The five trends I chose are independent of industry moving and shaking, however. They focus instead on key aspirations and objectives that I've picked up in my research and in conversations with organizations about their BI and analytics projects.

1. **Data discovery and visualization will reshape BI experiences.** Reporting, while essential, is passé; "discovery" is what an increasing



number of users want to do with their data. Rather than be limited to the known world of reports, users want to go beyond them to analyze detailed big data and unearth significant and unexpected insights. Their quest is the essence of analytics: to understand why something is happening and what can be done to improve outcomes.

Advanced BI and data discovery tools are giving business users greater ability to define variables and perform discovery analytics on their own. The analytics are being used to sharpen metrics to focus on aspects of performance that have the most impact on desired business outcomes. Discovery analytics are helping organizations see data relationships, such as between costs and outcomes, to understand whether the cost of taking certain actions to improve customer satisfaction is actually worth it from a financial perspective.

Critical to discovery is data visualization. Dashboards have improved upon traditional BI reports to provide a range of data visualizations, but dashboards are being tested by the growth of analytics. Moving beyond standard online analytical processing (OLAP) visualization to represent roll-ups and aggregations, discovery analytics demand scatter plots, heat maps, and more to be available on the fly. In fact, the hardest part of data

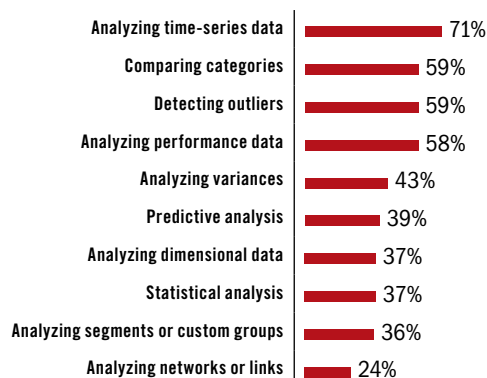
visualization may be to resist cramming too much “eye candy” into dashboards. Visualizations should be easy to understand and should support decision processes. Figure 1 offers a perspective on the importance of visualization for major analytical tasks.<sup>1</sup>

## SOCIAL BI IS THE NEW FRONTIER. BI SERVICES LOCATED IN THE CLOUD TYPICALLY PLAY A BIG ROLE.

Data visualization and analytics come together to give us “visual analytics,” in which nontechnical users explore data through graphical representations without having to get their hands dirty with the data itself (unless they want to). Visual analytics are essential as organizations begin to tap big data to unearth trends, anomalies, and behavior that might indicate new customer preferences, fraud, or other significant events. BI and discovery tools that enable visual analytics are helping organizations sow the seeds of an analytics culture that applies data insights effectively to shape decisions and actions.

2. **Social and collaborative BI innovations will break the BI mold.** With so many ways to communicate with friends and colleagues these days, it can take a minute to decide whether to text, call, chat, e-mail, Skype, tweet, write a wall post, or even just walk

Rate the degree of usefulness of visualization for the following analytical tasks.



**Figure 1.** Source: TDWI, January 2011. Based on 210 respondents who ranked visualization's importance as "high."

<sup>1</sup> For a more detailed discussion, see the TDWI Best Practices Report *Visual Reporting and Analysis: Seeing Is Knowing*, available on [tdwi.org](http://tdwi.org).



over and talk in person. BI has traditionally been a “personal” application that is difficult to share. To support collaborative decisions and actions, BI has to change.

## MAKING INFORMATION ACTIONABLE SHOULD NOT BE A LICENSE TO INCREASE THE NOISE LEVEL SO THAT EMPLOYEES ARE LOST IN A CACOPHONY OF BLINKING ALERTS AND URGENT MISSIVES.

In recent years, BI tools have incorporated primitive collaboration features such as commenting, annotation, and the ability to e-mail links to specific reports. Now, however, the leading edge of BI is moving to the next stage to become more fully integrated with collaboration systems and social media. The first step has been to improve integration with corporate applications such as Microsoft SharePoint and IBM Lotus as well as dominant cloud services like Google. Tools are also integrating communication, collaboration, and BI at the glass, using mashup services accessed from clouds to turn dashboards into “mashboards” that allow users to pick their own components and widgets.

Social BI is the new frontier. BI services located in the cloud typically play a big role. Social BI capabilities allow users to disseminate reports, dashboards, and analyses through social media platforms and wikis that can be either public or private. Social BI enables users to receive faster and broader response to the BI artifacts they are sharing. Finally, social BI can open up projects to collaborative, crowdsourcing approaches for data analysis and reporting. Social BI is an important area of innovation in how single users experience BI and, more important, how collaborative and often globally distributed groups can share BI artifacts and capabilities.

3. **Performance management will become more actionable and integrated.** Performance

management involves technology and practices for improving the alignment of decisions and actions with strategic goals across the enterprise. After initial progress, however, many organizations find that performance management can fall into a trough of disillusionment if efforts are limited to static scorecards and metrics that allow for little or no interaction. Organizations are finding greater success if they can deploy dashboards that deliver a call to action in the form of checklists or other actionable visual representations so that employees are encouraged or required to do something with the information. Making BI and performance management actionable will become even more important as mobile device adoption grows; TDWI Research found that performance management metrics and key performance indicators (KPIs) are considered the most important features to have in mobile BI applications.

Making information actionable, however, should not be a license to increase the noise level so that employees are lost in a cacophony of blinking alerts and urgent missives. In addition, if performance management systems are implemented departmentally and not integrated across the enterprise, the KPIs and metrics may lead different parts of the organization to take actions that are at cross-purposes. Disconnected performance management systems may not even implement consistent data or metrics definitions, making results across departments hard to compare.

Thus, to increase success and return on investment, organizations need to make performance management both actionable and integrated across the enterprise. These two imperatives are not unrelated; cross-functional business objectives such as improving customer satisfaction require executives and managers to use metrics to understand how actions in one business function such as marketing impact results in another function, such as a customer service contact center. In 2012, we expect

to see leading organizations improve performance management through more actionable functionality and better enterprise integration.

4. **Mobile BI will spur creativity to satisfy ease-of-use demands.** Apple iPads and similar mobile devices are all the rage, and not just among consumers interested in playing games and watching movies. Business executives—and soon, frontline sales, service, and support personnel—are excited about their potential for BI. TDWI Research just published a new Best Practices Report, *Mobile Business Intelligence and Analytics: Extending Insight to a Mobile Workforce*. In our analysis of the report's research survey results, we found that while executives are currently the dominant users of mobile BI, "improved customer sales, service, and support" leads the way as the business benefit most organizations seek to achieve (see Figure 2).

Mobile BI and analytics applications can give sales, service, and support personnel on the go an information resource that has previously been unavailable. Rather than lug around binders full of outdated, paper-based materials, mobile BI applications can enable sales, service, and support personnel to interact with data through frequently updated reports and dashboards. If mobile BI applications are integrated with communication and collaboration systems,

personnel on the go can share information—such as photographs and audio files—with colleagues so that customer issues may be resolved sooner.

Mobile BI will put pressure on the BI community to develop and deploy applications and services that are easier to use, with clear self-service features that do not frustrate nontechnical users. These necessities will drive a new era of creativity in the visual presentation of data on mobile devices. Already, leading organizations such as GUESS? Inc., a 2011 TDWI Best Practices Awards winner, are deploying dashboards that go well beyond standard presentations of data graphs and tables to provide visual icons, such as pictures of products that are familiar to nontechnical users. Mobile BI adoption will increase the emphasis in BI development on easy, visual interaction with data by a wide range of users.

5. **BI technologies and development methods will improve agility.** High-quality and timely data access and analysis are becoming essential to every decision and process. Organizations are under pressure to adjust strategies and tactics to fast-changing markets and economic situations, and they are depending on BI and analytics to help them make good decisions. These factors make delivering business value sooner and being adaptable to change two of the most important objectives for BI

**What are the business benefits that your organization seeks to achieve from implementing BI and analytics on mobile devices? (Please select all that apply.)**



**Figure 2.** Source: TDWI, January 2011. Based on 1,306 responses from 406 respondents; a little over 3 responses per respondent, on average.

projects. Unfortunately, the complex task of assembling all the “moving parts” in BI systems plus use of traditional waterfall development cycles have made it hard to meet these objectives.

*Agility* is about being flexible and responding quickly to change. Making BI systems more agile so they can support the dynamic needs of businesses requires significant changes in development approaches. Leading organizations are now employing agile software development methods, either strictly or in spirit. With these methods, developers can work more closely and iteratively with users so that changing requirements can be incorporated and users gain value sooner.

Methodology is not the only answer for increasing agility. Technology evolution toward analytic appliances, tightly integrated and prepackaged data warehousing and integration, and options for BI and data warehousing in the cloud are making it possible for organizations to be more dynamic and reduce long configuration and provisioning steps. Self-service capabilities in BI and data discovery tools are easing the burden on IT developers and are enabling users to implement the tools they actually need. In 2012, agility will be an important goal for many organizations; fortunately, technology and practices are becoming more capable of supporting that goal.

#### **INFORMATION INSIGHTS TO COME**

Exciting times lie ahead in BI as innovations in analytics, visualization, collaboration, and other areas mature to become core components of BI systems and services. This article has covered five major trends shaping the future of BI; I am sure readers could identify more. TDWI would love to hear what you think. We invite you to participate in our [LinkedIn community](#) or communicate with us in more traditional ways to let us know what trends are most important to your organization’s implementation of BI, analytics, and data warehousing in the New Year.

---

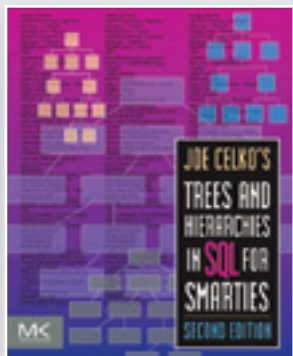
*David Stodder is director of TDWI Research for business intelligence. He focuses on providing research-based insights and best practices for organizations implementing BI, analytics, data discovery, data visualization, performance management, and related technologies and methods. Stodder has provided thought leadership about BI, analytics, information management, and IT management for over two decades. Previously, he headed up his own independent firm and served as vice president and research director with Ventana Research. He was the founding chief editor of Intelligent Enterprise and served as editorial director for nine years. He was also one of the founders of Database Programming & Design magazine. You can reach him at [dstodder@tdwi.org](mailto:dstodder@tdwi.org).*

# Morgan Kaufmann is your Source for the Best Business Intelligence References

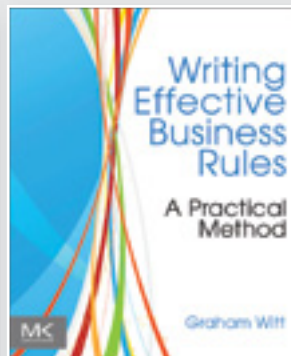
## NEW AND NOTABLE



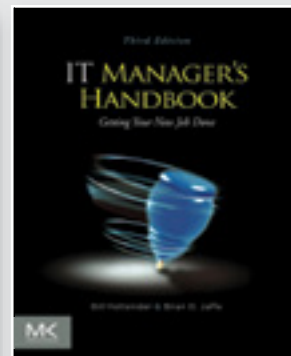
ISBN: 9780123850171



ISBN: 9780123877338



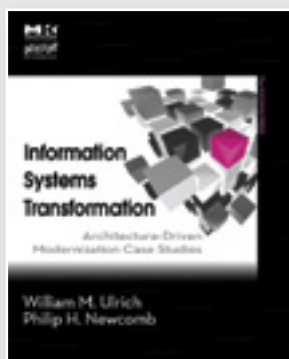
ISBN: 9780123850515



ISBN: 9780124159495



ISBN: 9780123737175



ISBN: 9780123749130



ISBN: 9780123819727

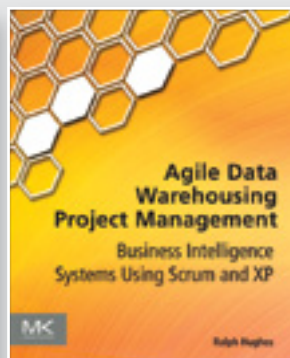
## COMING SOON



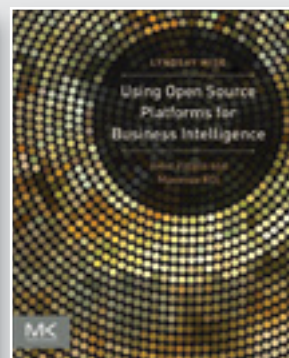
ISBN: 9780124160446



ISBN: 9780124158290



ISBN: 9780123964632



ISBN: 9780124158115



Save 20% when you order at [mkp.com](http://mkp.com)  
Enter discount code TDWI20 at checkout. Free shipping in the US!



TDWI RESEARCH

## THE STATE OF OPERATIONAL DATA WAREHOUSING

BY PHILIP RUSSOM

### Status of OpDW Implementations

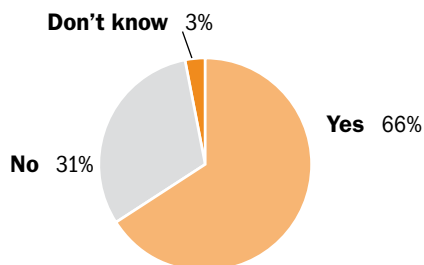
This discussion of operational data warehousing and related matters may make you wonder how many user organizations are actually doing it. To that end, this report's survey asked respondents: "Does your organization practice some form of operational data warehousing today?" (See Figure 1.)

**Two-thirds of organizations practice some form of OpDW today.** Sixty-six percent of survey respondents answered "yes," which gives OpDW users a clear majority over the 31% who answered "no." Note that the 66% includes a wide range of technologies and practices, whether operating in real time, near real time, on demand, intraday, or latently overnight. The wide range explains why so many respondents claim they are doing OpDW today. By comparison, far fewer are pushing OpDW to the extremes of real time today, as revealed by the 17% figure cited earlier.

**OpDW awareness is high.** Survey respondents seem well-informed about OpDW; only 3% said they don't know whether their organization is doing it. This suggests that users are discussing the various forms of OpDW with

their peers and management; otherwise, they wouldn't know whether it's being used. In fact, the results of this survey in general show a high level of awareness of OpDW and its variations, more so than with TDWI surveys about other topics.

#### Does your organization practice some form of operational data warehousing today?



**Figure 1.** Based on 446 responses.

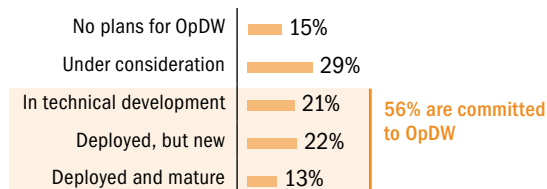
To get a rough sense of what life cycle stages organizations are in with OpDW, the survey asked: “What’s the status of your organization’s OpDW implementation?” (See Figure 2.)

#### Roughly half of organizations are committed to OpDW.

More than half of user organizations surveyed (56%) have made a commitment to some form of operational data warehousing, whether their solution is currently in development or already deployed. Given the wide margin of accuracy of this type of survey, the 56% committed to OpDW in Figure 2 is in the same ballpark as the 66% claiming the use of OpDW in Figure 1. Both figures corroborate that OpDW is commonly used today, but with ample room for growth in the future.

**More organizations are likely to commit soon.** A large percentage of organizations currently have OpDW under consideration (29%), and no doubt many of these will progress into OpDW usage. Furthermore, relatively few organizations have no plans for OpDW (15%). For these reasons (and for the reasons stated earlier in “Why Care About OpDW Now?”), TDWI Research feels confident that OpDW usage will increase.

#### What’s the status of your organization’s OpDW implementation?



**Figure 2.** Based on 446 responses.

### Benefits of OpDW

OpDW can result in benefits that positively affect data, the enterprise, and the management of each. To get a sense of which benefits are more likely than others, this report’s survey asked: “Which of the following would improve in your organization if you implemented some form of OpDW?” The most likely benefits (seen at the top of Figure 3) are those most often selected by survey respondents, and the likelihood of a benefit declines as the list proceeds downward.

#### Business intelligence tasks are the top beneficiaries of OpDW.

Almost half of respondents selected business decisions and strategies (46%) as things that would benefit from OpDW, and these are typically enabled by BI tools and techniques. Related to BI, data-driven corporate objectives (14%) and views of the business via data (13%) are also potentially benefited by OpDW.

#### Business operations are likely to gain from OpDW.

Survey responses indicate strongly that OpDW can improve business performance and execution (39%) and operational excellence (24%). In related operational issues, OpDW can also contribute to the efficiency of business operations (33%) and perhaps even lower the cost of business operations (19%).

#### Data currency improves when OpDW operates in real time.

As established earlier in this report, integrating operational data in real time or with intraday frequency is a common technical component of OpDW, though not required for all applications. Given that most OpDW implementations include these capabilities, it’s no surprise that survey responses identified data freshness or timeliness (35%) as an area improved by OpDW. When the currency of data improves, so does employees’ ability to be proactive (17%).



**Data sharing is expanded by OpDW.** With that in mind, OpDW can improve the business leverage of data assets (31%), data sharing across business units (23%), and cross-unit business processes (16%).

**Improvements to most data attributes is unlikely with OpDW.** Data currency and sharing aside, other attributes of data and its management ranked somewhat low in the survey. According to respondents’ perceptions, OpDW is not as likely to improve data’s quality (19%), governance (16%), architecture (15%), metadata (9%), or models (9%).

To summarize, the leading potential benefits of OpDW are improvements to business intelligence, business operations, data currency, and data sharing.

**Which of the following would improve in your organization if you implemented some form of OpDW?**

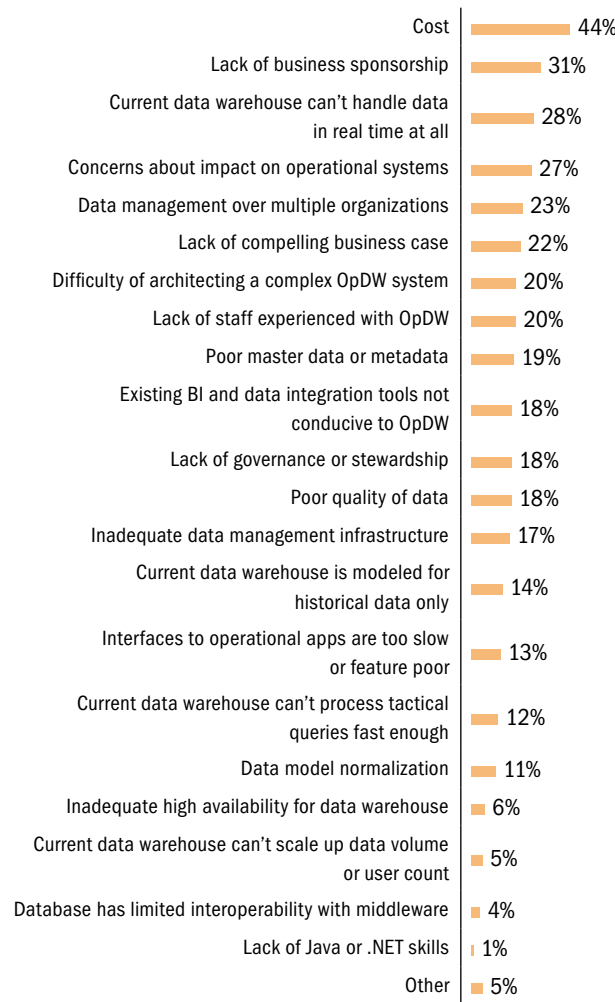


**Figure 3.** Based on 1,914 responses from 446 respondents, 4.3 responses per respondent on average.

**Barriers to OpDW**

OpDW has its benefits, as we just saw. Yet, it also has its barriers. Again, to get a sense of which barriers are more likely than others, this report’s survey asked: “In your organization, what are the top potential barriers to implementing OpDW?” The most likely barriers (seen at the top of Figure 4) are those most often selected by survey respondents, and the likelihood of a barrier declines as the list proceeds downward.

**In your organization, what are the top potential barriers to implementing OpDW?**



**Figure 4.** Based on 1,687 responses from 446 respondents, 3.8 responses per respondent on average.

**Cost (44%) is the most likely barrier to OpDW adoption and success.** This is natural, given that economies around the world are struggling to recover from one of the worst economic recessions in recent centuries. Of course, cost consciousness is still heightened, and many IT budgets are still cut, locked down, or tightly controlled. But there's more to it than the economy.

In the recent past—say, as late as 2002 or so—implementing OpDW usually meant purchasing and implementing additional technologies for real-time operation and interoperability between operational applications and BI/DW tools. The situation today is very different. As we'll see in the next section of this report, organizations that have kept up-to-date with BI/DW platforms and IT infrastructure already have all or most of what they need for many configurations of OpDW. Yet, the outdated perception persists that OpDW demands the acquisition of new and expensive tools.

**As with most initiatives, OpDW won't succeed without the business behind it.** Perhaps that's what survey respondents were thinking when they selected certain barriers, such as the lack of business sponsorship (31%), lack of compelling business case (22%), and lack of governance or stewardship (18%).

**The weaknesses of a data warehouse platform or team can be impediments to OpDW.** For many users, the sad fact is that their current data warehouse can't handle data in real time at all (28%). TDWI's 2009 research into next-generation data warehouse platforms showed that migrating to a platform that is inherently real time is the leading reason for replacing a data warehouse platform. Closely related barriers arise when existing BI and data integration tools are not conducive to OpDW (18%) and the current data warehouse can't process tactical queries fast enough (12%). Some users worry about the difficulty of architecting a complex OpDW system (20%), which is exacerbated by their lack of staff experienced with OpDW (20%) and the fact that the current data warehouse is modeled for historical data only (14%).

**Problems with enterprise data infrastructure can inhibit OpDW implementations.** These include poor master data or metadata (19%), poor quality of data (18%), inadequate data management infrastructure (17%), and interfaces to operational applications that are too slow or feature poor (13%).

To summarize, the leading potential barriers to OpDW are cost (whether actual or perceived), the lack of business support, weaknesses in the current data warehouse platform, and problems with enterprise data infrastructure.

---

*Philip Russom is the research director for data management at The Data Warehousing Institute (TDWI), where he oversees TDWI's research-oriented publications, services, and events. He is a well-known figure in data warehousing and business intelligence, having published over 500 research reports, magazine articles, opinion columns, speeches, Webinars, and more. Before joining TDWI in 2005, Russom was an industry analyst covering BI at Forrester Research and Giga Information Group. He also ran his own business as an independent industry analyst and BI consultant and was a contributing editor with leading IT magazines. Before that, Russom worked in technical and marketing positions for various database vendors. You can reach him at [prussom@tdwi.org](mailto:prussom@tdwi.org), [@prussom](https://twitter.com/prussom) on Twitter, and on LinkedIn at [linkedin.com/in/philiprussom](https://www.linkedin.com/in/philiprussom).*

This article was excerpted from the full, 32-page report, *Operational Data Warehousing: The Integration of Operational Applications and Data Warehouses*. You can download this and other TDWI Research free at [tdwi.org/bpreports](http://tdwi.org/bpreports).

The report was sponsored by EMC, HP Business Intelligence Solutions, SAP, SAS, Syncsort, and Teradata.

**TDWI** RESEARCH

## INTRODUCTION TO NEXT GENERATION DATA INTEGRATION

BY PHILIP RUSSOM

Data integration (DI) has undergone an impressive evolution in recent years. Today, DI is a rich set of powerful techniques, including ETL (extract, transform, and load), data federation, replication, synchronization, changed data capture, data quality, master data management, natural language processing, business-to-business data exchange, and more. Furthermore, vendor products for DI have achieved maturity, users have grown their DI teams to epic proportions, competency centers regularly staff DI work, new best practices continue to arise (such as collaborative DI and agile DI), and DI as a discipline has earned its autonomy from related practices such as data warehousing and database administration.

To help user organizations understand and embrace all that next generation data integration (NGDI) now offers, this report catalogs and prioritizes the many new options for DI. This report literally redefines data integration, showing that its newest generation is an amalgam of old and new techniques, best practices, organizational approaches, and home-grown or vendor-built functionality. The report brings readers up-to-date by discussing relatively recent (and ongoing) evolutions of DI that make it more agile, architected, collaborative,

operational, real-time, and scalable. It points to new platforms for DI tools (open source, cloud, SaaS, and unified data management) and DI's growing coordination with related best practices in data management (especially data quality, metadata and master data management, data integration acceleration, data governance, and stewardship). The report also quantifies trends among DI users who are moving into a new generation, and it provides an overview of representative vendors' DI tools.

The goal is to help users make informed decisions about which combinations of DI options match their business and technology requirements for the next generation. But the report also raises the bar on DI, under the assumption that a truly sophisticated and powerful DI solution will leverage DI's modern best practices using up-to-date tools.

## Ten Rules for Next Generation Data Integration

Data integration has evolved and grown so fast and furiously in the last 10 years that it has transcended ancient definitions. Getting a grip on a modern definition of DI is difficult, because "data integration" has become an umbrella term and a broad concept that encompasses many things. To help you get that grip, the 10 rules for next generation data integration listed on the next page provide an inventory of techniques, team structures, tool types, methods, mindsets, and other DI solution characteristics that are desirable for a fully modern next generation DI solution. Note that the list is a summary that helps you see the newfound immensity of DI; the rest of the report will drill into the details of these rules.

Admittedly, the list of 10 rules is daunting because it's thorough. Few organizations will need or want to embrace all of them; you should pick and choose according to your organization's requirements and goals. Even so, the list both defines the new generation of data integration and sets the bar high for those pursuing it.<sup>1</sup>

1. **DI is a family of techniques.** Some data management professionals still think of DI as merely ETL tools for data warehousing or data replication utilities for database administration. Those use cases are still prominent, as we'll see when we discuss TDWI survey data. Yet, DI practices and tools have broadened into a dozen or more techniques and use cases.
2. **DI techniques may be hand coded, based on a vendor's tool, or both.** TDWI survey data shows that migrating from hand coding to using a vendor DI tool is one of the strongest trends as organizations move into the next generation. A common best practice is to use a DI tool for most solutions, but

augment it with hand coding for functions missing from the tool.

3. **DI practices reach across both analytics and operations.** DI is not just for data warehousing (DW). Nor is it just for operational database administration (DBA). It now has many use cases spanning across many analytic and operational contexts, and expanding beyond DW and DBA work is one of the most prominent generational changes for DI.
4. **DI is an autonomous discipline.** Nowadays, there's so much DI work to be done that DI teams with 13 or more specialists are the norm; some teams have more than 100! The diversity of DI work has broadened, too. Due to this growth, a prominent generational decision is whether to staff and fund DI as is, or to set up an independent team or competency center for DI.
5. **DI is absorbing other data management disciplines.** The obvious example is DI and data quality (DQ), which many users staff with one team and implement on one unified vendor platform. A generational decision is whether the same team and platform should also support master data management, replication, data sync, event processing, and data federation.
6. **DI has become broadly collaborative.** The larger number of DI specialists requires local collaboration among DI team members, as well as global collaboration with other data management disciplines, including those mentioned in the previous rule, plus teams for message/service buses, database administration, and operational applications.
7. **DI needs diverse development methodologies.** A number of pressures are driving generational changes in DI development strategies, including increased team size, operational versus analytic DI projects, greater interoperability with other data management technologies, and the need to produce solutions in a more lean and agile manner.
8. **DI requires a wide range of interfaces.** That's because DI can access a wide range of source and target IT systems in a variety of information delivery speeds and frequencies. This includes traditional interfaces (native database connectors, ODBC, JDBC, FTP, APIs, bulk loaders) and newer ones (Web services, SOA, and data services). The new ones are critical to next generation requirements for real time and services. Furthermore, as many organizations extend their DI infrastructure, DI interfaces need to access data on premises, in public and private clouds, and at partner and customer sites.

<sup>1</sup> For a similar list with more details, see *TDWI Checklist Report: Top Ten Best Practices for Data Integration*, available on [tdwi.org](http://tdwi.org).



9. **DI must scale.** Architectures designed by users and servers built by vendors need to scale up and scale out to both burgeoning data volumes and increasingly complex processing, while still providing high performance at scale. With volume and complexity exploding, scalability is a critical success factor for future generations. Make it a top priority in your plans.
10. **DI requires architecture.** It's true that some DI tools impose an architecture (usually hub and spoke), but DI developers still need to take control and design the details. DI architecture is important because it strongly enables or inhibits other next generation requirements for scalability, real time, high availability, server interoperability, and data services.

## Why Care About NGDI Now?

### **Businesses face change more often than ever before.**

Recent history has seen businesses repeatedly adjusting to boom-and-bust economies, a recession, financial crises, shifts in global dynamics or competitive pressures, and a slow economic recovery. DI supports real-world applications and business goals, which are affected by economic issues. Periodically, you need to adjust DI solutions to align with technical and business goals for data.

**The next generation is an opportunity to fix the failings of prior generations.** For example, most lack a recognizable architecture, whereas achieving next generation requirements—especially real time, data services, and high availability—requires a modern architecture. Older ETL solutions, in particular, are designed for serial processing, whereas they need to be redesigned for parallel processing to meet next generation performance requirements for massive data volumes.

**Some DI solutions are in serious need of improvement or replacement.** For example, most DI solutions for business-to-business (B2B) data exchange are legacies, based on low-end techniques such as hand coding, flat files, and file transfer protocol (FTP). These demand a serious makeover—or rip and replace—if they're to bring modern DI techniques into B2B data exchange. Similar makeovers are needed with older data warehouses, customer data hubs, and data sync solutions.

**Even mature DI solutions have room to grow.** Successful DI solutions mature through multiple life cycle stages. In many cases, NGDI focuses on the next phase of a carefully planned evolution.

**For many, the next generation is about tapping more functions of DI tools they already have.** For example, most DI platforms have supported data federation for a few years now, yet only 30% of users have tapped this capability. Also to be tapped are newer capabilities for real time, micro-batch processing, changed data capture (CDC), messaging, and complex event processing (CEP).

**Unstructured data is still an unexplored frontier for most DI solutions.** Many vendor DI platforms now support text analytics, text mining, and other forms of natural language processing. Handling non-structured and complex data types is a desirable generational milestone in text-laden industries such as insurance, healthcare, and federal government.

**DI is on its way to becoming IT infrastructure.** For most organizations, this is a few generations away. But you need to think ahead to the day when data integration infrastructure is open and accessible to most of the enterprise the way that local area networks are today. Evolving DI into a shared infrastructure fosters business integration via shared data.

**DI is a growing and evolving practice.** More organizations are doing more DI, yet staffing hasn't kept pace with the growth. And DI is becoming more autonomous every day. You may need to rethink the headcount, skill sets, funding, management, ownership, and structure of DI teams.

---

*Philip Russom is the research director for data management at The Data Warehousing Institute (TDWI), where he oversees TDWI's research-oriented publications, services, and events. He is a well-known figure in data warehousing and business intelligence, having published over 500 research reports, magazine articles, opinion columns, speeches, Webinars, and more. Before joining TDWI in 2005, Russom was an industry analyst covering BI at Forrester Research and Giga Information Group. He also ran his own business as an independent industry analyst and BI consultant and was a contributing editor with leading IT magazines. Before that, Russom worked in technical and marketing positions for various database vendors. You can reach him at [prussom@tdwi.org](mailto:prussom@tdwi.org), [@prussom](https://twitter.com/prussom) on Twitter, and on LinkedIn at [linkedin.com/in/philiprussom](https://www.linkedin.com/in/philiprussom).*

This article was excerpted from the full, 32-page report, *Next Generation Data Integration*. You can download this and other TDWI Research free at [tdwi.org/bpreports](https://tdwi.org/bpreports).

The report was sponsored by DataFlux, IBM, Informatica, SAP, Syncsort, and Talend.

TDWI RESEARCH

# DATA VISUALIZATION TECHNOLOGY

BY WAYNE ECKERSON AND MARK HAMMOND

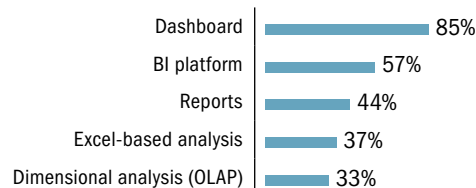
There are two main categories of data visualization technology: visual reporting and visual analysis.

- **Visual reporting.** Visual reporting uses charts and graphics to depict business performance, usually defined by metrics and time-series information. The primary type of visual report is a dashboard or scorecard, which gives users a visual snapshot of performance. The best dashboards and scorecards enable users to drill down one or more levels to view more detailed information about a metric. In essence, a dashboard is a visual exception report, highlighting performance anomalies using visualization techniques.
- **Visual analysis.** Visual analysis, on the other hand, enables users to visually explore data to discover new insights. While visual reporting structures the navigation of data around predefined metrics, visual analysis provides a much higher degree of data interactivity. With visual analysis, users can visually filter, compare, and correlate data at the speed of thought. Visual analysis tools also often incorporate forecasting, modeling, and statistical, what-if, and predictive analytics.

## Visual Reporting

**Dashboards.** By a sizable margin, dashboards are the preferred medium for data visualization. Nearly 85% of respondents ranked the importance of visualization as “high” in dashboards. On the other end, only 33% considered visualization to be highly important in an OLAP tool. (See Figure 1.)

Rate the importance of visualization in the following categories of BI tools.



**Figure 1.** Based on 210 respondents who ranked visualization's importance as “high.”

There are a near-infinite number of ways to design the visual elements of a performance dashboard. Most dashboards arrange a series of related charts in a grid template, usually two-over-two or three-over-three, and use multiple tabs or radio buttons to segment charts by category. They also usually display filters above or



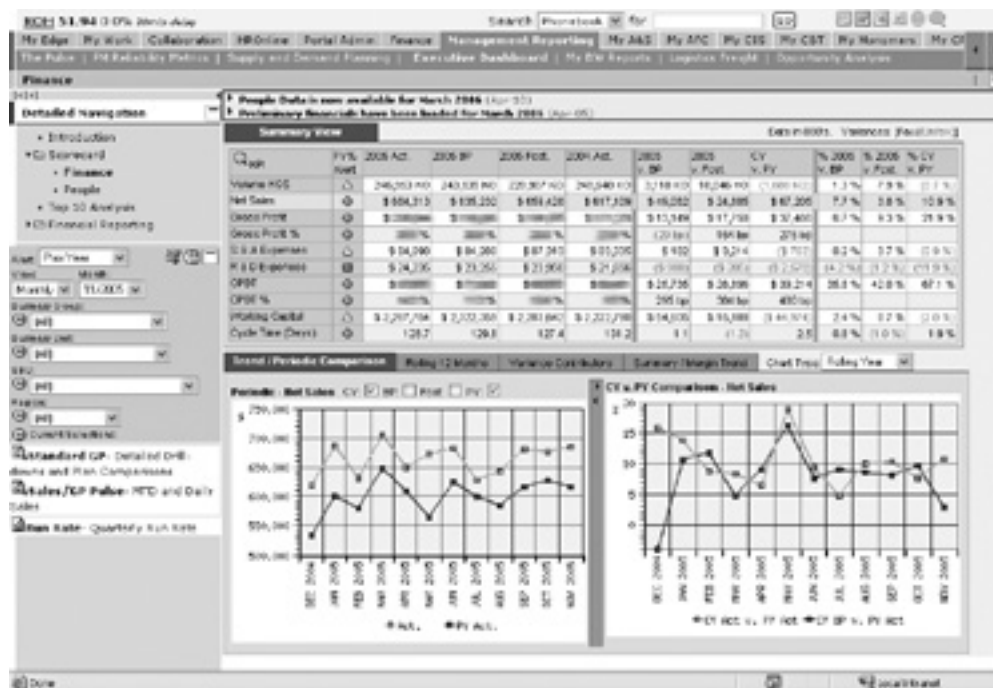


Figure 2. Rohm and Haas enterprise dashboard

beside the charts they apply to, as well as links to related dashboards or reports. The best dashboards display summary data graphically so it can be consumed at a glance and then provide access to any detailed information a user might need within three clicks.

**Sample Dashboard.** For example, the dashboard from Rohm & Haas (now owned by Dow Chemical) is embedded in the company's corporate portal, which has links (arrayed on top) to other enterprise content as well as other dashboards housed by the portal. (See Figure 2.) The dashboard itself consists of a table of 10 key performance indicators (established by top executives) that apply to every business unit and region in the company, along with pertinent targets (last year, variance, percentage change, and so on). Next to each metric are visual stoplights, which indicate the status of performance for the given metric against a selected target. Stoplights are perhaps the most common way to visually highlight exception conditions in a dashboard because they attract a user's attention quickly.

Below the grid are two somewhat interactive charts that show a time-series trendline for the metric highlighted by the user's cursor above. The left-hand panel contains the navigation path to the current view, and below that, a set of filters that users can use to change the alert in the grid and drill down to view performance along the same metrics at lower levels of the organization. (These filters are "universal"

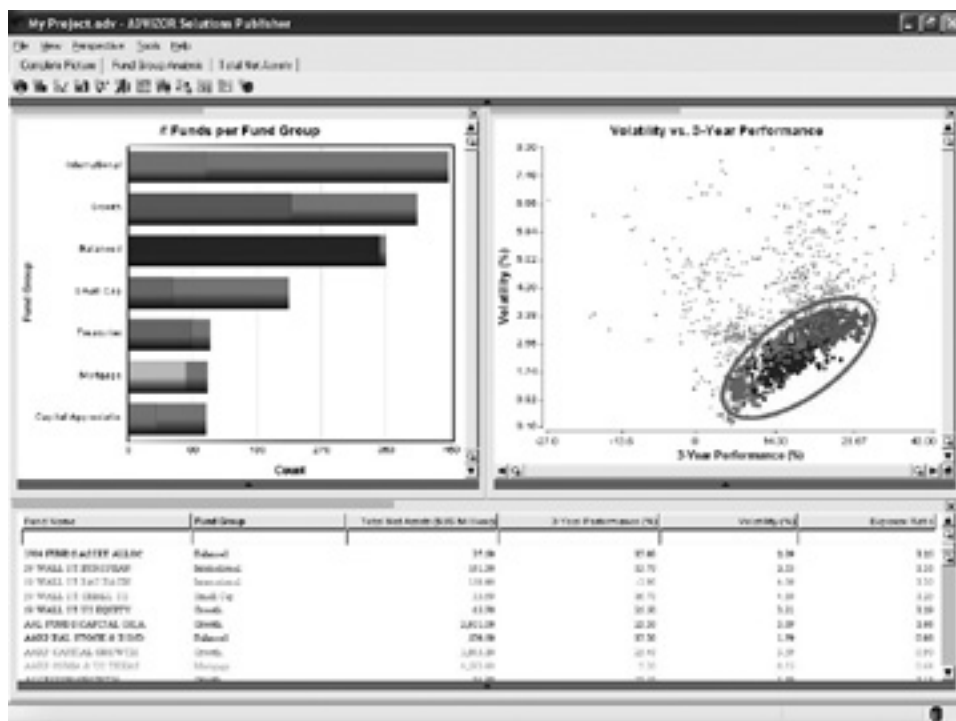
in that they apply to all objects on the screen instead of a single object.) The bottom of the panel contains hard-coded links to related dashboards and reports.

As you can see, the Rohm & Haas enterprise dashboard gives executives and managers a snapshot of performance for their domains with alerts to highlight exceptions and moderate levels of interactivity to drill into details and view related information. With a glance, executives and managers can see the status and trend of performance in their areas and how it compares to major benchmarks. Many companies are adopting this type of visualization to better monitor and manage performance.

## Visual Analysis

Visual analysis tools enable power users and business analysts (such as financial, marketing, and sales analysts) to explore data sets visually and identify trends and anomalies. These tools usually work with data stored in memory and expose rich navigational features that let users explore data at the speed of thought. Many also incorporate some form of statistical or predictive analytics.

Visual analysis tools compress and store data in memory, providing sub-second response times for any action taken against the data (such as filtering, drilling, calculating, sorting, and ranking). Visually, analysts point and click to interact with charts, apply filters, and change views.



**Figure 3.** Pointing and clicking can generate custom groupings.

For instance, analysts can use their mouse to “lasso” data points in a certain section of a scatter plot to create a new group and automatically filter other charts on the page. (See Figure 3.)

Compared to OLAP tools, visual analysis tools don’t require an IT person to design a dimensional data model. The tools use a “load-and-go” approach in which analysts load raw data from multiple sources and simply link tables along common keys to get a unified view of the data set. As a result, most visual analysis tools can be deployed in a few hours or a few days or weeks, depending on the number of data sources and their complexity and cleanliness.

Analysts or developers often use visual discovery tools to create and publish interactive, departmental dashboards for casual users. They often create the dashboards on desktop machines and then publish them to a departmental server for general consumption. When doing so, the developers generally strip out some analytical functionality and options that might overwhelm casual users.

**Two Environments.** It should be clear that visual reporting and visual analysis tools serve two different audiences and purposes. While visual reporting tools are designed to visualize performance against predefined metrics for executives and managers, visual analysis tools empower business analysts to explore trends and anomalies in data sets they create and publish views for others to consume.

## Visualization Technology

Both types of visualization solutions leverage emerging technology to enhance the visual experience of BI users. Here are key technologies driving the adoption of visualization in corporate environments.

- **64-bit systems and multi-core servers.** Charting engines chew up a lot of CPU cycles, especially if the charts are interactive. Rendering charts, especially in server-based environments, takes a lot of horsepower. Today’s 64-bit platforms and multi-core processors speed visual processing to give users more dynamic and interactive visual environments in which to view data.
- **RAM and compression.** Many visualization tools work with in-memory data to ensure speed-of-thought interactivity. With prices for RAM dropping, it’s easier for power users to analyze large data sets (up to 50 million records) held in memory. New compression techniques increase the amount of data that can be held in memory—but be cautious of decompression performance penalties.
- **Java applets/Active X controls.** These mini-applications run inside a Web browser and execute within a virtual machine or sandbox. Actions execute as fast as compiled code, making them an easy way to recreate

full-featured applications on the Web. However, they raise security concerns, so many IT administrators prevent users from downloading such controls through corporate firewalls, which limits their pervasiveness.

- **DHTML and AJAX.** A lighter-weight approach is to embed a scripting language inside HTML pages, such as JavaScript, that executes functions in the browser. Dynamic HTML (DHTML) uses scripting to animate a downloaded HTML page. For example, DHTML is often used to animate drop-down boxes, radio buttons, mouseovers, and tickers, as well as capture user inputs via forms. AJAX (asynchronous JavaScript and XML) takes this one step further and retrieves new content from the server in the background without interfering with the display and behavior of the page. Basically, AJAX enables users to add new data to the dashboard without having to reload the entire page. It can also be used to pre-fetch data, such as the next page of results.
- **Flash.** Another popular approach is to use multimedia development platforms, such as Adobe Flash, Java applets, Microsoft Silverlight, and Mozilla Scalable Vector Graphics (SVG), which add animation and movies to Web pages. Compared to Java scripting, these plug-ins provide stunning graphics and animation for displaying quantitative information, which makes the user interfaces very appealing to business users. They load both visualizations and data simultaneously in a single file rather than dishing up dozens or hundreds of pages. Although this makes the initial load slower than a comparable DHTML or AJAX application, performance thereafter is exceptionally fast, since the data required to display all components on a page resides locally.

**Vendor Advancements.** BI vendors have been scrambling to meet increasing demand for visualization. For instance, Oracle's release of Oracle Business Intelligence Enterprise Edition (OBIEE) 11g in mid-2010 addressed visualization weaknesses in earlier releases, Oracle officials said. Vendors such as MicroStrategy, ADVIZOR Solutions, and Tableau Software have recently emphasized new in-memory capacity for greater scalability. SAS (with its JMP visualization software) and DSPanel are among vendors incorporating the open-source R statistical programming language to mix visualization and data mining.

Corda and Dundas, which both provide charting components and dashboard tools, have expanded their tool sets to give developers greater flexibility. Microsoft is

aiming to elevate Excel's profile for BI visualization with the 2010 release of PowerPivot, an add-on that helps Excel accommodate large-scale data and extends its visualization capabilities, Microsoft officials said. Similarly, PowerPivot can leverage new visualization capabilities available through SharePoint 2010 integration with Visio, they said.

Many of these innovations are aimed at untethering business users from a reliance on IT so they can analyze data in a visual environment. "It's an evolutionary thing," said Doug Cogswell, president and CEO of ADVIZOR Solutions. "We're used to using BI to view reports or KPIs, and now people want to move beyond reporting to visual analysis."

---

*Wayne Eckerson has been a thought leader in the data warehousing, business intelligence, and performance management fields since 1995. He is the author of the best-selling book Performance Dashboards: Measuring, Monitoring, and Managing Your Business, second edition (John Wiley & Sons, 2010). Wayne is the former director of education and research at TDWI, and currently director of research at TechTarget and president of BI Leader Consulting. He can be reached at [weckerson@techtarg.com](mailto:weckerson@techtarg.com).*

*Mark Hammond is a veteran contributor to TDWI and TDWI Research, including a number of research reports, TDWI's Business Intelligence Journal, and the TDWI Marketplace. He researched and co-wrote E-Business Intelligence: Turning Information into Knowledge into Profit (McGraw-Hill, 2000) with the former CEO of Business Objects. An award-winning journalist, Hammond's work focuses on the use of data to improve business and organizational performance. You can reach him at [mhammond@tdwi.org](mailto:mhammond@tdwi.org) or [mhammond@comcast.net](mailto:mhammond@comcast.net).*

This article was excerpted from the full, 20-page report, *Visual Reporting and Analysis: Seeing Is Knowing*. You can download this and other TDWI Research free at [tdwi.org/bpreports](http://tdwi.org/bpreports).

The report was sponsored by ADVIZOR Solutions, Corda, DSPanel, Dundas, IBM, Microsoft, MicroStrategy, Oracle, SAS, and Tableau Software.

TDWI RESEARCH

## Introduction to Self-Service Business Intelligence

BY CLAUDIA IMHOFF AND COLIN WHITE

### DEFINITION OF SELF-SERVICE BI

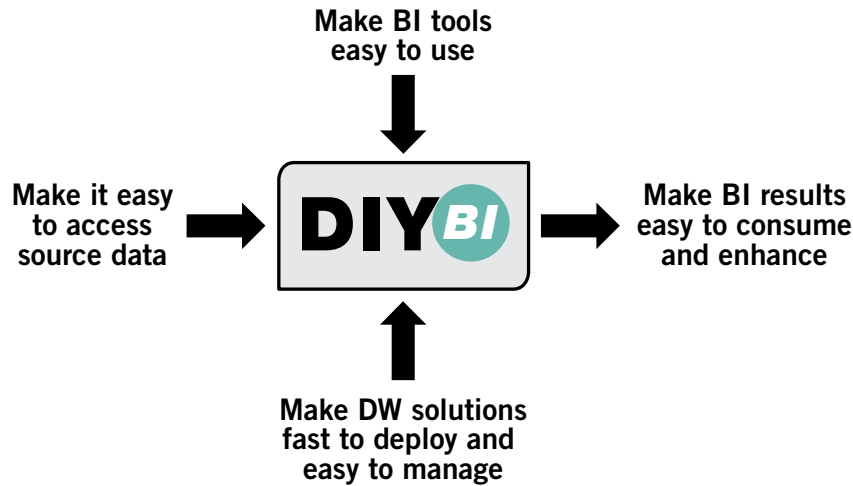
The facilities within the BI environment that enable BI users to become more self-reliant and less dependent on the IT organization. These facilities focus on four main objectives: easier access to source data for reporting and analysis, easier and improved support for data analysis features, faster deployment options such as appliances and cloud computing, and simpler, customizable, and collaborative end-user interfaces.

In today's economic environment, organizations must use business intelligence (BI) to make smarter, faster decisions. Business users must have better access to critical information at the right time and in the right format for comprehension. The business case for BI is well established; it gives companies their competitive edge and allows them to discover new business opportunities. Corporations and their employees need to be innovative and creative if they are to compete effectively. "Those with the imagination ... to invent smarter ways to do old jobs, energy-saving ways to provide new services, new ways to attract old customers, or new ways to combine existing technologies will thrive."<sup>1</sup> But in many organizations, countless decisions are still not based on business intelligence and analytics. Why? Certainly not from a lack of demand. Because of the changes in our economies, IT departments have been stripped down to the barest numbers, even as business users are demanding more control and faster access to BI and business data. From our survey of 587 technical and business professionals, we found that an overwhelming 78% of respondents stated that they needed a faster time to value from BI solutions.

To satisfy this demand and improve time to value, companies are looking for alternative approaches to BI. One approach is to set up an environment in which the information workers can create and access specific sets of

<sup>1</sup>Tom Friedman, *New York Times*, October 2009.  
<http://www.nytimes.com/2009/10/21/opinion/21friedman.html?ref=thomasfriedman>





**Figure 1.** The four objectives of do-it-yourself or self-service BI. (Courtesy of BI Research and Intelligent Solutions, Inc.)

BI reports, queries, and analytics themselves—without IT intervention. The purpose of this environment is to extend the reach and scope of BI applications to address a wider range of business needs and problems. At the same time, this extension must support the information workers' need for a personalized and collaborative decision-making environment. Information workers must become more self-sufficient by having a BI environment that is more usable and more consumable. It is these two themes—usability and consumability—that play crucial roles in a fully functioning self-service business intelligence (SS BI) environment.

## Self-Service Business Intelligence and Its Objectives

This section provides a brief overview and description of four key objectives of self-service BI (see Figure 1). The technologies supporting them will be covered in detail later.

### Make BI Results Easy to Consume and Enhance

This objective is probably the most important from the business community's perspective. Users must be able to grasp what the information presented to them means. A fire hose of information makes it difficult to determine where things are going off-kilter, where exceptions occur, or even get a handle on critical situations. SS BI must be an environment in which it is easy to discover, access, and share information, reports, and analytics. Information workers want to be able to personalize their dashboards or have automated BI capabilities so that the information

becomes actionable for their particular situations. It must also be in an easy-to-use format and delivered to a device and user interface of their choosing.

From a technical perspective, BI that is easy to consume and enhance requires clear business definitions that are easily accessible as well as data lineage that is tracked and documented. The organization improves its decision making by tracking interactions and decisions to discover, capture, and disseminate best practices. In addition, the information must be presented in a way that is easy for users to understand: data visualization and presentation are paramount for comprehension.

Finally, information workers increase knowledge content through interactions such as entering feedback on analytic results, models, and other BI results; adding business context on situations; and identifying related information such as external links, meteorological data, and other data that affects business activities. This allows the information workers to improve an organization's body of knowledge content. It also enables them to be more self-sufficient and make faster decisions. This feature helps BI implementers create environments that are appealing to business users and promotes the adoption of self-service.

### Make BI Tools Easy to Use

Not only do BI results need to be easy to consume and enhance, but the tools generating the results must also be easy to use. BI solution providers have focused on making these technologies easy to use for years and, for

the most part, vendors have succeeded in making them straightforward and simple. This is a significant factor in the success of any SS BI environment. It may help even novice information workers select their own reports and create simple analyses. It will certainly allow technologically savvy users to get what they need when they need it.

Although reporting and simple analytics interfaces have achieved a high level of ease of use, we still need to make the more complicated analytics easy to use, create, and publish. Sophisticated analytics are often too intricate, complex, or difficult to construct for many information workers. Support for such sophisticated analyses, as well as making results easy to publish in the required format, greatly improves the productivity of a company's information workers.

## Make Data Warehouse Solutions Fast to Deploy and Easy to Manage

Self-service business intelligence may mean looking at alternative deployment mechanisms to reduce costs, improve time to value, and support increasingly extreme data processing. Agile methodology, software-as-a-service (SaaS), and cloud offerings, as well as analytic DBMSs, all contribute to these goals. Key components of this objective include ensuring that the SS BI environment provides good performance and scalability for simple to complex analytical workloads and high data volumes. In addition, SS BI must support easy administration and enhancement of the environment in a timely manner.

Opening BI to access by the business community also enables applications to be built that may not have been possible with earlier architectures and technologies. Business units can deploy their own applications, tailored to their specific requirements, and on their own timetable. User satisfaction increases dramatically when this level of creation and management of reports and analytics is available to the business community.

## Make Data Sources Easy to Access

In our interviews, we heard a number of times that if you couldn't access the data, then nothing else mattered—whether it was traditional, IT-created BI components or SS BI. However, there is one significant difference with SS BI—not all the data accessed needs to be stored in a data warehouse. Data external to the data warehouse such as operational and external but relevant data (e.g., weather information, geographic, demographic, or psychographic

data) may need to be made available for access by the business community without IT assistance.

Self-service business intelligence may also require that all types of data be made accessible by the BI implementation team—not just traditional, structured data. This includes unstructured data such as comments or e-mails and even social media sources. The ability of the information workers to understand the full picture (which includes the context in addition to the content) is becoming mandatory. It may not be possible or even necessary to bring this contextual data into the data warehouse before it can be utilized by SS BI. If some of this data resides outside the warehouse, it means the BI environment must have some means or mechanism to federate data—bringing it together virtually from different sources for analysis and access.

The BI implementation team's job is to create the infrastructure that permits the free flow of data from all these sources. They can then monitor access and utilization of the data, ensure the environment's optimal performance, implement appropriate security and privacy procedures, and provide support to the business community where needed in the construction or publication of BI reports, analytics, and so on.

---

*Claudia Imhoff, Ph.D., is a popular analyst and dynamic speaker on business intelligence. She is the president of Intelligent Solutions, Inc., a data warehousing and BI consultancy. She has co-authored five books on these topics and writes articles and research papers for technical and business magazines. She is a TDWI Fellow and founded the Boulder BI Brain Trust. You can reach her at [cimhoff@intelsols.com](mailto:cimhoff@intelsols.com).*

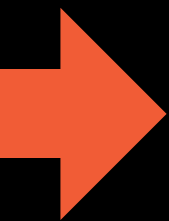
*Colin White is the founder of BI Research. As an analyst and educator, he is well known for his in-depth knowledge of data management, information integration, and BI technologies and how they can be used for building the smart and agile business. With many years of IT experience, he has consulted for dozens of companies throughout the world and is a frequent speaker at leading IT events. You can reach him at [cwhite@bi-research.com](mailto:cwhite@bi-research.com).*

This article was excerpted from the full, 36-page report, *Self-Service Business Intelligence: Empowering Users to Generate Insights*. You can download this and other TDWI Research free at [tdwi.org/bpreports](http://tdwi.org/bpreports).

The report was sponsored by Birst, IBM, Infobright, Neutrino Concepts, PivotLink, Quest Software, SAP, SAS, and Tableau Software.

# The only thing that SAP and Oracle suites integrate is the invoice

Birst is 1/3 the cost, time, and staff of SAP Business Objects and Oracle. And did we mention it is a single, unified solution delivering easy-to-use, enterprise-class capabilities?



Get your FREE copy of a ***Agile Business Analytics: How a New Generation of Business Intelligence is Changing the Market Landscape*** report at [www.birst.com/agile](http://www.birst.com/agile)



Agile Business Analytics



# TEN MISTAKES TO AVOID

## When Setting Your Cloud Business Intelligence Strategy

BY SHAWN ROGERS

### FOREWORD

One of the hottest technologies in the business intelligence (BI) space is the growing success of BI in the cloud. This new model has quickly been embraced by the business side of companies and has gained popularity for the way it can reduce capital and operating costs of business intelligence projects. Cloud BI is often identified as a fast way to implement projects, and its ability to grow and contract to meet the needs of BI environments is a plus. The upsides to cloud business intelligence are easy to identify, but like any other solution, you need a strategy before you begin your research to identify the right solution for your company. Skipping the due-diligence phase with cloud solutions can be just as costly as with traditional on-premises solutions.



### CONFUSING PRICE AND TOTAL COST OF OWNERSHIP

Cloud BI vendors offer a wide variety of pricing models, including per-seat licenses (similar to on-premises pricing models but generally lower in overall cost), data-constrained licenses (with charges based on the size or amount of data transferred to and from the system), and licenses based on the maximum number of concurrent users who need to access the system. Many companies make the mistake of getting locked into models that will not financially scale within their budgets. Over time, successful business intelligence programs grow, and as they do, the costs mirror that growth.

It's critical that growth be factored into the total cost of ownership (TCO) of a project. If the program is expected to exist beyond three years, it's possible that per-seat license models (especially those with significant growth) will end up costing as much as or more than traditional on-premises projects.

Another consideration specific to pricing is identifying how the prospective vendor will raise fees. Many vendors offer introductory pricing that escalates in the second or third year of a project. Concurrent-pricing models offer the most control and generally include unlimited seat licenses. The inclusion of a larger end-user community in a project can often lead to stronger adoption, but will eventually add pressure to system performance, requiring greater levels of concurrency.

Overall, it's critical to calculate your long-term project TCO and engage a pricing model that fits your immediate needs *and* your anticipated growth.





## IGNORING THE SECURITY CHALLENGE

When cloud business intelligence first arrived on the enterprise technology scene, security was a key hurdle to adoption. Many IT professionals were reluctant to relinquish control or power over their data security. This was especially true for companies that handle highly regulated data, such as financial services and healthcare.

Over the past several years, many pioneers in the cloud space have proven that the cloud is secure and ready for prime time. In many cases, the security infrastructure supplied by cloud BI vendors is more sophisticated than the data centers of their clients. Research from Enterprise Management Associates shows that security remains a top concern for companies when making investments in cloud technology, and it's critical that you take the proper steps to audit and confirm how much effort a cloud BI vendor has made to ensure data security.

Common security certifications include SAS70, SAS70 Type II, SysTrust, SSAE 16 Type II, ISAE 3402, and ISO 27001. Each offers different levels and perspectives on how data should be handled and secured. It's important that you identify the correct certification for your company.

In addition to these certifications, be sure to review the vendor's disaster and recovery procedures and backup and versioning processes, and whether these practices meet or exceed your company's compliance and regulatory guidelines.



## MISUNDERSTANDING FEATURES AND ARCHITECTURE

Assuming that cloud BI offers a generic plug-and-play environment and that one size fits all is a critical mistake when planning a cloud BI project. Every vendor offers different levels of features and configuration options. In the early stages of cloud BI, vendors offered 70 to 80 percent of the features delivered by traditional on-premises solutions. This is no longer the case; many leading firms are delivering similar feature sets to those delivered by traditional vendors.

Cloud vendors are generally able to deliver services at a lower cost than their on-premises competition by better leveraging hardware and infrastructure. Many software-as-a-service (SaaS) BI vendors utilize multi-tenant architectures and shared code bases to leverage a cost advantage. This economy of scale allows them to serve many customers from

a single platform but can, in turn, restrict companies that need to highly customize SaaS platforms.

Many SaaS vendors have created flexible frameworks that allow customization, but in the end, these platforms have limited flexibility. It's important to thoroughly investigate your customization needs prior to subscribing to a solution. You must define the costs involved in customization; these fees can add significant overhead to a SaaS project and thus can impact budgets significantly.



## FAILURE TO TEST PLATFORMS THROUGH PROOF-OF-CONCEPT PROGRAMS

Because many cloud BI vendors employ a multi-tenant production environment, provisioning a new instance or test environment is generally a trivial task. This provides an excellent opportunity for you to test before you buy.

Under traditional on-premises circumstances, it's not as easy or inexpensive to test projects as it can be in the cloud. As part of an overall cloud BI strategy, it's a mistake to skip this step. A proof of concept reduces the risk of the project and can also contribute to adoption once it's in production (because many stakeholders will have already used the system and had significant input during testing).

Depending on the scope of the project, many vendors are willing to provide free proof-of-concept programs to prospective clients. Taking advantage of test implementations can help you eliminate unexpected issues by testing processes, data integration, user acceptance, and service-level agreements.



## MISSING ENTERPRISE-LEVEL SLAS

Service-level agreements (SLAs) are a common component of any vendor/customer contract. When engaging a cloud service provider, apply the same standards of performance.

Speed of delivery is the cornerstone of these agreements, but other important considerations include a thorough understanding of how the cloud provider will deliver service-level monitoring, reporting, and issue resolution. Working your way through these policies and procedures is a critical part of the new paradigm of managing your IT environment outside the firewall.

Other service issues you must review in advance include how new users are added to the system, licensing management, and system elasticity. An advantage of software-as-a-service

(SaaS) over traditional on-premises solutions is that SaaS providers push system upgrades more often, dynamically updating the environment, so everyone benefits. It's important to explore the details about the schedule and process so you don't experience compatibility issues.

Don't just *expect* strong customer service. Not all vendors are created equal, so be sure to explore staffing and customer service assets to ensure service levels match your corporate needs. In the planning stages, be sure you understand the customer service hours, service ticket processes, and case escalation and reporting processes; they are essential to managing a successful cloud-based BI project.



### SKIPPING TEAM TRAINING

Don't skip training for end users and IT staff managing the cloud BI ecosystem.

An upside to working with the new breed of cloud BI vendors is that many platforms have been purpose-built for the cloud, enabling vendors to create the most up-to-date user interfaces and feature sets. In order for your end users to gain the highest level of value from the environment, you must support them with the best training possible.

Most vendors provide entry-level training programs that include virtual workshops or feature-centric classes. More sophisticated vendors are supporting online communities, discussion forums, and blogs. Many have invested in user groups that meet annually, onsite seminar services, and conferences. Matching your training needs to the services the vendor provides is a key component of long-term success for all business intelligence programs, especially those delivered via cloud technologies.



### LACK OF TRANSPARENCY

A unique aspect of cloud BI is communication. Moving system hardware, infrastructure, and the application layer to an outside vendor brings with it the need for new project management skills.

Regular meetings and performance reporting are important to keep the project on course, but require a high level of transparency from the vendor. This transparency must be in place as part of the vendor's culture, not just a simple policy. Processes for clear communication are required so you understand system performance, uptime, and relevant issues as they arise.

Research the vendor to see how they have historically handled failures. Request real-world examples of customer communications from recent system failures; beware of vendors that claim to have a perfect track record. No IT environment is failure free. Determine what level of detail the vendor shares and whether it's made public in a transparent fashion or cloaked in secrecy. Examine the vendor's public-facing communication mechanisms to determine if the vendor is forthcoming with the level of detail that will satisfy your company's needs. When the worst possible scenario occurs, you will be glad to have invested significant time and research into communication and transparency.



### VENDOR WEAKNESS

With every new technology offering comes innovative start-up companies to meet the challenge and deliver the solutions. Unfortunately, some start-ups innovate and then evaporate. This is an especially critical issue when dealing with cloud BI firms because your data and entire project are hosted on systems beyond your direct control.

If you invest in a start-up on-premises vendor, your risk is somewhat lower as you have complete control over the hardware, infrastructure, data, and application. Start-ups are generally privately held companies, so researching their financial stability can be difficult. If venture-capital firms are involved with the vendor, it's prudent to review those firms. If your investment so merits, request a meeting with key board members and investment representatives of the company.

Other key indicators of strengths and weaknesses can be found by reviewing the backgrounds of the executive team. Furthermore, research the company's partner ecosystem. Firms with limited partner support can be an indication of weak technology or limited resources.

The most important step in researching the strength of cloud BI vendors is to request domain-specific references. Don't skip this step under any circumstances. I recommend using this *Ten Mistakes to Avoid* as your guide when talking with reference accounts.



## ACCEPTING LIMITED TECHNOLOGY

Once you are utilizing a cloud business intelligence platform, you will want to link aspects of that platform to other enterprise systems. Sharing information from a cloud platform can be easy or difficult depending on the openness and the array of application programming interfaces (APIs) supplied by the solution provider.

Most purpose-built cloud solutions easily enable interoperability and system integration. Fast and simple access to the system can support security and compliance concerns and generally make the solution a more valuable asset to your overall BI ecosystem. Beware of platforms that are closed and offer limited API access. These systems will bring additional costs when you have integration challenges. Investigate whether there are additional charges for connectors to other popular platforms—these “hidden” costs can add up significantly and will increase development time over the course of your project.

As more enterprise applications migrate to the cloud, the ability to link and integrate the platforms and their data will be crucial to your overall success.



## NO EXIT STRATEGY

An exit strategy is one of the most overlooked parts in a successful cloud BI strategy. Many enterprises are so eager to engage cloud technology that they jump in without understanding how to jump out.

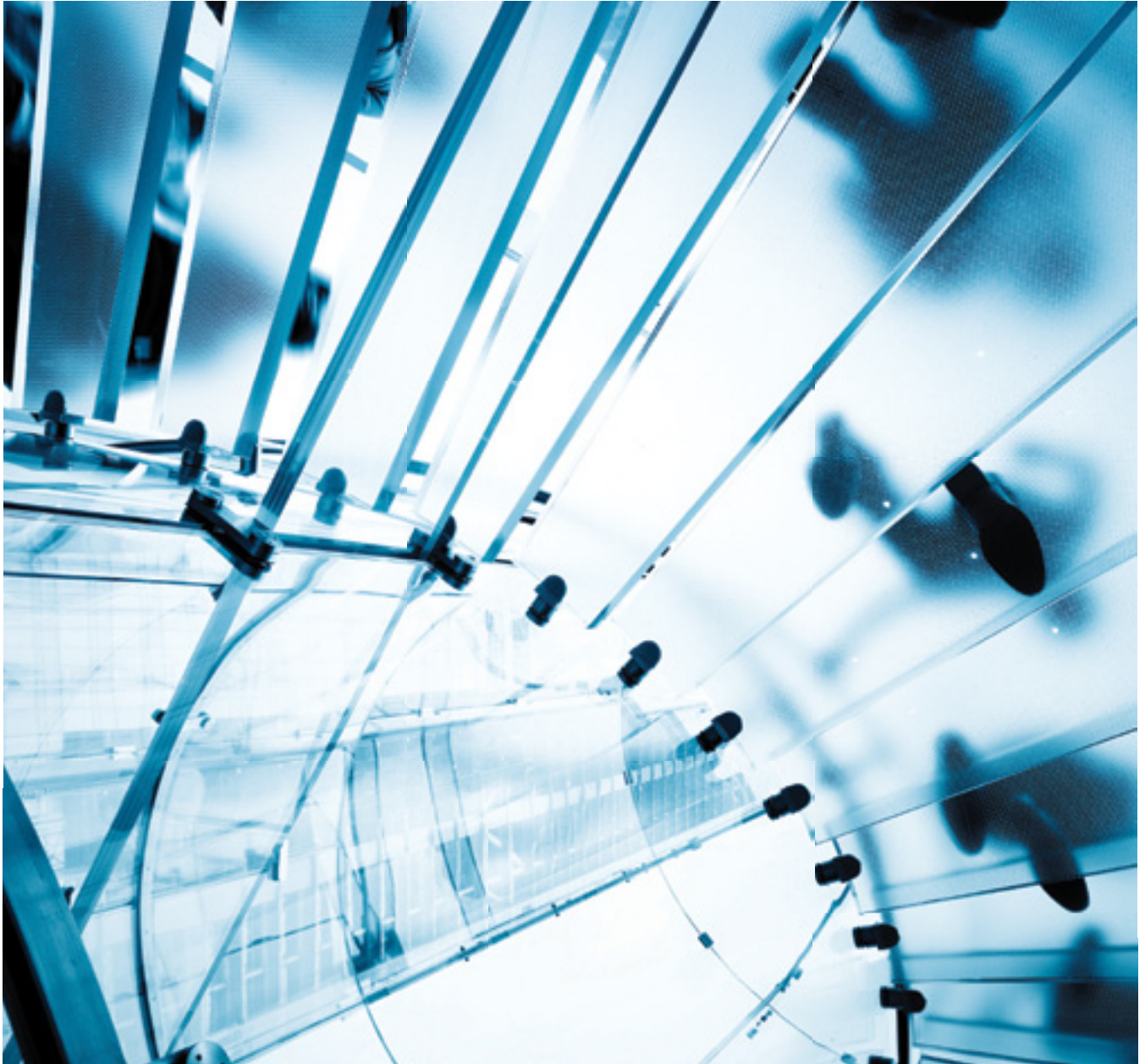
It's key to have an exit strategy, whether it's to move a project behind the corporate firewall or move to a vendor's competitor. Most cloud vendors don't publish their exit process, so you must understand what fees may be charged and whether there are any penalties to transition away from a cloud platform. In rare cases, disputes can arise over data ownership; understand in advance what the transition policies are. A well-documented work process for data migration and report migration should be defined up front. Without a clear exit strategy, you may experience significant problems later, from adding unforeseen costs to ending or moving a project.

*Shawn Rogers is vice president of research for business intelligence and data warehousing at Enterprise Management Associates, a research and consulting firm in Boulder, Colorado. Shawn lectures and teaches at conferences worldwide and has cofounded two successful media companies, DMReview magazine (now Information Management) and the BeyeNETWORK and its international community of Web sites focused on business intelligence and data warehousing. Shawn can be reached at [srogers@EMAusa.com](mailto:srogers@EMAusa.com).*

This issue of *Ten Mistakes to Avoid* was published in the third quarter of 2011. *Ten Mistakes* is an exclusive TDWI Premium Member benefit. For more information about TDWI Premium Membership, visit [tdwi.org/premiummembership](http://tdwi.org/premiummembership).

# Analytics that power business performance

- Self-service analytics for business users
- Visualization of process / performance issues
- Continuous monitoring and auditing
- Case management capabilities for issue resolution



Lavastorm Analytics is a global business performance analytics company that enables companies to analyze, optimize, and control the performance of their business processes. Using next-generation data discovery analytics, the Lavastorm Analytics Platform analyzes processes, finds defects, and builds controls to optimize complex financial, operational, and customer experience processes.

[www.lavastorm.com](http://www.lavastorm.com)

**LAVASTORM**  
analytics







# An Introduction to BI Life Cycle Management

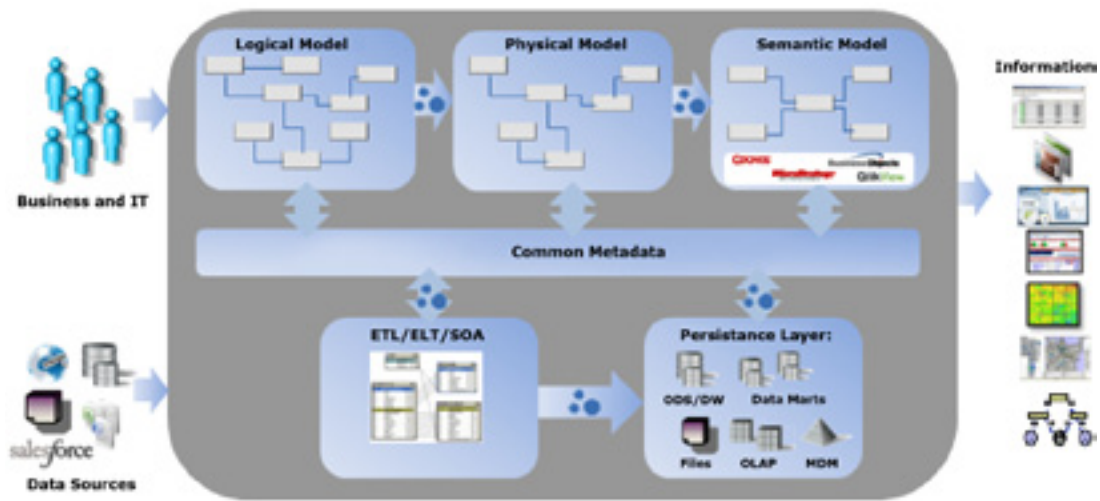
BY STEVE DINE

It's no secret that business intelligence (BI)<sup>1</sup> projects are time and resource intensive, often suffer from poor communication between the business and IT, and are usually inflexible to changes once development has started. This is due in large part to the method in which BI projects are traditionally implemented.

Regardless of the project management methodology you employ, a successful BI iteration requires business requirements identification, data analysis, data architecture and modeling, data integration (e.g., ETL, ELT, data virtualization), front-end development, testing, and release management—at the least. Whether you choose to integrate testing with development or employ prototypes and sprints, it doesn't change the fact that design, development, and testing are part of the process.

The problem with the traditional implementation of BI projects is that the design and development steps across the architecture aren't integrated and insulated from direct input by the business. In addition, because the tools usually employed for design and development aren't integrated, both initial development and subsequent

<sup>1</sup> Note: In this context, BI refers to both the data integration layer as well as the front-end tools for reporting, analyzing, and disseminating the information.



**Figure 1:** The BI life cycle process

changes require a significant level of manual effort and change management. If we want to improve the traditional BI development process, we need to approach this process as a business-driven life cycle and integrate and automate as much of the development process as possible.

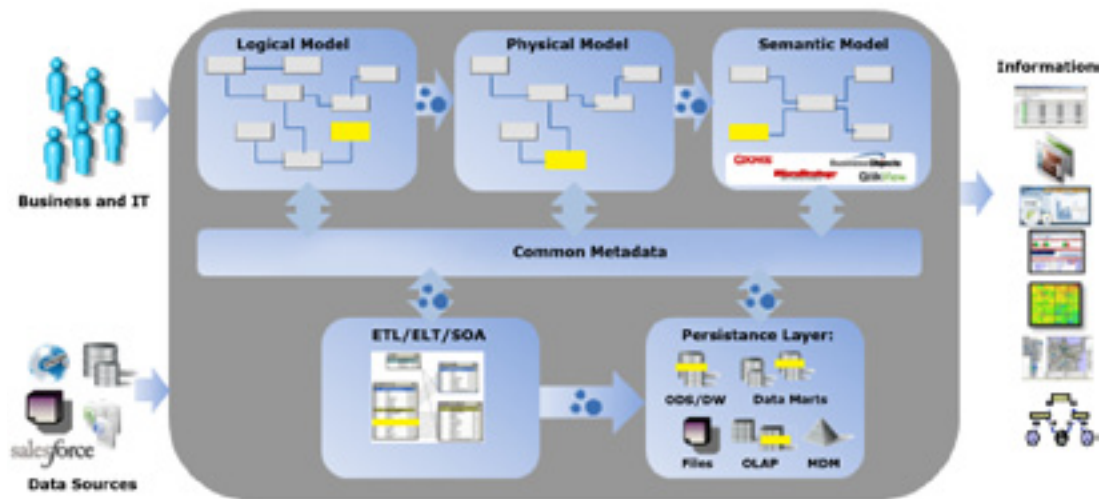
BI life cycle management is a design, development, and management approach to BI that incorporates business users into the design process by generating data models, database objects, data integration mappings, and front-end semantic layers based on input from business users. The challenge is that traditional BI projects leverage multiple, disparate tools used throughout the BI architecture to capture requirements; document business rules; perform source system analysis; model the data warehouse; and transform, store, report, and analyze the target data. In many environments, metadata management applications, data quality tools, and advanced analytic applications are also employed. Rarely do these tools share metadata, which makes it challenging to automate development and difficult to determine the impact when changes are required. In addition, business input is indirect because it must be captured, disseminated, and often recommunicated to the development team.

Fortunately, a number of BI life cycle management tools can facilitate this approach. With these tools, users are able to input project requirements, logical entities, relationships, business rules, source attributes, target attributes, and business metadata. These elements are

inputs to the logical model, which can be defined and then reviewed by the business. Once approved, a physical model is generated from the logical model, the database objects are generated automatically from the physical model, and the data integration (i.e., ETL/ELT/SOA) mappings are created, along with the objects in the persistence layer. (See Figure 1.) Some tools also generate the semantic models for front-end tool sets from the metadata captured during logical data modeling. Each step in the process is reviewable by the business and IT through a workflow process, and dependencies ensure that a change is reflected across the architecture.

The consistency between layers of the architecture is provided by the common metadata layer. Consider it an inventory of BI architecture assets with defined relationships. Because all the attributes, business rules, physical objects, and other elements are being captured, the dependencies can be identified and traced. This enables change-impact analysis and facilitates the testing process. The common metadata layer also creates a linkage between all the layers of the architecture so that each layer can be generated dynamically for rapid prototyping, which provides context to users throughout a BI implementation.

If a change is required to one entity in your logical model (in yellow, Figure 2), you can also see what objects would require changes across the entire architecture. In addition, with workflow and version management, changes can be controlled, tracked, prototyped, implemented, and rolled



**Figure 2:** BI life cycle change impact analysis

back if necessary. You could also view changes to the entire environment over time. Workflow automation ensures that the changes are approved at each step along the way.

The business community has long complained that BI takes too long, is too IT focused, and doesn't respond easily to change. If BI is to become more agile, we must approach the implementation process as an integrated life cycle rather than loosely connected steps. Although most BI life cycle tools are still immature (i.e., they only address a few areas of the complete process), they provide a glimpse into what's possible.

However, you don't need to purchase a BI life cycle tool to get started. BI projects and programs can incrementally add BI life cycle capabilities into their processes using existing tools. For example, many ETL applications support mapping definitions that can be generated from Excel spreadsheet templates, which can be populated by a data modeling tool. Although this doesn't address the entire life cycle, it does help you view the design and development process as a business-driven, integrated life cycle, and begin to change the way you implement BI.

*Steve Dine is the managing partner at Datasource Consulting, LLC. He has extensive, hands-on experience delivering and managing successful, highly scalable, and maintainable data integration and business intelligence solutions. Steve is a faculty member at TDWI and a judge for the TDWI Best Practices Awards. Follow Steve on Twitter: @steve\_dine. Contact him at [SDine@datasourceconsulting.com](mailto:SDine@datasourceconsulting.com).*

This article appeared in TDWI FlashPoint e-newsletter April 7, 2011. TDWI FlashPoint is an exclusive TDWI Premium Member benefit. For more information about TDWI Premium Membership, visit [tdwi.org/premiummembership](http://tdwi.org/premiummembership).



## Business Requirements for BI

BY DAVE WELLS

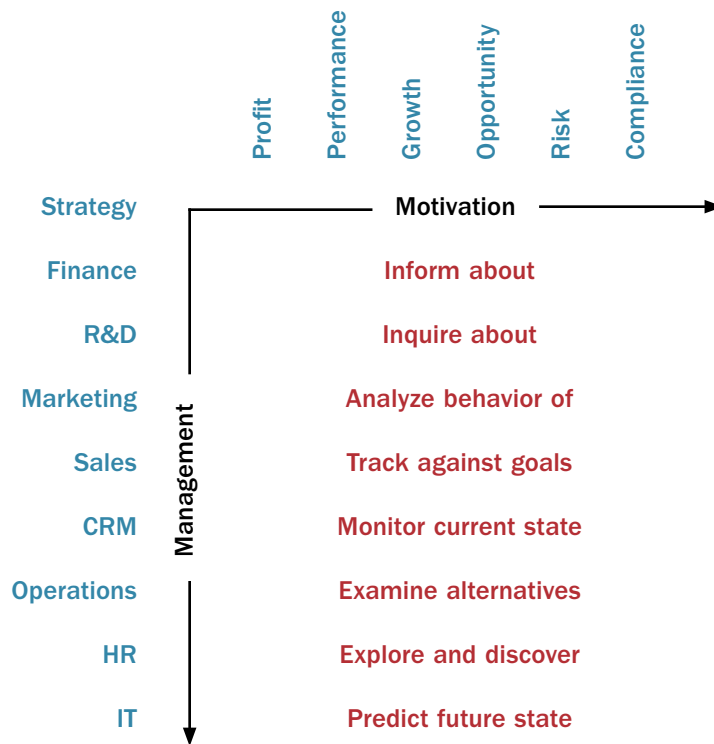
Getting good business requirements is one of the really hard problems in BI. Real and meaningful requirements are elusive. Business stakeholders often don't know what they need, can't express needs in BI terms, and revert to describing requirements as reports. Requirements analysts also struggle with BI requirements. The questions and methods that work for more traditional information systems projects just don't seem to work well for BI projects.

We need a different approach for BI requirements, but are mostly limited to the traditional techniques to elicit requirements—primarily, interviewing. To get good requirements, then, we need to change the interviewing process by asking different kinds of questions, and this article sets forth pragmatic questions that can help you capture real business requirements for BI.

### Setting the Scope

Asking the right questions begins with context, which you can create prior to requirements interviews by identifying the management functions and motivations that frame a BI project. Figure 1 illustrates this concept. The management axis lists common business functions—the *things* that we manage. The motivation axis lists typical goals of management—the *reasons* that we manage. Tailor both lists to industry and organization specifics. Engaging BI stakeholders in refining these lists is a good first step in asking the right questions.





**Figure 1.** Scope of BI business requirements

Within the structure of management and motivation, determine the kinds of business capabilities needed—inform, analyze, monitor, explore, predict, and so on. At the center of Figure 1 are common business capabilities that BI enables. Combining the three elements—management, motivation, and capabilities—creates clear and meaningful statements about the scope of a BI project, such as “monitor sales growth.”

After setting contextual scope, you can conduct a different kind of interview. Often the hard part of eliciting requirements is asking questions that shift thought processes from reporting-oriented views of information. The interviewer must ask new kinds of questions so the interviewee thinks about information delivery in different ways. Ultimately, we need to identify the BI products (e.g., reports, queries, scorecards, dashboards) that must be built. Getting to these products involves asking several kinds of questions: big picture, personalization, current state, and future state.

### Big-Picture Questions

These are “surveying the landscape” questions that help both interviewer and interviewee get into the right frame of mind to explore and describe real BI requirements. Breaking away from “just another report” thinking means that you need to step back far enough to view information within the context of how it is used.

- What are the expected goals of your area?
- How do you measure results?
- What are the critical success factors of your job?
- How do you identify opportunities and problems?
- What business dimensions are central to your analysis and decision processes?
- What are your current sources of information?

## Making It Personal

These are introspective questions for individuals. Getting each person to “see themselves” in BI can dig out real requirements. Use these questions to start the conversation about a personal view of business, information, and BI.

- What are the most important business goals?
- How does your job contribute to meeting those goals? How do you personally contribute?
- How does information help you contribute?
- How do you know if the goals are being achieved?

## Current State of Data and Analysis

This set of questions about spreadsheet use is an effective way to discover how people work with information and perform analysis. Excel is a good indicator of how individuals meet their own information needs. You’ll get different responses if you ask “How do you use Excel?” instead of examining how recent files are used. Begin with the recent files list, and extend your conversation to other uses. Some questions to apply this approach are:

- Please tell me about the content and purpose of each file on the recent files list.
- Where do you get the data for each file?
- What kinds of analyses do you perform? For what reasons?
- Which files are one-time-use files and which do you use frequently?
- What more would you like to do with this information? What are the barriers?
- Do you use any database or data analysis tools besides Excel?

## Current State of Reporting

Eventually you’ll need to explore existing reports to elicit BI requirements. Examine existing reports to understand what they represent as information and analysis needs—not to simply replicate the reports with different technology. Good reporting questions include:

- What existing reports do you rely on?
- How do you access existing reports?

- What is the frequency for each of the reports?
- What reports go together to make them relevant?
- What other tools do you use to analyze data from these reports?
- Are there specific features that you particularly like or dislike?
- Are the goals of the existing reports still relevant?
- What matters about the organization of each report (e.g., sequence, format, totals)?
- What reports do you cross-check for consistency and correctness?
- How do reports help you do your job?

## Future State

Future state returns to the question of capabilities. Ultimately, the requirements-gathering process needs to identify needs for capabilities at a deeper level than the scope setting. However, it doesn’t work to ask, “What capabilities do you need?” Nor can you find these requirements by providing a list—inform, inquire, analyze, track, monitor, examine, explore, predict—as if it is a restaurant menu.

You need to ask about the work of an individual or group and translate the responses into needs for business capabilities. Questions such as those listed below readily translate to business capabilities.

- What kinds of information do you depend on others to provide?
- What kinds of information do you provide to others?
- Do you see any trends in the kinds of analysis you perform?
- What kinds of business behaviors or results do you analyze?
- What is your role in setting and monitoring performance goals?
- What indicators do you watch on a daily or weekly basis?
- What needs do you have for forecasting and simulation?

BUSINESS CAPABILITY	BI PRODUCT
Inform about ...	Scheduled reports Ad hoc reports
Inquire about ...	Managed query Ad hoc query
Analyze behavior of ...	OLAP
Track ... against goals	Scorecards
Monitor current state of ...	Dashboards
Examine alternatives for ... Simulate behavior of ...	Analytic models
Explore patterns and trends of ... Discover hidden insights of ...	Data mining
Predict future state of ...	Predictive analytics

## Product Requirements

Business requirements for BI should detail the things you'll build to deliver capabilities—the BI products to be produced. The table below illustrates common mapping of business capabilities to BI products. Choosing the right product depends on understanding who will use the product and for what purpose.

A BI project will likely produce more than one product—for example, several products of the same type, such as many different scorecards. Different product kinds are also likely to be needed—a combination of dashboards and scorecards, for example.

When working with multiple products, consider how they are related. Common relationships among BI products include:

- A cascading hierarchy of dashboards or scorecards
- Drill-down from dashboards to scorecards or from scorecards to OLAP
- Drill-across from OLAP to OLAP

## Beyond Business Requirements

Requirements gathering that leads to identifying BI products will help you define functional requirements for each product. Once you know what you need to build, defining functional and technical BI requirements becomes much easier.

---

*Dave Wells is a consultant, teacher, and practitioner in information management with strong focus on practices to get maximum value from your data: data quality, data integration, and business intelligence. Contact him at [dwells@infocentric.org](mailto:dwells@infocentric.org).*

This article appeared in *TDWI FlashPoint* e-newsletter June 6, 2011. *TDWI FlashPoint* is an exclusive TDWI Premium Member benefit. For more information about TDWI Premium Membership, visit [tdwi.org/premiummembership](http://tdwi.org/premiummembership).

# Where will your data take you?



They say, "It's the journey, not the destination." When your data supports your business, you can enjoy both. Better data enhances business processes, helping you find ways to rapidly increase revenues, reduce costs and mitigate risks.

Innovative data management solutions from DataFlux help you transform data into a strategic asset. Business and IT users can collaborate to build the foundation for a smarter, more agile business. With award-winning data quality, data integration and master data management (MDM) solutions, DataFlux enables better, faster decisions – based on accurate and trusted data.

When better data drives your organization, the possibilities are endless.



Your data.  
Your business.  
One solution.

info.us@dataflux.com  
(877) 846-3589

www.dataflux.com

Data Quality | Data Integration | Master Data Management





## How Gamification Will Change Business Intelligence

BY MARSHA BURKE AND TROY HILTBRAND

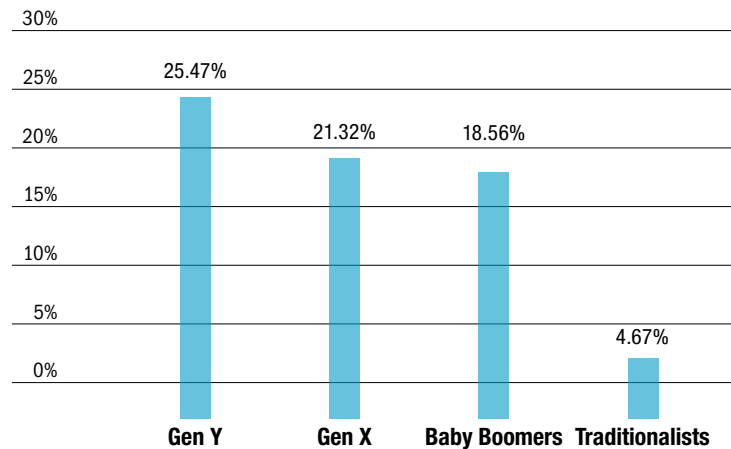


In 1958, William “Willy” Higinbotham was part of the Brookhaven National Laboratory (BNL) Instrumentation Division. In those days, BNL hosted a “visitors day” each fall. This was an opportunity for the public to tour the lab and get a better understanding of the capabilities and operations at a nuclear research facility. William Higinbotham wanted to do something that would impress the thousands of visitors as well as demonstrate the capabilities of the instrumentation division. His idea was to allow the audience to play a simulated game of table tennis. Using an analog computer and an oscilloscope, Higinbotham created the first publicly available video game.



Courtesy of Brookhaven National Laboratory

The game was very simple. It simulated the side view of a tennis court, showing the edge of the floor and the perpendicular edge of the net on a screen. Each player had a button and a rotating knob. Turning the knob changed the angle of the bar controlling the ball; pressing the button sent the ball toward the opposite side of the court. If the ball hit the net, it rebounded at an unexpected angle. If the ball went over and was not hit back, it would hit the floor and bounce again at a natural angle. If it disappeared off the screen, a reset button could be pressed, causing the ball to reappear and remain stationary until the player pressed the button to send the ball flying back across the net.



**Figure 1.** World population percentages by generation. Source: U.S. Census Bureau, 2009, International Data Base

Higinbotham recounted that “it was a simple design. Back then, analog computers were used to work out all kinds of mechanical problems. They didn’t have the accuracy of digital computers, which were very crude at the time, but then you don’t need a great deal of precision to play TV games.”

The game was wildly successful and Higinbotham could tell from the crowd’s reaction that he had developed something very special.

From the days of Higinbotham’s rudimentary version of the classic game Pong, video games have matured and become much more complex. The technology advanced, allowing for better graphics, sound, and game play, and significant research has been done in the area of game mechanics.

Game mechanics are rule-based systems that facilitate and encourage a user to explore and learn the properties of their possibility space through the use of feedback mechanisms.

— Raph Koster

As these game mechanics become better understood, they are being applied to areas outside of video game development. Game mechanics have become the basis for understanding the psychological drivers of users in the digital world. Business intelligence (BI), as a part of this digital experience, can learn some great lessons in how to engage users and revolutionize the user experience through the application of game mechanics.

In the tradition of BNL’s innovative spirit and as a sister laboratory in the Department of Energy laboratory complex, Idaho National Laboratory (INL) is exploring ways to enhance the user experience and meet the needs of our business through the effective application of game mechanics to BI.

### Why Gamification?

To understand why game mechanics concepts are emerging as important elements in all facets of business, it is important to understand the dynamics of the business environment. For decades, the demographics of the business workforce have consisted largely of baby boomers (those born between 1946 and 1964). As this generation nears retirement age, two generations will make up the majority of the workforce: Generation X (1965–1978) and Generation Y (1979–2000). See Figure 1.

Both Generation X and Generation Y view the workplace differently from their baby-boomer predecessors. Generation Y, also commonly referred to as “millennials” or the Net Generation, is of particular interest with respect to game mechanics. This generation has grown up with technology, and especially video games, as a part

of everyday life. These individuals don't remember a time when video games were not a staple in their lives. This means that game mechanics have had a prevalent impact on the way they solve problems.

In his book *Grown Up Digital*, Don Tapscott explains that:

Growing up digital has had a profound impact on the way this generation thinks, even changing the way their brains are wired. And although this digital immersion presents significant challenges for young people—such as dealing with a vast amount of incoming information or ensuring balance between the digital and physical worlds—their immersion has not hurt them overall. It has been positive. The generation is more tolerant of racial diversity, and is smarter and quicker than their predecessors. These young people are remaking every institution of modern life, from the workplace to the marketplace, from politics to education, and down to the basic structure of the family. ... To them, technology is like the air.

This changes the way that millennials engage with others and how they interact in the workplace. They come to work expecting the same engagement found in the digital world. The progression of workforce expectations means that challenges this group encounters will be overcome through gamification—the application of one or more reward structures to reinforce behavior driving the quality of and access to information.

### Success Stories

As we look at the application of game mechanics, we find that games have an astounding ability to engage people. As of the end of 2010, Farmville had more than 60 million users worldwide, or 1 percent of the world's population. At an average of 70 minutes played weekly, this is the equivalent of 70 million hours per week of engagement in a game where users maintain virtual farms. World of Warcraft, a popular massively multiplayer online role-playing game (MMORPG), has more than 11.5 million users who each pay \$14.99 per month to play. The population of this game is larger than that of Cuba or New York City. Each day, subscribers participate in 16.6 million quests and buy and sell 3.5 million virtual items via player auctions.

### Desired Outcome

Before we apply gaming concepts to business intelligence, we have to define the outcome we are striving to achieve. At INL, the Information Management department has a motto: Right information, right people, right time, right context. This encapsulates the spirit of business intelligence. It is about making sure that the right people, inside

or outside the organization, obtain the right information in a timely manner and that delivery of that information happens in the right context to enhance decision making.

As we look at the implementation of game mechanics to business intelligence, the objective is to drive behaviors based on business requirements that provide valuable information in a way that drives the right business decisions. To achieve this goal, we need employees and partners to participate in the information life cycle in the following ways:

1. Ensure data quality
2. Continually enhance information
3. Disseminate information to key decision makers
4. Personalize the user experience
5. Interact with information

### Game Mechanics

Game mechanics are composed of multiple facets of “game play” that can be assembled to ensure that the user is engaged and keeps playing. SCVNGR, which is a mobile game with real-world challenges, has revealed nearly 50 different game mechanics, or dynamics, that make up their playbook. We will look at some of these and how they can be implemented as part of business intelligence to drive desired behaviors among our user community (Schonfeld, 2010).

#### Achievement, Bonus, Point Dynamics

The *achievement* game dynamic relies on achievements, which are a virtual or physical representation of having expended special effort to accomplish something the user views as significant.

The *bonus* game dynamic relies on bonuses as the reward for having completed a series of challenges or core functions. Bonuses can be earned by completing a combination of tasks or by completing a specific task.

The *point* game dynamic relies on rewarding the user for completing any single action or combination of actions with a numeric value that is added to their overall point total. The cumulative nature of points drives users to continue to remain active.

These three game mechanics are closely related. They reward the user for performing a desired task or set of tasks. When we look at the paradigm of delivering the right information at the right time in the right context, there are three behaviors that we are trying to elicit.

First, to achieve the right information, a joint effort is required between information technology (IT) and the user base. The IT department is responsible for aggregating

information from multiple sources, passing it through data quality processes, and delivering it in a meaningful way to the user. The challenge remains that searching for information in today's business environment often results in information overload for key decision makers.

By applying a reward system that includes achievements, bonuses, or points, we can provide virtual rewards to users who increase the visibility of specific, important morsels of information key to decision making. Providing mechanisms to facilitate and track the dissemination of information through the BI platform allows rewards to be associated with those end users who disseminate the information.

The next desired behavior is information interactivity. For example, one challenge that persists in nearly every organization occurs each time an end user pulls raw data into Microsoft Excel and manipulates it to generate results for consumption by others. Because Microsoft Excel is not an enterprise reporting tool, the functionality of the data analysis is limited to an individual or a small group. It also creates multiple versions of the truth, which is counterproductive to providing the right information at the right time. When users are rewarded for interacting with data in ways that support the proliferation of information throughout the business and not only within their workgroup, they become accustomed to utilizing the tools that are available through the BI platform for the betterment of the organization. Their behavior is reinforced as they earn points for slicing, dicing, and drilling into information while maintaining its integrity.

The third desired behavior is information enhancement, which is initiated through the automated process of information aggregation and data quality enhancements. Achieving the full context of the business requires users' participation. Rewarding individuals for making comments on metrics or performing information classification or information value assessments can drive the appropriate user behavior as well as provide motivation to expend extra effort.

The rewards used to teach and incent user behavior can be simple, such as points awarded for each desired activity. The rewards can also be more random and fun, such as pieces of flair or bling or spontaneous awards for performing specific activities. Imagine a coworker receiving the elusive sparkly purple hammer, and the associated bragging rights, for completing their thousandth information classification action this month.

Users can also be rewarded by earning exclusive access to certain information after completing a set of predefined tasks. If individuals know that they can achieve access

to privileged information, they will act with a sense of urgency to get at that information.

These game mechanics rely on the user's desire to have bragging rights derived from accomplishing something important. These accomplishments do not have to be completed by an individual. They can also result from the efforts of a group, either organizational or virtual. This reinforces the mentality of joining forces as a team to achieve an end goal.

### Envy Dynamic

The *envy* game mechanic is based on the user's desire to have what others have. To be effective, visibility into what other people have (voyeurism) is essential.

BY APPLYING A REWARD SYSTEM THAT INCLUDES ACHIEVEMENTS, BONUSES, OR POINTS, WE CAN PROVIDE VIRTUAL REWARDS TO USERS WHO INCREASE THE VISIBILITY OF SPECIFIC, IMPORTANT MORSELS OF INFORMATION KEY TO DECISION MAKING.

Envy is closely related to achievement. This game mechanic requires the visibility of others' achievements. When users can see what other people have accomplished, they share in the desire to achieve the same. The social nature of game mechanics relies as much on perceived value as it does on actual value. The achievement, points, and bonuses might or might not have any physical value, but if the perception of the achievement is sufficient, it will establish a value that is often perceived to be of greater worth than a physical reward because of the envy others develop over that achievement.

In World of Warcraft, players obtain pieces of armor to enhance their characters and protect them from harm. The more difficult the game becomes, the higher the perceived value of the armor a player might win. By applying experience and skill, players can immerse themselves in the pursuit of rare, high-value items and eventually become the envy of the virtual realms. The mere chance of winning an epic item keeps millions of people motivated to organize 10-person and 25-person raid groups on a weekly basis.



### Leveling Up, Progression Dynamics

Levels are a system, or “ramp,” by which users are rewarded for their skill or effort by increasing their value as an individual. The process of “leveling up” is based on an accumulation of points or experience.

IF INDIVIDUALS KNOW THAT THEY CAN ACHIEVE  
ACCESS TO PRIVILEGED INFORMATION, THEY  
WILL ACT WITH A SENSE OF URGENCY TO GET  
AT THAT INFORMATION.

The *progression* game dynamic provides a visual representation of progress as users complete an itemized set of predefined tasks. As users make progress, they gain a sense of accomplishment. LinkedIn has utilized this dynamic to entice its subscribers to complete their profiles, adding value and further developing the content of the LinkedIn network. To progress, a LinkedIn subscriber has to upload a résumé, establish a network, and give and get recommendations before they can reach the magical 100 percent target.

This reward system can be employed to get users to customize their own experience. For instance, progress or levels could work as an incentive for users to set up the parameters that will drive their information portals and personalize their information experience. This could motivate them to set delivery preferences, customize their own dashboards, and arrange portlets on their home pages to provide the information they need.

*Leveling up* is a variation of progression. With progression, the concept is to hit a target. With leveling, once users achieve a target, they level up and receive a new set of tasks required to achieve the next level. This can be driven by point totals or by specific requirements. For example, when a user fills in her profile information, she could achieve the level of “Information User.” When she runs a specific number of reports, she could become a “Power Information User.” Finally, when she builds her own report, she could reach the level of “Information Builder.” If she is constantly adding value, she could attain the level of “Grand Master Information Czar.”

Once users reach a level, their accomplishment is celebrated and they enjoy the satisfaction of a job well done. The feeling is short-lived, though, because they are presented with their next set of challenges, which motivates them to move on.

### Community Collaboration, Companion Gaming Dynamics

The *community collaboration* game dynamic rallies an entire community to work together to solve a riddle, resolve a problem, or overcome a challenge. The concept behind this is to engage multiple people who feed off the energy of others in the group to keep moving forward. This dynamic is immensely viral and fun.

The *companion gaming* dynamic requires games that can be played across multiple platforms. Participation in the community collaboration can take place from a user’s desk or remotely on a mobile device. This dynamic perpetuates the continuous drive to accomplish a shared goal.

In late 2009, the Defense Advanced Research Projects Agency (DARPA) announced a challenge to test the way social networking could be utilized to accomplish a large-scale, time-sensitive task. The challenge was to find 10 eight-foot red weather balloons fixed in various locations across the U.S. in one day. The balloons were spread out so no single individual could find all of them. The top prize was \$40,000 to the team that found all 10; \$2,000 was given to each individual who first reported the coordinates of an individual balloon. A team from MIT won the challenge, identifying all 10 balloons in just nine hours by setting up a social network over the Internet and providing a recursive incentive mechanism to compensate those who participated. Again, this pseudo-multi-level marketing scheme was developed and executed in *less than nine hours* (Pickard, et al, 2010).

This example shows that when a group comes together, it can achieve things that are far beyond the capabilities of any single individual. By introducing challenges that must be addressed by community collaboration, we can harness both the combined efforts of diverse groups of individuals and the drive synergies that occur when groups feed off each other’s energy.

These groups do not have to be tied to an organizational structure. They can be both virtual and cross-functional, and they should be dynamically created to realize the energy that comes from the challenge itself. In the DARPA challenge, the winning team opened up participation through the Web to anyone who wanted to contribute. It was partially due to the open boundaries of their dynamic network that they were able to achieve results so quickly.

### Leaderboard Dynamic

This dynamic relies on generating competition between individual users through their desire to move up on a leaderboard.

I admit that I am addicted to Bejeweled, a game on the popular social networking site Facebook. Each and every day, I log on to Facebook to play. It is not a complex game; it requires a repetitive motion of swapping bricks to form lines of similar color, making the bricks explode and disappear. Part of the drive behind this is the leaderboard, which shows representatives from my Facebook network of friends. Every couple of weeks, the scores reset to zero and our specific group races to make it to the number one spot on the leaderboard. Many are friends I have not seen in years and do not share real-life interactions with, but because they are constantly on the leaderboard with me, I feel an affinity toward them. By merely seeing their scores, I am continually driven to achieve a higher score. I am consistently earning three times more points now than I did just a few months ago because of this ongoing drive to improve.

As BI users earn points and advance in their levels, displaying this to their peers creates a virtual competition that drives each individual to strive to achieve more. Keeping the leaderboard fresh and competitive keeps this game dynamic viable. Options to do this might include showing leaderboards based on groups (either organizational or virtual) or by focusing a leaderboard on a certain level of experience. Once users change levels, they can be bumped up to a higher-level leaderboard where the competition is more intense.

### **Appointments, Countdown, Extinction Dynamics**

Within the *appointments* dynamic, rewards are given when users participate at a predetermined time and virtual place.

The *countdown* dynamic affords players only a certain amount of time to accomplish something. This drives users to accomplish the task before it is moved into extinction.

The *extinction* dynamic refers to the action of ceasing to provide a reward.

Information delivery is often cyclical because of the scheduling mechanism for information aggregation. Once the information has been aggregated, it is often imperative to perform information value assessments and enhancements as quickly as possible so the data can provide the greatest value to consumers. By establishing a specific time when the information will be made available and specific rewards associated with performing these time-sensitive tasks, users can be encouraged to remain engaged and perform tasks in a timely manner.

A reward system of this type could include providing double points to users when they participate between 8:00 a.m. and 9:00 a.m. and then have the points drop dramatically after that time frame—or go away completely.

In cases when the action must be performed prior to a deadline, the ability to perform the activity could be completely removed, or moved into extinction, after a predefined milestone. Visible countdowns can reinforce that a task is time sensitive and requires an enhanced sense of urgency.

BY INTRODUCING CHALLENGES THAT  
MUST BE ADDRESSED BY COMMUNITY  
COLLABORATION, WE CAN HARNESS BOTH  
THE COMBINED EFFORTS OF DIVERSE  
GROUPS OF INDIVIDUALS AND THE DRIVE  
SYNERGIES THAT OCCUR WHEN GROUPS  
FEED OFF EACH OTHER'S ENERGY.

### **Epic Meaning Dynamic**

This dynamic motivates users because they believe they are working to achieve something great, something awe-inspiring, and something bigger than themselves.

With business intelligence, users are often provided with key measures and metrics but cannot envision how these metrics fit into the bigger picture of achieving business success. When larger goals are set and presented as the “epic meaning,” users begin to view that metric as a small piece of the business’s larger goals. Displaying a sales number for the week or the previous day is an important element in understanding operations, but seeing the same number in the context of the year’s sales goal gives it a strategic context and provides the user with the opportunity to engage in something that has epic meaning.

### **Ownership Dynamic**

The *ownership* dynamic represents a positive, sustained connection to an entity that leads to a feeling of shared ownership. This is often reinforced with a visual representation.

One of the challenges that most BI initiatives face is to propagate the ownership of data quality. IT is frequently blamed for the data “not being right,” when the ultimate responsibility of data quality is often a function of the business itself. The ownership associated with the quality of data can be integrated throughout the process to achieve a positive end result. If an end user runs a report and sees a photo of the data owner associated with that report, there is a direct tie to ownership. Another option

is to tie this to the leaderboard dynamic and provide a leaderboard of data owners. Feedback mechanisms can be incorporated to allow end users to indicate if the data has quality issues and provide rewards for leaders with fewer data quality issues.

## THE EPIC MEANING DYNAMIC MOTIVATES USERS BECAUSE THEY BELIEVE THEY ARE WORKING TO ACHIEVE SOMETHING GREAT, SOMETHING AWE-INSPIRING, AND SOMETHING BIGGER THAN THEMSELVES.

### Free Lunch Dynamic

The *free lunch* dynamic provides players with achievements they have not earned. In this way, they feel they are getting something for free because someone else has done the necessary work. To avoid breaching trust in the scenario, it is critical that work is perceived to have been completed. The player must be made to feel that they've lucked into something.

When additional information is “unlocked” thanks to the efforts of an individual, this dynamic could be used to provide other members of the team with the same unfettered access to the exclusive information. They would not have earned it themselves, but would get access merely by association with the individual doing the work. This creates a self-sustaining motivation for the individual performing the work because he or she gets recognition from the other members of the team for having accomplished something. The users receiving the free lunch remain engaged because they are given a privilege not open to others.

### Pitfalls

Game dynamics are not always a positive motivator. They can be detrimental if used incorrectly.

### Cross-Situational Leaderboards Dynamic

The *cross-situational leaderboards* dynamic occurs when one ranking mechanism is applied across multiple (unequal and isolated) gaming scenarios. Users often perceive that these ranking scenarios are unfair because not all players may have been presented with an “equal” opportunity to win.

With leaderboards, differing levels of task complexity can create a feeling of inequality. In the Bejeweled example, everyone is playing the same game with the same repetitive

moves, so there is a perception of equality. As we apply this to BI, the tasks are not all equal, and points need to be distributed in a way that enhances the perception of equality in the leaderboards so that fairness isn't questioned. Otherwise, leaderboards will move from being motivators to demotivators.

### Moral Hazard of Game Play Dynamic

The *moral hazard of game play* dynamic recognizes that there is a risk that, by rewarding people manipulatively in a game, you remove the actual moral value of the action and replace it with an ersatz, game-based reward. The risk is that by providing too many incentives to take an action, the incentive of actually enjoying having taken the action is lost. The corollary to this is that if the points or rewards are removed, the person loses all motivation to take the action, even if it was once fun.

Engaging users in a manner that plays on their desire to participate in a game sets up a paradigm where the game play must continue or the motivation to continue to perform learned behaviors is lost. Blizzard Entertainment, the maker of World of Warcraft, has maintained player engagement and lured new players to the game through a series of product expansions. Over the course of six years, Blizzard has released three expansions and kept millions playing. The gamification of business intelligence can quickly drive the desired performance and behaviors of the end users, but must be carefully planned to ensure its sustainability. Otherwise, it can have an adverse affect on the engagement of the user community.

### Privacy Dynamic

This is the concept that certain information is private and not for public distribution. This can be a demotivator (“I won't take an action because I don't want to share this”) or a motivator (“By sharing this, I reinforce my own actions”).

Just as the public-facing nature of gamification on the user interactivity with information can be a driver of certain desired activities, it can also be a demotivator. If the achievements gained by participation become negatively viewed as wasting time, users will shy away from doing anything with the information because of a perception that they are simply engaging for the sake of entertainment.

### Summary

As Generation Y takes the lead in organizations, the dynamics of business will change. These individuals will expect that the actions they perform at work should mimic the way that they engage to solve problems outside of work. As we move forward, the gamification of business intelligence will become more prevalent and must be understood and managed carefully to be successfully implemented.

ENGAGING USERS IN A MANNER THAT  
PLAYS ON THEIR DESIRE TO PARTICIPATE  
IN A GAME SETS UP A PARADIGM  
WHERE THE GAME PLAY MUST CONTINUE  
OR THE MOTIVATION TO CONTINUE TO  
PERFORM LEARNED BEHAVIORS IS LOST.

## References

- DOE R&D Accomplishments [2011]. Retrieved January 8, 2011, from "Video Games—Did They Begin at Brookhaven?" <http://www.osti.gov/accomplishments/videogame.html>
- International Data Base (IDB) [n.d.]. Retrieved January 8, 2011, from U.S. Census Bureau. <http://www.census.gov/ipc/www/idb/>
- Koster, R. [2004]. *A Theory of Fun for Game Design*, Paraglyph Press.
- Pickard, G., Rahwan, I., Pan, W., Cebrian, M., Crane, R., Madan, A., et al [2010]. "Time Critical Social Mobilization: The DARPA Network Challenge Winning Strategy." Retrieved January 8, 2011. <http://www.cs.umass.edu/~wallach/workshops/nips2010css/papers/pickard.pdf>
- Schonfeld, E. [2010]. "SCVNGR's Secret Game Mechanics Playdeck," August 25. Retrieved January 8, 2011, from TechCrunch. <http://techcrunch.com/2010/08/25/scvngr-game-mechanics/>
- Tapscott, D. [2008]. *Grown Up Digital: How the Net Generation is Changing Your World*, McGraw-Hill.

---

*Marsha Burke is high performance workplace analyst for Idaho National Laboratory. Contact her at [marsha.burke@inl.gov](mailto:marsha.burke@inl.gov).*

*Troy Hiltbrand is enterprise architect and IT strategy lead for Idaho National Laboratory. Contact him at [troy.hiltbrand@inl.gov](mailto:troy.hiltbrand@inl.gov).*

This article appeared in Volume 16, Number 2 of the *Business Intelligence Journal*, an exclusive TDWI Premium Member benefit. For more information about TDWI Premium Membership, visit [tdwi.org/premiummembership](http://tdwi.org/premiummembership).



# Integrating Structured and Unstructured Data

BY BARRY DEVLIN, HARRIET FRYMAN, MICHAEL GONZALES,  
DAVID SCHRADER, AND DAVID STODDER



Nanci Johnson is the very successful BI director at Teen Cosmetics, a company that—as the name suggests—specializes in the creation, manufacture, sale, and distribution of cosmetics for

teenagers. It sells its products through department stores and a company Web site. Under Nanci's leadership, Teen Cosmetics has developed a customer-centric data warehouse and a comprehensive set of online reports, dashboards, and other applications, including campaign management.

As is the case in most organizations, the BI focus has been on the storage and use of structured data. Nanci believes that this focus needs to change and be expanded to include unstructured data such as e-mail, information from conversations between customers and call center personnel, social media data, and so on. She believes that unless she takes leadership in this area, other organizational units such as marketing (which is already using Facebook, Twitter, and viral marketing) will claim this territory—or at least take some initial actions that will ultimately have to be supported by her team. For example, she has already had conversations with marketing staff interested in using sentiment analysis to understand how customers are reacting to new product introductions.

This is all new territory for Nanci. She knows enough to have a high-level conversation about the use of unstructured data, but that is about all. Can you help answer some of her questions?

1. Should the collection, storage, and analysis of unstructured data be the responsibility of the BI group, or is it better left to other departments such as marketing? What factors should affect where the initiative is organizationally housed?
2. How should she start the conversation with management about how to organize and best use unstructured data?
3. As Nanci tries to learn more about the use of unstructured data, what are some resources she should consider?
4. What are some of the technologies or tools she should be familiar with that are used with unstructured data?
5. What are some of the major differences between working with structured and unstructured data that Nanci is likely to find?



## Barry Devlin

Oh, dear! What trouble we've caused in the IT industry with the unfortunate terms "structured" and "unstructured" data. Not only does this suggest that these are two fundamentally incompatible categories, but it also leads to wild assumptions about differing technological, organizational, and implementation decisions whenever the topic arises.

Let's go back to basics. It's all information that we're dealing with! As I described in a past article ("Beyond Business Intelligence," *Business Intelligence Journal*, Vol. 15, No. 2), the hardened data we gather in operational systems and move into warehouses is information that has been translated into a highly structured form suitable for computer processing. This transformation of soft information was driven by many reasons, of which some (storage and processing limitations) are less relevant today while others (accuracy, consistency, and computability, and so on) remain vital.

Labeling some information as "unstructured" is not just technically incorrect—by definition, unstructured information is noise—but it also distracts us from the reality that this new, soft information such as e-mail, customer conversations, and social media data is actually a mixture of soft and hardened information from a wide variety of often unreliable, unverified, and unaccountable sources. It is disparate, continually variable in content and structure, and arrives in increasing volumes, already far larger than any we currently process in BI.

By now, I see Nanci rolling her eyes and asking how any of this helps answer her terribly practical questions. Let's address them in turn.

## THE PROCESS OF EXTENDING THE BI ENVIRONMENT TO SOFT INFORMATION IS SIMPLY ANOTHER STEP ON THE BI ROAD MAP.

1. In populating the warehouse, the BI group's prime responsibility is to provide the processes and technology needed to ensure the cleanliness, consistency, and reliability of information for users. That responsibility extends to all "decision support" information, irrespective of its source.

Collection and storage is most certainly the responsibility of the BI group. Analysis is a different matter. Best practice for traditional BI generally places responsibility for analytics in the business. Addressing different information does not change that practice. Analysis and reporting on this new information should be done by business users in departments such as marketing, with technical support from the BI group.

This initiative should be seen in all respects as an extension of existing BI efforts. As such, executive responsibility lies with the business. The process of extending the BI environment to soft information is simply another step on the BI road map. Teen Cosmetics' current organizational structure for managing and prioritizing that road map can be extended to include the groups using this information.

2. Nanci's conversation with business management has already begun, at least with respect to marketing's interest in sentiment analysis and social media. Nanci's role here is twofold: (1) to lead the conversation about issues such as information quality and consistency, ensuring that the costs and benefits of using external information are properly quantified, and (2) to provide guidance on new techniques and technologies emerging in the market that could provide specific competitive advantage to the business.

As BI director, Nanci has a central role in determining how this information should be organized, which leads me to her last three questions, which I'll take together.

3. Despite much innovation and early adoption, the use of soft information in BI still lacks an overarching structure and framework to reliably guide prospective implementers. There exist two broad schools of thought. One, coming mainly from BI tool vendors, proposes that soft information should be closely integrated into existing relational data warehouses and accessed through traditional BI tools. The second, largely advocated by content tool vendors, suggests that Google-like search techniques extended to traditional BI data will solve business users' needs.

In my view, both approaches are simplistic and stem from a lack of understanding of the needs of modern business to use all of the information available to it, regardless of source or structure, in a consistent, comprehensive, and coherent manner. I've dealt with this topic in a recent white paper ("Beyond the Data Warehouse: A Unified Information Store for Data and Content," May 2010, [http://www.9sight.com/uis\\_white\\_paper.pdf](http://www.9sight.com/uis_white_paper.pdf)).

The central point for Teen Cosmetics is that soft information must be linked to the data in the customer-centric warehouse. However, given the different technologies involved in the management and use of soft information, as well as the volumes of information anticipated, mass loading of soft information into the existing data warehouse should be avoided. Rather, indexes and other metadata generated when the external information is ingested into content management and search tools must be linked to the warehouse to enable analysis that spans both hard and soft information.

From a technological viewpoint, the major difference in working with soft information is that its meaning is embedded in the content itself and, due to its varied sources, that meaning may be vague and unreliable. Hard data meaning, on the other hand, is (or should be) clearly defined in separate metadata. Working with soft data therefore requires a “fuzzier” logic and produces more probabilistic results.

The most important point to remember is the one I started with: It’s all information!

### Harriet Fryman

“It is a very sad thing that nowadays there is so little useless information,” said Oscar Wilde back in the 19th century. Never were truer words spoken, as organizations such as Teen Cosmetics start to realize there is huge untapped potential in all those internal e-mail messages, documents, call-center notes, and external Twitter, Facebook, YouTube, news, and blog sites. Rather than dismissing the oceans of unstructured information growing with such velocity and volume as “useless,” organizations are starting to realize valuable analytics and actionable insights are there to be had.

Teen Cosmetics needs to tap into this wealth of useful information to answer questions such as, “What are teens hearing about our brand? Is negative chatter about our new product hurting our reputation? What consumer issues are our call centers dealing with? What attributes of our competitors’ products are teenagers raving about?” Such questions simply can’t be answered using the organization’s structured information.

The good news is that Nanci knows BI is all about enabling Teen Cosmetics to answer critical business questions for more effective decision making. She and her team have worked hard to deliver meaningful, structured information in such a way that it is easily accessed, understood, analyzed, and put to use by the business community. Now she is right to see her mandate expand

to incorporate the larger, more diverse, and more complex analytics world of unstructured data.

Better yet, she sees the opportunity in the consumer-facing departments of marketing, call center, and sales, where greater insight and consumer analytics will directly impact the top line. With access to unstructured data, marketing professionals and call-center teams can be more precise, agile, and responsive to consumer demands, enhance loyalty for the company’s products, and sustain Teen Cosmetics’ competitive advantage.

For Nanci, it is not a matter of “Is unstructured data something to tackle?” but “How do we get started?” The very best news is that the expertise she has built up in data sourcing, data warehousing, reporting, and analysis stand her in very good stead. Yes, unstructured data has different characteristics, much larger data volumes, and requires different analytics technology to answer those business questions, but the BI team’s skills in knowing what it takes to deliver information-centric analytic solutions remain invaluable.

First, Nanci should consider how search indexes can take their rightful place in her list of data sources for analytics. Everyone is familiar with Google search, so it would be easy for business users to picture the world of unstructured data made available to them through the simple navigation of keyword search. Because users are also familiar with drilling down through hierarchies of structured data, it will be easy for them to imagine refining the search results by selecting additional attributes.

Imagine Nanci has indexed her call-center records. A simple search of calls by time period shows that complaints increased during a four-week period in December. Further refinement identifies the complaints as centered in the northern United States and Canada and related to the new “Sleek Plum lipstick” product line. Links to the sales reports show a decline in sales in the following month and an increase in returns. If Nanci has also linked to Internet sites for sentiment analytics, her BI users will be able to identify chatter about Sleek Plum lipstick drying out and cracking in cold weather.

In this way, search indexes provide a powerful aid to analysis, and they take advantage of keywords from the text to provide the facets of navigation. Facets, unlike hierarchies, are not pre-defined or fixed, but are generated from the linguistic output of text analysis into attributes of the data. It works a little like the game of 20 Questions, where you use good questions to cut the number of potential answers in half each time.

Text analytics uses facets to narrow down answers, and when those facets are aligned with the dimensional world, together they provide seamless navigation between uncovering news in unstructured data and linking it to structured information for historical analysis and future prediction. With search indexes under the hood as a data source, and the new paradigm of keyword search and facet-led analytics, Nanci can marry traditional BI with unstructured data so users can gain insight into how the business is doing, why, and what actions they should take.

Next, Nanci should look at the scale of Internet data. The huge volume and heterogeneous nature of unstructured text, video, and audio information results in petabytes of information. The technologies of Apache Hadoop and MapReduce enable data-intensive distributed applications to work with thousands of nodes and petabytes of data.

How can this new technology be made available to a line-of-business user? One place to explore its potential is via an IBM emerging technologies initiative called Big Sheets. Think of it as a spreadsheet interface that enables you to pull in millions of rows of text and numeric data millions of columns wide, and then apply sentiment analytics to find insight in data sources such as Twitter. Nanci can see how this might work at [youtube.com/ibmetinfo](http://youtube.com/ibmetinfo).

Finally, let's look at where this type of analytics is taking us. This year we saw IBM's "Watson," a computer that can answer a complex, human-language question in less than three seconds, compete on *Jeopardy!* against the world's best human players. Watson's ground-breaking ability to understand the meaning and context of human language and rapidly process information to find precise answers to complex questions holds enormous potential to transform how computers help people accomplish tasks in business (as well as in their personal lives). For example, such technology could be applied at Teen Cosmetics to improve online self-service help desks or answer brand sentiment questions such as "What product of ours has the most positive buying sentiment in females age 13 to 18?"

To summarize, Nanci needs to expand her BI charter to include unstructured information, partnering with the business just as she does for other BI initiatives. Her team needs to grasp the new technologies of search indexing, blended file system and database storage, Hadoop and MapReduce, text and sentiment analytics tools, and new navigation and visualization approaches.

Nanci should stay laser-focused on a high-value, consumer-facing area of the business for her first foray. If she is asked to tackle unstructured data inside the enterprise first, I would recommend call-center analytics. However, given the choice, I strongly recommend that she run a

pilot with the marketing department to tackle social media sources and deliver brand sentiment analytics. This will deliver tangible insight with the biggest "wow" factor and greatest value for the organization, creating an immediate win.

### Michael Gonzales

Proceed with caution! Nanci's situation presents several issues. For instance, the BI group is considering taking on the responsibility of launching the initiative. No experience in sentiment analysis or dealing with unstructured data is described. No defined requirements are driving the adoption of sentiment analysis. Given all this, as opposed to directly answering Nanci's specific questions, I'd like to offer some general guidance from two different perspectives: (1) value versus risk and (2) data architecture.

#### Value versus Risk

The relatively low cost of storage is not a winning argument to sway executives about unstructured data. Risk mitigation and value creation offer more compelling reasons to adopt sentiment technology for two reasons: (1) sentiment analysis is expected to have only moderate benefit for organizations, and (2) as an emerging area, there is wide disparity between implementations, vendors, and solution maturity (Collins, 2010). This translates into risk.

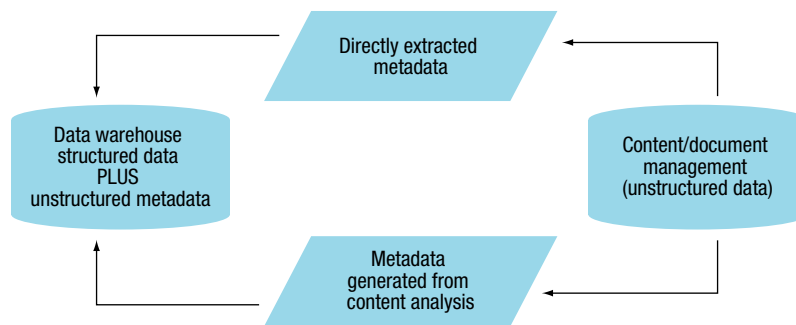
## NANCI SHOULD STAY LASER-FOCUSED ON A HIGH-VALUE, CONSUMER-FACING AREA OF THE BUSINESS FOR HER FIRST FORAY.

The risk is not limited to the effort required to glean value. Emerging areas are often characterized by a lack of formal best practices, mature vendors, and skilled resource pools (Collins, 2010). The BI group is probably not equipped to be responsible for the overall initiative; they likely lack skill or experience in the processes, architectures, or technology necessary for sentiment analysis, including areas such as natural language processing or content management.

Given Nanci's situation, I would recommend the following steps to increase value and reduce risk:

- **Define BI roles.** The BI group should resign itself to supporting specific areas of the sentiment analysis initiative, such as integrating metadata from unstructured data sources into the data warehouse and/or offering a blended presentation layer of structured and unstructured data for analysis.





**Figure 1.** Integrating structured and unstructured data.

- **Define success.** Decide what success and failure look like before you launch the initiative or specific projects.
- **Small scope.** Keep the scope of the initial efforts small and focused on specific applications.
- **Fail fast.** Identifying failure quickly affords an opportunity to determine cause and effect—not to mention save limited resources (McGrath, 2011).
- **Fail cheaply.** Design your initiative such that the impact of failure is relatively small. This can be achieved by clearly specifying your project effort and binding those efforts to predefined, expected, measurable value (McGrath, 2011).

### Data Architecture

From a BI/DW architecture standpoint, a core question to address is whether unstructured data should be modeled and stored within the data warehouse itself. Like all technologies and techniques in emerging markets, instead of formal best practices, there are active debates and differences of opinion. It is not surprising that some argue unstructured data should be integrated into the data warehouse. Naturally, this perspective is often espoused by BI and data warehousing practitioners (including Bill Inmon). However, those with a deep background in content management argue that unstructured data should be kept under a content management system and related database as opposed to being modeled within a monolithic data warehouse.

The information systems community has confirmed that information silos of structured data are simply not productive for an organization. However, with the exception of regulatory compliance, experience with unstructured data seems to suggest the opposite. Attempting to consolidate

multiple, unstructured data sources in order to address multiple, domain-specific applications interferes with functionality for two reasons: (1) there is a wide variety of unstructured data types, and (2) the subject matter of unstructured data is more diverse than what we typically find in structured data (Kuechler, 2007). As a result, there often is no reasonable value for consolidation of unstructured data sources. Some researchers have even suggested that it is naïve to attempt consolidation of unstructured sources into a monolithic environment (Rosenblatt, 2003).

Nevertheless, the benefits of integrating structured and unstructured data and presenting that blended content to end users are well documented. Although a lofty goal, it remains true that an enterprise information strategy should provide a comprehensive approach for handling structured and unstructured data as opposed to leaving it as an afterthought or merely cobbling together an improvised solution.

Several techniques are evolving as potential solutions. Enterprise content integration leverages middleware that connects centralized metadata in a federated approach to the integration of unstructured data. This is a classic approach where structured and unstructured data reside in respective environments that rely on the presentation layer to bind structured and unstructured data into a cohesive view for the user.

A more contemporary solution recommends the integration of metadata from unstructured data sources into the data warehouse (Figure 1). Once metadata is extracted from content, it is handled within the data warehouse like any other structured data, creating a data store accessible by all typical BI/DW tools and analytic technology, such as OLAP.

## References

Collins, Kimberly [2010]. "Hype Cycle for CRM Marketing Applications," Gartner Research. [http://www.gartner.com/DisplayDocument?doc\\_cd=205060](http://www.gartner.com/DisplayDocument?doc_cd=205060)

Kuechler, William L. [2007]. "Business Applications of Unstructured Text," Communications of the ACM. <http://portal.acm.org/citation.cfm?id=1290967>

McGrath, Rita G. [2011]. "Failing by Design," *Harvard Business Review*, April. <http://hbr.org/2011/04/failing-by-design/ar/1>

Rosenblatt, B. [2003]. "Enterprise Content Integration: A Progress Report," *The Seybold Report*.

## David Schrader

Let's look at Nanci's questions one at a time.

### 1. Should the collection, storage, and analysis of unstructured data be the responsibility of the BI group, or is it better left to other departments such as marketing? What factors should affect where the initiative is organizationally housed?

In most organizations with data warehouses, the collection and optimal storage of data (of every type) is usually placed in the IT/BI group. The use/analysis is a shared activity between sophisticated BI experts and domain-specific gurus—in this case, marketing and quite possibly customer service and merchandising. In general, data should be a shared corporate asset that is used multiple times. The BI/IT group should be a service group providing help to all other groups that need improved insights to make better, faster decisions based on any type of data. Perhaps a BI competency center could be of help for Nanci.

Marketing *may* have the skill sets to acquire the data and perhaps analyze it on their own, but when it comes to data quality, cross-subject data integration, sophisticated analytical tools, model building, or deep analytics, departments usually need the help of the BI/IT experts.

Of course, this is not always true. Sometimes the marketing group has deeper analytical skills than does the IT organization. Placement of responsibility can be influenced by where the relative sources of expertise sit in the organization—as well as sources of financial sponsorship and innovation.

### 2. How should she start the conversation with management about how to organize and best use unstructured data?

Nanci can assume management will not care much about the latest shiny technical thing. Social media analytics and unstructured data fit in this category. She needs some compelling business cases with plausible ROI that management can understand. In the fashion industry, product cycles are fast. Quick insights into "hot" and "cold" products can drive a variety of concrete business decisions about pricing, replenishments, and manufacturing/logistics supply chains.

Consumer allegiance can be fickle. Nanci should make the case for unstructured data using vocabulary and examples that management can easily understand. First, she needs a project game plan showing ROI; then, she needs to come through with actuals over time to build her credibility in any new areas.

Pain is a great motivator for action. If the company has excessive returns, a spike in the number of calls to customer care, and numerous complaint e-mails, then it probably also has negative word of mouth. All of these are predictors of negative business performance, and the clues are often based on unstructured data—phone calls, negative e-mails, bad comments in blogs or tweets ... things that should very much be on the minds of executives.

These business pain points can motivate executives to fund efforts to increase "voice of the customer" initiatives—sometimes quite literally in the case of voice recording complaint calls at the care center. As a way of securing funding for a project addressing unstructured data, Nanci might try proposing a Customer Focus Team that regularly looks at the customer inputs and starts automating the scoring of customer interactions.

Finally, watching what the competition is doing and sharing the highlights with executives can be quite compelling and another way to educate management, spark interest, and drive investments.

### 3. As Nanci tries to learn more about the use of unstructured data, what are some resources she should consider?

Many database organizations and vendors provide educational seminars, online Webinars, and white papers, and will even come to your location with demos to educate people at your company about what's possible.

For a good example of the kind of resources available, take a look at "The Case of the Retail Tweeters" on YouTube (keywords Teradata and BSI). This seven-minute story

shows the advantages of collecting and analyzing tweets, blogs, and feedback in the context of customer management at Brizio Fashions in terms that management can easily understand. It also shows mockups of ROI that can probably be adapted for Nanci's use.

#### **4. What are some of the technologies that she should be familiar with that are used with unstructured data?**

Sentiment analytics from companies such as Attensity can be used to move from "words" to "sentiment scores" that can be spliced together with other, traditional data elements (such as sales) to develop stronger attrition models.

Capturing and using voice inputs from automated survey systems and contact centers is a hot area right now. Linking voice recordings to customer records is relatively easy. Beyond that, emotion analytics are a leading-edge use of voice data to measure how emphatically customers make points during conversations with customer care agents.

#### **5. What are some of the major differences between working with structured and unstructured data that Nanci is likely to find?**

Structured data is much easier to analyze because there are good standards for data quality and better precision for analysis. For example, there is a range of "correct" values for point-of-sale transactions such as the total of a market basket. If you want to calculate a store's total sales for the day, there is really only one way to do so. The answer is either right or wrong.

Unstructured data is a mixed bag, depending on the data type. For voice, you can capture the bits and replay the recording, certainly, but how do you score it? Do you run voice-to-text algorithms, then examine the kinds of words used that have high emotion-laden content? Do you "listen" to the voice and rate increases/decreases in volume or pitch to extract the emotion?

The hottest area of all is social media network analytics, used to study viral marketing effects. Googling this term will result in hits on numerous vendors and open software sources of information. (A list of some of these is in the "How We Did It" PowerPoint deck that accompanies "The Case of the Retail Tweeters" case mentioned earlier.)

A final word of caution: It's tempting to perform viral marketing analytics in isolation, but it's far better to integrate the results back into the customer database and use insights such as viral influencer scores as part of an overall customer relationship management program. Although the technology is progressing to convert unstructured

objects into structured numbers that are easier for computations, it's still an art and rather imprecise. When you run a query on structured data, there is usually little ambiguity about how to write the query and interpret the result. It's not so easy with unstructured data. You can count the number of customer calls coming in, but what's the right answer if you want to query on "number of unhappy customers"? Treat the unstructured data capability as a work in progress. Experiment and build a fast-learn/fast-fail environment with the marketing organization.

#### **David Stodder**

Nanci can no doubt relate to the ancient Chinese curse, "May you live in interesting times." The BI paradigm is changing. The expanse of online data that customers generate on social networking and media sites and through e-mail, customer surveys, product registration, and warranty forms is an enormous temptation—and can be of enormous value—for marketing functions. Traditionally, this unstructured data, which most experts feel represents about 80 percent of the data available to organizations, has been difficult to access and analyze in a timely fashion. Frontline operations such as contact centers have had a hard time hearing "the voice of the customer" in anything close to the time interval within which organizations could most benefit by acting upon it. Organizations have not had the technology to effectively tap this kind of data.

Now, however, more of this data is recorded in digital form, and technologies such as text analytics and sentiment analysis are evolving to make it easier to crack these information treasures. What makes this interesting for BI is that as this content-oriented technology has been maturing, many organizations have been busy expanding BI out to operational users in marketing, sales, and service. BI dashboards have given such users previously unimaginable access to data and performance metrics. If these dashboards could now provide access to analysis, visualization, and alerts about social media trends or information derived from unstructured data sources, frontline users could have a significant competitive edge.

#### **Structure: BI's Hidden Gift**

Amid the rush of interest in social media and unstructured "big" data, it is easy for BI directors such as Nanci to fear that they are becoming marginalized because they only work with that limited "20 percent" of an organization's data sources. Putting aside the argument that this 20 percent share still contains the most important sources for decision making about profitability, budgets, and revenues, this view overlooks the hard-won structural gains of BI and data warehousing, and how these technology implementations have allowed users to be much more productive.

Organizations must not take their eyes off the ball by neglecting to continue to invest in advanced BI and analytics for structured data. At the same time, they should look at how the structural gains made with BI and data warehousing might be extended to the “Wild West” of social media data and unstructured content so that users can be as productive with these sources as they are with BI.

In most organizations, users continue to be bogged down by unstructured data that they must personally organize, manage, and share, or that they can access only through proprietary content management systems. To tap unstructured sources, BI directors and data warehouse managers definitely need to learn about taxonomies, “folksonomies,” and other concepts used by search engines and content management systems to find relevant terms and data. However, directors such as Nanci should also consider how their rich experience in providing access and analysis of heterogeneous sources to many different types of users could be valuable for unstructured data—and for integrating BI with these sources.

### Starting the Conversation

As with BI implementations, the discussion should begin with an effort to understand user requirements. What are Teen Cosmetics’ objectives in working with unstructured data? How are these different from the objectives with traditional BI? Is there a relationship between current BI reports and analysis and the knowledge sought from unstructured sources? Could structures used for BI and data warehousing to deliver timely information be applied to goals using unstructured sources?

The answers derived from this type of discussion will help organizations such as Nanci’s focus on what will enable users to be productive with the new data sources faster. Stressing what is known from their use of BI tools and data warehouses could also help them avoid the unmanaged creation of huge volumes of social media data and other unstructured content that are accessible only through proprietary methods and hand-coded programs. BI and data warehousing professionals know only too well that this practice usually leads to silos that become costly obstacles to future user access and analysis.

### Unified Information Access: A Helpful Technology Direction

Many organizations have experience with tools and techniques for enterprise search and content management, including the development of taxonomies. As mentioned, Nanci should seek out colleagues who have experience in these areas. In recent years, technology providers have made progress toward what some industry analysts call “unified information access,” or the integration of BI with search and discovery technology. This is an important area

for Nanci to evaluate; most major BI vendors, for example, have integrated search tools into their suites to compete with specialized unified information access vendors.

It would be wise for Nanci to examine these tools jointly with those in the organization responsible for enterprise search and taxonomies. If social media data is the key concern, she should make sure to get good answers from the vendors about text analytics and other capabilities for accessing and analyzing of this type of data.

Text analytics (or “text mining”) is the primary technology used for sentiment analysis, which gauges the intent and emotions behind the words used in social media and networks. An important sentiment analysis concept is “polarity,” where the writer’s expressions are ranked on a spectrum of positive to negative. Along with building predictive models, text analytics specialists will frequently apply statistical and linguistic techniques to better understand the emphasis behind and relationships between words and develop rules for when certain expressions merit action from users or automated processes.

## TEXT ANALYTICS TOOLS AND TECHNIQUES YIELD RESULTS THAT ARE BY NATURE INEXACT AND OPEN TO INTERPRETATION.

Nanci should learn more about these tools and techniques. She should keep in mind, however, that text analytics tools and techniques yield results that are by nature inexact and open to interpretation. In addition, sentiment analysis is not yet mature, which means that organizations should experiment with it rather than accept the results as gospel.

This last point suggests probably the biggest difference between BI and unstructured worlds that Nanci should keep in mind: that is, the difference between BI’s precision based on numbers and the inexact nature of unstructured data search and discovery. Social media and networks in particular are constantly in flux, with “influencers” riding high for a while as figures with the power to shape potential customers’ sentiments, opinions, and decisions, only to be replaced over time by other influencers. Social media data analysis is critical for organizations to manage their brands, shift marketing strategies, and generally participate in the ongoing discussion inside the universe of their products and services. However, the subjects of this analysis are constantly moving and changing, which means that the results should be taken with a grain of salt.



This is why I would strongly advocate looking at how BI and unstructured analysis can complement each other. Unified information access technologies are maturing to support this integration. By putting the two together, users will find higher value and productivity as they travel into new worlds of data and analysis.

---

*Barry Devlin is founder and principal for 9sight Consulting. Contact him at [barry@9sight.com](mailto:barry@9sight.com).*

*Harriet Fryman leads market strategy for BI software at IBM. Contact her at [hfryman@us.ibm.com](mailto:hfryman@us.ibm.com).*

*Michael Gonzales, Ph.D., is managing partner for dss42, LLC. Contact him at [mlg@dss42.com](mailto:mlg@dss42.com).*

*David Schrader is director of marketing for Teradata. Contact him at [david.schrader@teradata.com](mailto:david.schrader@teradata.com).*

*David Stodder is TDWI research director for business intelligence. Contact him at [dstodder@tdwi.org](mailto:dstodder@tdwi.org).*

This article appeared in Volume 16, Number 2 of the *Business Intelligence Journal*, an exclusive TDWI Premium Member benefit. For more information about TDWI Premium Membership, visit [tdwi.org/premiummembership](http://tdwi.org/premiummembership).



Analytics Everywhere™

# Vertica delivers actionable analytic insights from ALL your data. Now.

**Real-time loading and querying - 50 to 1000x faster**

**Scale-out MPP architecture - linear and limitless**

**Runs on standard hardware, virtual machines, or in the cloud**



## Big Data Made Easy.

Simply stated, Vertica is the leading provider of next-generation analytics platforms enabling companies to monetize all of their data. With over 500 customers, Vertica powers some of the largest organizations and most innovative business models in the world. Unlock the value hidden within your data with Vertica.



### Free 30-Day Trial

Test drive the Vertica Analytics Platform.

[www.vertica.com/evaluate](http://www.vertica.com/evaluate)



## How the iPad Experience Will Change the Face of BI

BY STEPHEN SWOYER

iPads. If they aren't yet ubiquitous, it sure *seems* as if they're everywhere: planes, trains, escalators, the executive suite.

Increasingly, the iPad is being used as a complement to—and in some cases, as a replacement for—conventional desktop or laptop computers.

It's a trend that promises to wholly transform the way information is presented and consumed in the enterprise. To their collective credit, many business intelligence (BI) players get this, in part thanks to the foresight—or stubbornness—of their executive leaders.

Leading BI executives have embraced the iPad, and executives—unlike rank-and-file workers—are the kind of users who can compel an inflexible IT department to accommodate and support their needs.

Kalido CEO Bill Hewitt uses an iPad. So does Michael Corcoran, vice president of marketing for Information Builders (IBI). Other BI executives tote both a laptop *and* an iPad. Most would probably prefer to tote only the latter.

Hewitt, for one, didn't initially intend to go laptop-less. Owing to a delivery snafu, he says, he canceled an order

for a new laptop and decided—in a fit of what might be described as exasperation—to try to tough it out on his iPad. "Tough," however, doesn't quite describe Hewitt's experience.

"I love it. I use it for everything. It's so much more convenient. I carry it with me all day, wherever I go, whatever I'm doing," he enthuses.

What did Hewitt's IT department think about his decision?

"I told them, 'This is my new computer. I'm going to use it for everything,' and that's what I've been doing," he says, smiling benignly.

Although Kalido's software isn't overly dependent on a pretty user interface (UI), a next-gen device like the iPad *can* show its business information modeler (BIM) to excellent effect, Hewitt observes. A user experience like that of the iPad, he suggests, can "enhance the way that's [i.e., BIM] used. If you think about the way [BIM is] designed, it promotes collaboration between IT and business. Just imagine [IT and the business] collaborating with an iPad, using [a] touchscreen instead of [a] mouse. It's a completely different experience."

IBI's Corcoran likewise uses his iPad exclusively. He loves it, says Jake Freivald, a product marketing manager for IBI who works with Corcoran. "He uses it for everything. He came in [with his iPad] and said, 'I got this, and I'm going to use it.' That's basically what [he's been able to do]."

Freivald, for his part, is bullish on the iPad—as well as on other mobile devices. In many ways, he argues, mobile or nontraditional computing devices are the perfect tools—and offer the perfect context—for the kind of lightweight, pervasive analysis that he, Corcoran, and other IBIers have been promoting for the last 18 months.

"Mobile BI tends to be a little less analytical because in many cases it's being pushed down to these [nontraditional user] constituencies," he says. "The people running those kinds of [mobile] apps are typically not people who are doing analysis. Part of a CMO's job is to ask lots and lots of questions. A guy who drives a truck and is worried about stocking shelves is only going to be asking questions that are specific to him," Freivald points out.

At the same time, he stresses, the iPad isn't the end-all and be-all of mobile devices.

"Last year [i.e., 2010] was the year of the iPad as an interactive client device. We would hear from our customers, 'We want a native app just for the iPad. We're only going to be iPad. But that tune has changed a bit over the last four or five months. What we're also seeing is that people are bringing their own devices into work, and while many these are iPads, some of them aren't,'" he explains.

When it comes to mobile, IBI, like many software vendors, has focused on delivering rich Web applications (based on HTML5) instead of native- or platform-specific applications. This jibes with a recent study from software development researcher Evans Data, which found that a majority of mobile developers are focusing on Web apps instead of native apps—in spite of the dominance of the iPad, or (in the small-form-factor mobile arena) the iOS or Android platforms.

### **Risky Business?**

While mobile is undoubtedly here to stay, it poses unique risks of its own, such as "juice-jacking."

At the recent DefCon conference in Las Vegas, several hundred attendees—the bulk of them, presumably, security-conscious folk who should've known better—plugged their cell phones into a phony charging kiosk.

This exploit (which experts dub "juice-jacking") was sponsored by information security specialist Aires Security and was intended as a cautionary proof-of-concept

demonstration. Attendees who plugged in received the following message: "You should not trust public kiosks with your smartphone. Information can be retrieved or downloaded without your consent. Luckily for you, this station has taken the ethical route and your data is safe. Enjoy the free charge!"

It's one of many reasons companies will have to think long and hard about how much they're going to expose—and how they're going to expose it, says IBI's Freivald. "It really will depend on the company. Maybe there are cases where you don't want persistent data, or maybe you want to make it so that you can only go back to the server live—[in other words] you can't save it on the device—these are things that the companies themselves are going to have to determine," he comments.

"The way that we handle [security] in [a] mobile [context] isn't radically different from how you'd protect a [desktop or laptop] computer—things like passwords and expiration dates. If you're concerned about giving away information accidentally or maliciously, you have to worry about someone taking screen captures [of the device]. There are so many different [vectors]."

On the other hand, maybe it isn't a question of having the *most mobile device*—like an iPad or iPhone—but (conversely) of having a *more mobile laptop*.

After all, even executives who don't tote an iPad seem to have a weakness for an iPad-like form factor. WhereScape CEO Michael Whitehead, for example, sports the latest, lightest, and leanest version of Apple Inc.'s MacBook Air. Whitehead used to use a netbook-form-factor PC laptop that—withstanding the compactness of its length and width—was almost an inch thick. His new iPad-sized MacBook Air? Just over a quarter inch at its thickest point.

---

*Stephen Swoyer is a technology writer based in Nashville, TN. Contact him at [stephen.swoyer@spinkle.net](mailto:stephen.swoyer@spinkle.net).*

This article appeared in *BI This Week* e-newsletter October 18, 2011. For more information or to subscribe, visit [tdwi.org/pages/publications/newsletters](http://tdwi.org/pages/publications/newsletters).





## 7 Tips for Working with Data Analysts

BY TED CUZZILLO

Data analysts tend to feel like odd ducks. Crucial as they can be to business success, many say that clients just don't know how to work with them for the best results—failing to explain context, for example, or micromanaging, or just weighing them down in any number of ways.

It takes two to analyze—and just a few simple things can make the difference between disappointment and golden eggs.

Though it's hard to generalize about this emerging group, I've noticed several trends. First, many data analysts have had little or no formal training for the job. They may be "good at math"; they may even know statistics inside and out; and they're all skilled with one or more tools. Many, though, don't know how to tell a story—with all its power to give context and set facts in memory.

Second, the data analyst's closest relative in business may be the journalist. The best analysts spin stories from their observations of data, and they ask a lot of questions—some of them inconvenient. They know a little about a lot, they follow trails from one clue or question to another, and they need elbow room.

"There's a distorted idea of what data analysts do," said advanced-analytics consultant Theresa Doyon, Ph.D. Her doctorate is in social psychology, and she trained later to use SAS and other tools. She's seen job listings that ask for high-end data scientists or a degree in computer science for database marketing analysis, which actually requires far lower qualifications than that.

True, analysts are a kind of quant. They're also a kind of artist, which is usually not the kind of beast that finds an easy nest in most businesses.

Here are seven tips to help you get the best results from your data analyst.

**Tip 1: To launch a project, include in the first meeting someone who knows the data and will consume the analysis. Sit down with the analyst, explain the context, and go into depth with the questions.**

What's your intent? How did you form the questions? The better the analyst understands your situation, the better he or she can provide actionable results.

"I want to work with the people who live and breathe this data, I want to talk to the audience," said Joe Mako, an analyst at Cincinnati-based S2.

**Tip 2: Ask honest questions and don't micromanage.**

"There's always someone who seems to think they know what you ought to be doing and how you ought to be doing it," said Doyon. One client, for example, asked her to segment his customer base and was unhappy when she found different segments than he'd decided on beforehand. "What was in [his] head didn't exist," she recalled.

Given enough depth and room to roam, a good analyst can find the data's story and spin it for insight and memorability. "People are so used to the idea that analysts are cold and analytical," said Doyon. They don't recognize one mark of superior analysts: they're storytellers. "Data is their paint."

**Tip 3: Explain the data's lineage.**

What's been done to the data? Is it raw, or has it been transformed in some way, and how? Explain business rules. One analyst, who asked for anonymity, recalls discovering an apparently deliberate misspelling of Connecticut in a large set of transactions. Why? In another project, one of a retail store's cash registers showed zero sales on every third week. Often only the client knows the significance. Such puzzles should be explained as much as possible at the start.

**Tip 4: Allow for failure.**

Analysts try one approach and then another—and sometimes they miss. If failure's not allowed, they play it safe and paths go unexplored. Traditional business culture has little appetite for "test and learn" and usually opts for "fail and get out!"

**Tip 5: Respect their time.**

Because analysts' work is often not well understood within organizations, many seem to assume it's free. Save everyone's time by learning basic terminology before you meet to hear results. Little else kills a meeting more than letting 1 of 11 people in the room stall for 10 minutes to understand a term such as "KPI."

**Tip 6: Change priorities judiciously.**

Analysis, like most projects, runs better when there's a long road map laid out at the start.

**Tip 7: Learn to perform data analysis yourself.**

You can learn how analysts think. You might even start performing some analysis of your own with the new generation of easy-to-use desktop tools.

Bonus Tip: If somehow you've started off on the wrong foot with a data analyst, there's always Ben and Jerry's ice cream—or beer.

---

*Ted Cuzzillo is a journalist and industry analyst focused on analysts' tools and needs as well as the environments in which they work. You can contact him at [analysts@datadoodle.com](mailto:analysts@datadoodle.com). If you're a data analyst, he'd appreciate your participation in [his survey](#); you'll receive a free preview of his report when it's complete.*

This article appeared in *BI This Week* e-newsletter May 4, 2011. For more information or to subscribe, visit [tdwi.org/pages/publications/newsletters](http://tdwi.org/pages/publications/newsletters).



## How Agile Requirements-Gathering Can Conquer Analysis Paralysis

BY MICHAEL A. SCHIFF

We live in an imperfect world, yet some people still believe that if something can't be done perfectly, it shouldn't be attempted at all. Although this may be an extreme view and may be totally appropriate in some situations (such as building a nuclear power plant or performing surgery), it should not be a showstopper for data warehouse implementations.

How many times have you encountered situations where a feasibility study of a major system (operational or analytic) continually missed deadlines because of the supposed need for additional research and analysis? Furthermore, when the analysis was completed (or management insisted that the project go ahead without additional delay), the implementation date almost always passed the initial estimate.

This phenomenon is sometimes referred to as "analysis paralysis," and I have personally observed it all too many times. I still remember an incident in which a marketing director, when estimating product demand, delayed an entire team's associated efforts as he continually recomputed his numbers. We finally realized that he was attempting to estimate percentage demand across channels down to two decimal places even though he was instructed that management wanted "ballpark" estimates within 5



percent. He (wrongfully) thought that his overly cautious attitude would reduce the possibility of an error and was oblivious to the fact that by the time his analysis was completed (if ever) and decisions made, any associated benefits might be long gone.

Data warehouse practitioners need to remember that before the term *business intelligence* was coined, organizations built what was known as a decision support system (DSS) to facilitate and support better decisions. Because a primary goal of a data warehouse is to support business intelligence and make information available to those who need it when (and where) they need it, we cannot get bogged down by analysis paralysis. Instead, we must strive for agility and create an environment that allows our users to receive timely, relevant information.

We need to remember that there are costs associated with delayed decisions and that once an opportunity is lost, it may never be available again. We have all experienced situations where users' initial specifications didn't match actual user needs, even after formal functional specifications were approved. In most cases, the specifications are a starting point that usually leads to requests for additional details, additional data, and/or other report formats.

Analysis is frequently an iterative process and it is safe to assume that most answers will lead to additional questions; questions that were not thought of until the prior questions were answered. "What about..." may begin the most frequent user comments we hear after users are shown initial results.

One approach to facilitating an agile data warehousing and business intelligence environment that deserves special attention is prototyping. It can help us ensure that our reports and analyses are actually what our users need—which may or may not be what they initially thought they wanted!

Although sometimes the ability to prototype is constrained because the data we need does not already reside in our data warehouse, we can use techniques such as data virtualization or enterprise information integration (EII) to access this data from operational systems. Although the data in the operational systems may not be of the same quality as data formally loaded into a data warehouse, the approach can provide a quick-and-(sometimes very)-dirty solution. If satisfactory, the data can ultimately be formally cleansed and included in the data warehouse.

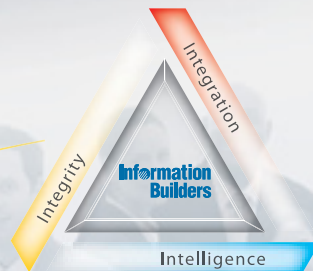
Even if our initial analysis provided all the answers our users needed, their requirements and associated queries will evolve over time. One of the best ways to deal with this is to design our data warehousing and business intelligence applications with agility as a major objective and not get bogged down by analysis paralysis in an attempt to perfectly solve initial requests that may (or may not) reflect what our users really need.

---

*Michael A. Schiff is the founder and a principal consultant at MAS Strategies. He can be reached at [mschiff@mas-strategies.com](mailto:mschiff@mas-strategies.com).*

This article appeared in TDWI Experts e-newsletter July 21, 2011. For more information or to subscribe, visit [tdwi.org/pages/publications/newsletters](http://tdwi.org/pages/publications/newsletters).





## Everyone Makes Decisions.

Do all of your employees, customers, and partners have the information and analytics needed to make your business succeed?

With our WebFOCUS and iWay Software products, your organization can:

- Deliver an information-rich customer experience
- Drive greater performance from employees and partners
- Create a single view of your customers and products
- Ensure information integrity and accuracy
- Identify fraud, waste, and abuse

Information Builders ranks first in customer satisfaction because we help our clients drive more value to more users, more rapidly, and at lower cost than any other vendor. Find out the five things you need to know about your users before deploying BI in our free white paper:

[informationbuilders.com/5things](http://informationbuilders.com/5things)



## Best Practices Awards 2011

---

TDWI's Best Practices Awards recognize organizations for developing and implementing world-class business intelligence and data warehousing solutions. Here are summaries of the winning solutions for 2011.

For more information, visit [tdwi.org/bpawards](http://tdwi.org/bpawards).



### TDWI thanks this year's panel of expert judges:

**Barb Wixom**, Associate Professor, University of Virginia

**Chris Adamson**, DW Specialist, Oakton Software LLC

**Colin White**, President and Founder, BI Research

**Daniel R. Evans, CBIP**, Sr. Group Manager/Solution Lead, Avanade, Inc

**David Stodder**, Director, TDWI Research

**Evan Levy**, Partner, Baseline Consulting

**Hugh Watson**, Professor of MIS, University of Georgia

**Jill Dyché**, CBIP, Partner, Baseline Consulting

**John Thompson**, CEO, Marketing Sciences, LLC

**Jonathan G. Geiger**, CBIP, Executive Vice President, Intelligent Solutions, Inc.

**Laura Reeves**, Principal, StarSoft Solutions

**Mark Madsen**, President, Third Nature

**Mark Peco**, CBIP, Partner, InQvis

**Nancy Williams**, CBIP, Vice President, DecisionPath Consulting

**Patty Haines**, President, Chimney Rock Information Solutions

**Philip Russom**, Director, TDWI Research

**Ralph Hughes**, Chief Systems Architect, Ceregenics, Inc.

**Sid Adelman**, Principal, Sid Adelman & Associates

**Steve Dine**, President, Datasource Consulting, LLC

**Steve Williams**, President, DecisionPath Consulting

## ► Advanced Analytics Con-way Freight

Con-way Freight is a leading less-than-truckload (LTL) freight transportation company that provides guaranteed, day-definite regional and transcontinental service through a single, unified network of more than 300 service centers in the U.S., Canada, Mexico, and Puerto Rico. Con-way Freight is a subsidiary of Con-way Inc., a \$5 billion diversified freight transportation and logistics services company.

Customer retention is an important element of Con-way Freight's strategy. In 2009, a company study showed midsize customers defected at a double-digit rate; this churn accounted for several million dollars in lost profits per quarter. Con-way's solution, the Defector Project, uses predictive analysis to identify customers likely to defect in the next three months. A report is published to Salesforce.com, where automatic notifications are received by account executives who implement corrective business processes to retain those customers.

The project leveraged Con-way's BI platform to understand why customers were leaving. It uses more than 20 metrics to identify profitable customers likely to defect in the near future. The sales and operations teams collaborated to build a customer-retention process, which starts with automated notifications in Salesforce.com and includes steps to proactively engage the at-risk customers, address their concerns, and retain their business.

Since May 2010, Con-way Freight has seen a 30 percent decrease in customer defections, saved nearly 120,000 shipments to date, and saved tens of millions of dollars. The solution was low cost, highly automated, and flexible enough to adapt to future problems.

## ► BI on a Limited Budget Alcatel-Lucent

Alcatel-Lucent is a trusted partner of service providers, enterprises, and governments worldwide, providing solutions to deliver voice, data, and video communication services to end users. A leader in broadband networking and IP technologies, applications, and services, Alcatel-Lucent leverages the technical and scientific expertise of Bell Labs, one of the largest innovation powerhouses in the communications industry.

As the result of a company-wide reorganization in 2009, Applications Group's BI platform had to be updated to support the business on a larger scale. The previous BI

platform focused mainly on project management; the new platform had to offer a 360-degree view of the business. Alcatel-Lucent also wanted a user-centered, collaborative solution rather than a traditional BI platform in which users passively view reports and dashboards.

The organization's BI solution, R&D 360, combines the latest open source technologies, Web 2.0 technologies, and BI best practices to monitor and control research and development activities. Leveraging BI, data warehousing, and data management technologies, R&D 360 provides up-to-date information and helps business users access and visualize near-real-time aggregated data buried in multiple operational systems—all in a single location.

R&D 360 supports Alcatel-Lucent's processes and deployment of best practices. The solution helps solve business issues, provides monitoring gauges to the organization, and allows management to be proactive and make appropriate decisions based on consolidated and accurate data. It has helped the company save \$500,000.

## ► Emerging Technologies and Methods GUESS? Inc. (co-winner)

GUESS?, Inc. is a well-known apparel company with more than 1,421 retail stores worldwide. A company known for its innovative style, GUESS continues to dress and accessorize the world with fashion-forward apparel, handbags, watches, shoes, and more.

Using emerging technologies in current mobile BI, GUESS?, Inc. created an innovative, next-generation BI iPad app that pushes actionable and mobile information to traditional and nontraditional BI users, allowing GUESS BI users to make better decisions. The content equivalent of 12 dashboards was turned into one multifaceted app to help merchants and executives be more productive during store visits, at corporate, or at home.

The app's rich data and easy workflow allows it to serve multiple user communities, including executives, merchants, and designers. Users can tap a given style on their screens to open new screens with additional data, empowering decision makers with business-critical information whether they are traditional or nontraditional BI users. Because it is designed to encourage users to be inquisitive about the data, the app is transforming the BI culture of the company.

This latest cultural shift is revolutionizing how BI is utilized at GUESS by allowing users to explore business data without becoming technically overwhelmed as they

navigate down and across to different levels of detail. The iPad app is easy to use, engaging, visually impactful, and business relevant with its embedded workflows.

### ➤ Emerging Technologies and Methods Nykredit (co-winner)

Nykredit is one of Denmark's leading financial services groups involved in mortgage lending, banking, insurance, pensions, and estate agency businesses.

Nykredit's project goal was to establish the technical framework to become an advanced credit institute under the Capital Requirements Directive (CRD) of Basel II. The company needed to develop risk assessment models that the Danish FSA would approve.

The Nykredit team developed a metadata framework that would define the behavior of the BI application and could be used actively in calculating all the measures and the generation of the actual reports for business users to view and approve before sending the finalized reports to the Danish FSA. Nykredit used the open source LAMP platform, with PHP as the programming language and an Oracle database to store the raw data, metadata, and resulting measures.

In November 2007, Nykredit obtained approval from the Danish FSA to apply the internal ratings-based approaches under Basel II. The framework's first challenge came when the Danish FSA changed the majority of the report layout definitions and some of the formulas within a month of the first parallel reporting deadline. The changes were handled by only one worker in less than a week by simply verifying and changing metadata, proving that the metadata-driven framework is strongly adaptive, reusable, and easy to maintain and expand in a modular approach. It provides a competitive advantage: faster time to market and lower manpower requirements, and thus a high return on investment.

### ➤ Enterprise BI Westfield Insurance

Westfield Insurance is a 160-year-old provider of commercial and personal insurance and surety services via a network of 1,200 independent insurance agencies.

Westfield began its business intelligence program in 2007. It sought to attain customer intimacy to increase profitable growth and retention; develop insight into insurance risks, claims, agency/producer distribution, and customer

satisfaction; drive product and service innovation using customer data; integrate and consolidate business and analytical intelligence processes across the enterprise; and drive common understanding and consistent usage of information.

The company adopted an iterative, sprint-based approach to report development. Reports were prototyped in joint application design (JAD) sessions with business analysts. Data governance was another key factor in delivering applications that meet the needs of different business units. The certification and publication of a BI data glossary resulted in a high degree of confidence in BI data quality and a high rate of adoption.

Westfield established an Analytics Resource Center as a center of excellence for business analysts. It plays a critical role in establishing enterprisewide analytical priorities and translating business questions into meaningful metrics and dimensions. A dedicated change manager addressed the risks associated with the required cultural change, and "change agents" from each business unit helped manage the process for their areas.

These four elements—collaborative and iterative design; enterprisewide data governance; analyst center of excellence; and change management discipline—combined to create a broadly supported, integrated effort that drove successful adoption of BI at Westfield.

### ➤ Enterprise Data Management Strategies Quicken Loans Inc. (co-winner)

Quicken Loans is the nation's largest online home mortgage lender and one of the 10 largest retail lenders in the country. The company is affiliated with sister companies in the title and real estate industries.

Quicken Loans' 7.5 TB data warehouse is focused on providing visibility to the constantly changing business. Vast amounts of operational data have made enterprise data management strategies necessary, including automation, optimizing visibility into data and processes, and database administration. The company's real-time BI goals are driven by fluctuating interest rates, increased compliance, and banker-licensing requirements.

Over the years, Quicken has developed a solid architecture that has streamlined the way it integrates data from any source. It has identified the exact steps the data must go through. In addition, the rules and data-cleansing steps are coded into the templates for data warehouse population. In a data warehouse environment of this size, managing



metadata is a necessity. Quicken has automated the collection of data regarding all database-related objects, including tables, views, columns, cube dimensions and measures, and stored procedures.

Quicken Loans' successful implementation of a real-time EDW and ETL/OLAP architecture exemplifies a good foundation, and its tools show consistent innovation on the platform. The company's bankers, directors, and marketing analysts understand and trust the information they view, which helps them make the best decisions and helps keep the company ahead of fluctuating capital markets and interest rates.

### ► Enterprise Data Management Strategies Thrivent Financial for Lutherans (co-winner)

Thrivent Financial for Lutherans is a faith-based, not-for-profit, *Fortune* 500 membership organization with nearly 2.5 million members and over \$73 billion in assets.

Thrivent Financial began to design and implement its enterprise information management strategy in 2007. The goal: to build a foundation of trusted, accessible, and documented data by leveraging existing data, complying with standards, and contributing to data sharing and reuse. The strategy covers data warehousing, business intelligence, data governance, master data management, data services, and enterprise content management.

The implementation of the strategy drastically changed the data approach from silo project-by-project development to a cohesive, enterprisewide approach. This enabled several key business initiatives, improved data quality and security, and yielded multimillion-dollar annual savings. Execution of the strategy supported business growth, reduced risk, and increased efficiency by providing common, unified, and trusted data to operational and analytical applications.

The subject area framework has gained widespread acceptance and is now used to provide a common framework and language across business and IT, establish clear boundaries of data ownership and stewardship, and identify common data needs across projects. It has also been employed by business users to communicate their data needs to IT.

Thrivent Financial created a unique approach that interconnects key data capabilities and is based on a common subject area framework adopted by business and IT. This approach is consistently implemented by all enterprise projects so that data solutions are aligned across the enterprise.

### ► Enterprise Data Warehousing WellPoint, Inc.

SOLUTION SPONSOR: COGNIZANT TECHNOLOGY SOLUTIONS

WellPoint is the largest health benefits company in the U.S., with approximately 33.5 million members in its affiliated health plans. WellPoint offers a broad portfolio of integrated healthcare plans and related services, and specialty products such as life and disability insurance benefits; pharmacy, dental, and vision plans; and long-term care insurance.

WellPoint wanted to establish a centralized, near-real-time enterprise data warehouse by integrating multiple sources of data into a single enterprisewide data model and promoting standardized platforms, tools, and definitions. The initiative EDWard (enterprise data warehouse reporting depot) provides integrated healthcare information with latency of less than 24 hours from over 30 operational entities. In addition, it provides a high-end analytical data platform (with consolidated views of its 33.5 million members) to internal business entities and external partners, driving business initiatives that lower care costs and improve care quality.

EDWard played the critical role of information backbone of several enterprisewide business initiatives at WellPoint, including standardizing health and wellness programs, an online tracking and self-serve incentives reporting solution, and reducing the cost of care programs.

After the implementation of EDWard, WellPoint is expected to retire about 20 information assets and realize significant cost savings by eliminating infrastructure and maintenance. The EDW also supports WellPoint's compliance management processes and helps build beneficial partnerships with healthcare entities. Business departments within WellPoint consider EDWard a key strategic asset and adopted it in many business processes and programs.

### ► Government and Non-Profit Internal Revenue Service

The Internal Revenue Service (IRS) is responsible for collecting taxes, administering our nation's tax laws, and providing service to taxpayers.

The IRS Compliance Data Warehouse (CDW) supports these efforts by providing data, tools, and computing services to research analysts in the IRS and Department of Treasury for statistical analysis, mathematical modeling,

simulation, prediction, and other research activities aimed at improving tax administration. As the biggest database in the IRS, the CDW has set the standard for large-scale data analytics in the world's largest tax agency. Brought online in 1997 with less than one terabyte (TB) of data and a handful of users, it has since grown into a nearly one petabyte (PB) system serving hundreds of analysts across multiple agencies.

The CDW manages a massive amount of data used to estimate the U.S. tax gap, develop models of taxpayer burden, simulate the effects of policy changes on taxpayer behavior, identify abusive tax schemes, and support other strategic initiatives. Over time, it has become the preferred system for both short- and long-term analytics, and is often the last resort when data does not exist in a usable format anywhere else.

The CDW has surpassed its goal of providing data to the IRS Research community. It now maintains over 15 years of data from more than 20 legacy sources and supports analysts in the Treasury Department, Joint Committee on Taxation (U.S. Congress), General Accountability Office, and the Treasury Inspector General for Tax Administration.

## ► Organizational Structures National Instruments Corporation

National Instruments, a leader in virtual instrumentation for 34 years, delivers LabVIEW graphical software and modular instrumentation hardware for product design, simulation, control, testing, and manufacturing. The Austin, Texas-based company has 5,200 employees worldwide, direct operations in over 40 countries; and \$873 million in 2010 revenue.

National Instruments' BI/DW program develops enhanced capabilities to meet the needs of global stakeholders, including increased self-service to integrated, actionable information that provides answers and insights to increasingly sophisticated business questions.

Since 2007, NI has adopted a new EDW architecture (with gradual retirement of its legacy architecture), implemented a new enterprise reporting platform, and adopted or improved critical processes.

NI employs a BI/DW virtual team model where three individual teams work in partnership. Each team has specific core competencies, career paths, and dedicated management. The three teams consist of BI program managers and analysts, DW business analysts, and DW programmer analysts. The virtual team organizational structure enables an agile, enterprise engagement

framework, which produces an effective tool to identify and accomplish the needs of the business.

This organizational structure offers uncommon benefit and scale, as each of the three teams can focus resources to evolve their own departmental initiatives, proficiencies, and vision. No single team is responsible for enterprisewide information management. Instead, the BI team specializes in crafting strategic and tactical solutions for the business, while the business and programmer analyst departments specialize in implementation.

## ► Performance Management Cleveland Clinic

SOLUTION SPONSOR: STRATEGIC TECHNOLOGY SOLUTIONS

Located in Cleveland, Ohio, the Cleveland Clinic is a not-for-profit, multi-specialty academic medical center that integrates clinical and hospital care with research and education. It employs 40,000 workers, including more than 2,800 physicians and scientists in 120 medical specialties and subspecialties in clinics around the world.

Cleveland Clinic developed a comprehensive strategy for an enterprise business intelligence (EBI) program to help it focus on defining, measuring, and improving the key performance indicators necessary to fulfill strategic performance management objectives. The EBI program addresses people, processes, governance, and technology as it improves decision making by using automated enterprise data warehouse information in fact-based analytics.

The company's solution allows leaders to gain real-time views of actual performance and help focus scarce resources, accelerate change, and achieve performance management excellence. It enabled Cleveland Clinic to create a comprehensive set of management dashboards with timely information, adding quantifiable value in less than a year. The solution also helped Cleveland Clinic reduce costs and enhance quality of care, even during periods of national and local economic turmoil.

Cleveland Clinic's enterprise performance management program demonstrates that comparisons between actual performance and strategic objectives can enable healthcare organizations to achieve rapid organizational change. This change has reduced costs, increased revenues, and improved the quality of patient care at Cleveland Clinic. Key lessons learned include the importance of executive championship, dedicated resources, openness to resolving data anomalies, clarity and appropriate communication, and scalable technology architecture.

➤ **Right-Time BI**  
**Station Casinos, Inc.**

SOLUTION SPONSOR: INFORMATICA CORPORATION

Founded in 1976 in Las Vegas, Nevada, Station Casinos provides gaming and entertainment for residents of the Las Vegas metropolitan area. Station's properties include hotels and restaurants, entertainment venues, movie theaters, bowling and convention/banquet space, and traditional casino gaming offerings.

Station Casinos wanted a better understanding of customer activities and preferences and the ability to offer personalized incentives to increase its customer base. Station Casinos undertook an initiative to gather data from its 18 resorts and casinos and provide a 360-degree, real-time view of customer activities. The data would include all customer "touch points," including financial, gaming, hotel, food and beverage, and other data to understand customer trends, usage patterns, and the overall casino experience.

To get this level of detailed information, Station Casinos deployed the Informatica Platform, which provides the ability to access, clean, and deliver critical data in real time, including integrated customer, gaming, and finance information from as many as 500 heterogeneous sources and 9.2 million customer records.

The solution is fully operational and has achieved its goals using real-time data integration and analytics across numerous properties and venues and customer touch points to power more rewarding customer service and build customer loyalty. The project shows how it is possible to overcome hurdles to real-time BI such as massive numbers of siloed data sources, high volumes of rapid-fire database changes, and development of new services to enhance customer service and build customer loyalty.



## **Birst**

[www.birst.com](http://www.birst.com)

### **BI CATEGORY:** Analytics and Reporting

Birst™ is the leader in agile business analytics. It offers a single place to manage all of a business's analytics and agility to answer questions that span departments, data sources, and deployments—across public or private clouds. Birst gives users the fastest way to answer their most pressing business questions—and then all the ones they didn't know to ask. At one-third the cost, time, and staff of traditional big BI, Birst brings the benefits of analytics and fact-based decision making to a much broader audience.

# **BI SOLUTIONS**

## Transforming Technologies

---

Our sponsors present their solutions in the following business intelligence categories:

- Analytics and Reporting
- Big Data
- Business Intelligence
- Dashboards, Scorecards, and Visualization
- Data Governance
- Data Integration
- Data Management
- Data Quality
- Data Warehousing
- Enterprise Business Intelligence
- Master Data Management
- Open Source Business Intelligence
- Predictive Analytics





## DataFlux

[www.dataflux.com](http://www.dataflux.com)

**BI CATEGORIES:** Data Governance, Data Integration, Data Management, Data Quality, Master Data Management

DataFlux provides a robust and complete solution for companies looking to implement a master data management (MDM) program, but aren't sure where to start. DataFlux technology is comprehensive and easily deployed—plus DataFlux offers professional services and learning programs to help support your program. It's everything you need to get your program up and running, and stay competitive in the year of MDM.

Of all the factors that drive an organization's success, customers and products are at the top of the list. To increase revenue, it's critical to know who your customers are, what they purchase, and how to sell more to them. And you need to know what you're spending on materials across all business units to control costs.

DataFlux helps organizations of all sizes and various industries create a single, accurate, and unified view of corporate data, integrating information from various data sources into one master record. This master data is then used to feed information back to the applications, creating a consistent view of information across the enterprise. MDM-savvy companies are able to enjoy more revenue, less risk, better customer relationships—all while uncovering new opportunities to maximize spend.



## Information Builders

[www.informationbuilders.com](http://www.informationbuilders.com)

**BI CATEGORIES:** Business Intelligence, Enterprise Business Intelligence

Information Builders provides software and services that transform data into business value—and our customers into market leaders—through information intelligence, integration, and integrity.

Our flagship WebFOCUS product line empowers every employee, partner, and customer to make smarter decisions with business intelligence and analytics. Our iWay integration products simplify the orchestration of complex data and application environments to deliver projects on time and on budget, and our iWay integrity products capture lost revenue and eliminate hidden costs generated by inaccurate and inconsistent data.

But great software is nothing without great people. Our corporate culture places a high value on getting smart employees who can develop deep customer relationships so that, together, we can deliver game-changing results. Every major research firm—including BARC, Bloor, Butler, TDWI, Gartner, Ventana, and others—has seen the results of this from our customers and has recognized our superior service. We have also achieved high honors from CRM Magazine, the SSPA, and the American Business Awards.

As a result, our customers have received substantial information technology and business awards for their accomplishments. Over 50 have even had their information systems inducted into the Smithsonian Institute for superior information technology achievement. Work with us, and you'll see why.



## Lavastorm Analytics

[www.lavastorm.com](http://www.lavastorm.com)

### **BI CATEGORY:** Analytics and Reporting

Lavastorm Analytics is a global business performance analytics company whose solutions are used by thousands of business and IT professionals and are applicable to almost any industry, including communications, financial services, insurance, utilities, and healthcare.

The company's flagship product, the Lavastorm Analytics Platform, is a discovery-based analytic platform that enables companies to analyze, optimize, and control the performance of complex business processes. The platform offers the following advantages:

- **Self-service analytics:** A visual, step-by-step interface provides both business and technical users the self-sufficiency and confidence to rapidly join dissimilar data in a fraction of the time normally required by traditional approaches.
- **Visualization of process/performance issues:** A library of over 100 configurable analytic building blocks gives users the speed to assemble and execute ad hoc analyses.
- **Continuous monitoring and auditing:** Analytics created with the platform can be used as persistent automated controls that continuously monitor the efficiency and compliance of business processes.
- **Case management capabilities for issue resolution:** The platform provides visualization, reporting, alarming, workflow, and case management capabilities that deliver insights and enable automatic interception and resolution.

*Fortune* 500 organizations and individuals alike rely on the Lavastorm Analytics Platform to improve complex financial, operational, and customer experience processes.



## Morgan Kaufmann

[www.mkp.com](http://www.mkp.com)

### **BI CATEGORIES:** Business Intelligence, Data Governance, Data Integration, Data Management, Data Quality, Data Warehousing, Master Data Management, Open Source Business Intelligence, Predictive Analytics

Morgan Kaufmann has been bringing the knowledge of experts to the computing community since 1984. Our goal is to provide timely yet timeless content to research and development professionals, business leaders and information technology managers, everyday practitioners, and academia. Through superior print and digital content, our leading authors educate readers and inspire future innovation. We publish textbooks that often redefine the trajectory of computer science education and technical references that help professionals stay ahead of the curve.



## SAND

[www.sand.com](http://www.sand.com)

### BI CATEGORY: Data Warehousing

The SAND Analytic Platform is a columnar analytic database platform that achieves linear data scalability through massively parallel processing (MPP), breaking the constraints of shared-nothing architectures with fully distributed processing and dynamic allocation of resources. SAND supports thousands of concurrent users with mixed workloads, infinite query optimization (requiring no tuning once data is loaded), in-memory analytics, full text search, and SANDboxing for immediate data testing. The SAND Analytic Platform focuses on complex analytics tasks, including customer loyalty marketing, churn analytics, and financial analytics.



## Sybase, An SAP Company

[www.sybase.com](http://www.sybase.com)

### BI CATEGORY: Big Data

The big data onslaught has IT professionals running scared. Everyone wants access to business intelligence but the fear of wreaking havoc on systems has IT worried. However, Sybase thinks the big data crisis is a big lie. It's simply an evolution of dire predictions that go back 60-plus years. Sybase IQ delivers a big data solution that cost-effectively scales performance and users with your enterprise needs. That's because our PlexQ analytics grid was built to answer to thousands of simultaneous ad hoc and complex queries against terabytes or petabytes of data. No other solution can take your big data efforts from one-time projects to enterprisewide analytics. Only Sybase IQ delivers the capability to handle any type of analytic workload against big data—both structured and unstructured—from lightning-fast predictive analytics to new analytic paradigms like a native MapReduce API advanced and flexible Hadoop integration, Predictive Model Markup Language (PMML) support, and an extensive library of statistical and data mining algorithms that leverage the power of distributed query processing across a PlexQ grid. Don't let the big data onslaught bring you down. Look to Sybase IQ to turn your massive data into actionable intelligence that's usable throughout your enterprise.



## TIBCO Spotfire

spotfire.tibco.com

**BI CATEGORY:** Dashboards, Scorecards, and Visualization

TIBCO Software Inc. provides infrastructure software for companies to use on-premises or as part of cloud computing environments. TIBCO Spotfire Analytics is the company's enterprise analytics platform that helps users quickly get to actionable insight. Spotfire Analytics helps users at every level of the organization seize new business opportunities and evade risks by providing unmatched speed and flexibility so data can be visually analyzed quickly, without requiring IT intervention.

### BENEFITS

- **Fastest to actionable insight:** Allows organizations to increase productivity by providing users throughout the organization with unmatched speed and flexibility to anticipate, answer, and act on answers to relevant questions at the speed of thought
- **Visibility into the unknown:** Enables users to uncover trends, patterns, and unexpected insights hidden in data that represent business opportunities or threats through intuitive visualizations, analytic dashboards, and applications
- **Self-service discovery:** Increases agility and productivity by drastically reducing reliance on IT and eliminating time related to data preparation, report building, and spreadmart creation
- **Universal adaptability:** Spotfire Analytics is a single highly scalable analytics platform that empowers a broad spectrum of business and technical users across a wide range of analytic use cases and processes to act on data-driven insights quickly



## Vertica, An HP Company

www.vertica.com

**BI CATEGORY:** Data Warehousing

Vertica, An HP Company, is the leading provider of next-generation analytics platforms enabling customers to monetize all of their data. Vertica's elasticity, scale, performance, and simplicity are unparalleled in the industry, delivering 50-1000 times the performance of traditional solutions at 30 percent the total cost of ownership. Vertica powers some of the largest organizations and most innovative business models globally including Zynga, Groupon, Twitter, Verizon, GUESS?, Inc., Admeld, Capital IQ, Mozilla, AT&T, and Comcast.

The Vertica Analytics Platform is easy to use and deploy, so users across an organization (not just DBAs) can get up and running quickly and immediately analyze mission-critical data. And Vertica offers the flexibility to deploy business intelligence where it's needed—on off-the-shelf hardware, virtual machines, or in the cloud. At its core is a columnar database, purpose-built to take advantage of a massively parallel processing (MPP) architecture, and offering a familiar interface and hooks into the analytics ecosystem—from ETL tools to visualization—for easy integration with a company's existing analytics environment.



## About TDWI

TDWI, a division of 1105 Media, Inc., is the premier provider of in-depth, high-quality education and research in the business intelligence and data warehousing industry. TDWI is dedicated to educating business and information technology professionals about the best practices, strategies, techniques, and tools required to successfully design, build, maintain, and enhance business intelligence and data warehousing solutions. TDWI also fosters the advancement of business intelligence and data warehousing research and contributes to knowledge transfer and the professional development of its members. TDWI offers a worldwide membership program, five major educational conferences, topical educational seminars, role-based training, onsite courses, certification, solution provider partnerships, an awards program for best practices, live Webinars, resourceful publications, an in-depth research program, and a comprehensive Web site, [tdwi.org](http://tdwi.org).



TDWI Education has even more to offer. Visit [tdwi.org/education](http://tdwi.org/education) for a full lineup of Solution Summits, Solution Spotlights, Forums, and BI Executive Summits.

### PREMIUM MEMBERSHIP

[tdwi.org/premiummembership](http://tdwi.org/premiummembership)

In a challenging and ever-changing business intelligence and data warehousing environment, TDWI Premium Membership offers a cost-effective solution for maintaining your competitive edge. TDWI will provide you with a comprehensive and constantly growing selection of industry research, news and information, and online resources. TDWI offers a cost-effective way to keep your entire team current on the latest trends and technologies. TDWI's Team Membership program provides significant discounts to organizations that register individuals as TDWI Team Members.

### WORLD CONFERENCES

[tdwi.org/conferences](http://tdwi.org/conferences)

TDWI World Conferences provide a unique opportunity to learn from world-class instructors, participate in one-on-one sessions with industry gurus, peruse hype-free exhibits, and network with peers. Each six-day conference features a wide range of content that can help business intelligence and data warehousing professionals deploy and harness business intelligence on an enterprisewide scale.

### SEMINAR SERIES

[tdwi.org/seminars](http://tdwi.org/seminars)

TDWI Seminars offer a broad range of courses focused on the skills and techniques at the heart of successful business intelligence and data warehousing implementations. The small class sizes and unique format of TDWI Seminars provide a high-impact learning experience with significant student-teacher interactivity. TDWI Seminars are offered at locations throughout the United States and Canada.

### CHAPTERS

[tdwi.org/chapters](http://tdwi.org/chapters)

TDWI sponsors chapters in regions throughout the world to foster education and networking at the local level among business intelligence and data warehousing professionals. Chapter meetings are open to any BI/DW professional. Please visit our Web site to find a local chapter in your area.

### ONSITE EDUCATION

[tdwi.org/onsite](http://tdwi.org/onsite)

TDWI Onsite Education is practical, high-quality, vendor-neutral BI/DW education brought to your location. With TDWI Onsite Education, you maximize your training budget as your team learns practical skills they can apply to current projects—with Onsite training tailored to their specific needs.

### CERTIFIED BUSINESS INTELLIGENCE PROFESSIONAL (CBIP)

[tdwi.org/cbip](http://tdwi.org/cbip)

Convey your experience, knowledge, and expertise with a credential respected by employers and colleagues alike. CBIP is an exam-based certification program that tests industry knowledge, skills, and experience within four areas of specialization—providing the most meaningful and credible certification available in the industry.

### WEBINAR SERIES

[tdwi.org/webinars](http://tdwi.org/webinars)

TDWI Webinars deliver unbiased information on pertinent issues in the business intelligence and data warehousing industry. Each live Webinar is roughly one hour in length and includes an interactive question-and-answer session following the presentation.