

The background of the entire cover is a vibrant, high-contrast image of an explosion. A bright, glowing yellow and orange core is at the center, with numerous sharp, radiating lines of light in shades of purple, magenta, and red extending outwards to the edges of the frame. The overall effect is one of intense energy and dynamic movement.

tdwi's
**Best of Business
Intelligence**
A Year in Review

VOLUME 10

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

2012 IN REVIEW:

THE BIG DATA EXPLOSION

PLUS 2013 FORECAST:

**USERS CONTINUE TO ADOPT NEW BI TOOLS,
DESIRE FLEXIBILITY AND SPEED WITH DATA**

RESEARCH EXCERPTS

Applying Technologies for Social Media Data Analysis
Introduction to High-Performance Data Warehousing

INSIGHTFUL ARTICLES

Big Data and Big Analytics
The 2020 Workplace
Ten Myths about Hadoop

tdwi
THE DATA WAREHOUSING INSTITUTE



A smarter planet is built on smarter analytics.

From the Industrial Age to the rise of the globally integrated enterprise, every new era has been driven by a fundamental change in how organizations relate to the world. But it's how businesses adapt to these changes that defines who will succeed and who will be left behind. Mounting evidence suggests that we're entering a new era right now, and the change that's driving it is the rise of big data—the 2.5 quintillion bytes being produced each day, which represents the collective output of every person, every organization and every instrumented thing.

In a world defined by big data, it's hard to overstate the importance of analytics, which makes it possible to find opportunities in the noise, to see the patterns hidden in the data and, ultimately, to predict what's coming next and act on it.

In fact, a 2011 *MIT Sloan Management Review* and IBM study found that organizations that embrace analytics are 2.2 times more likely to substantially outperform industry peers.

But not all approaches to analytics are equally smart. A smarter approach means the ability to apply analytics to *all* of an organization's data—not just the structured data from databases or applications, but also the more elusive data that exists outside a business's four walls: things like weather patterns, economic trends, even tweets.

This isn't just theory. Top businesses are putting it into practice, moving beyond *using* analytics to actually *becoming* analytic.

For Belgium's leading ticket vendor, Tele Ticket Service, this has meant changing one of the most fundamental parts of their business: their pricing model. By using analytics to “read” trends in ticket sales, they've been able to introduce flexible pricing that

automatically adjusts the prices of tickets in different seating areas based on trends in demand.

Other leading businesses are using analytics to transform their financial processes. By leveraging tools like real-time dashboards, financial close process automation and financial forecasting, Omnicom Group was able to reduce its investment in working capital by \$200 million.

And for McKesson, a pharmaceutical distributor, analytics is reshaping the very core of their business. With a holistic view of their entire supply chain, they can now cost out services based on product line, mode of transportation, even their carbon footprint, allowing them to be more responsive to customer needs and to reduce their working capital by more than \$100 million.

These aren't isolated examples. In more than 20,000 engagements over the last three years, IBM has helped organizations internalize a smarter approach to analytics—giving IBM the unique vantage point to track patterns across industries and apply those insights to every new client engagement. This commitment to analytics has resulted in well over 500 analytics-related patents every year, contributing to IBM's record as the number one recipient of U.S. patents for 19 years running.

As we move deeper into the era of big data, IBM will continue to push the boundaries of what's possible with analytics, including pioneering new systems like Watson that not only are able to assess and predict, but also are able to learn and reason. On a smarter planet, the opportunities are as vast as the data.

Let's build a smarter planet. Join us and see what others are doing at ibm.com/smarteranalytics



SPONSOR INDEX

Birst

IBM

Lavastorm Analytics

Melissa Data

Tableau Software

Vitria

TABLE OF CONTENTS

FEATURES

- 5 2012 in Review: The Big Data Explosion
STEPHEN SWOYER
- 10 2013 Forecast: Users Continue to Adopt New BI Tools,
Desire Flexibility and Speed with Data
PHILIP RUSSOM, DAVID STODDER

TDWI BEST PRACTICES REPORTS

- 19 Applying Technologies for Social Media Data Analysis
DAVID STODDER
- 24 Introduction to High-Performance Data Warehousing
PHILIP RUSSOM

TEN MISTAKES TO AVOID SERIES

- Ten Mistakes to Avoid When Validating Your BI/DW Direction
JONATHAN G. GEIGER

TDWI FLASHPOINT

- 32 TDWI Salary Survey: Wages, Bonuses on the Rise
MARK HAMMOND
- 34 What's Really New about Big Data and Big Analytics?
COLIN WHITE, CLAUDIA IMHOFF

BUSINESS INTELLIGENCE JOURNAL

- 36 The 2020 Workplace and the Evolution of
Business Intelligence
HOLLIS HENRY, TROY HILTBRAND
- 42 The Requirements for Being an Analytics-Based Organization
HUGH J. WATSON

BI THIS WEEK

- 45 Busting 10 Myths about Hadoop
PHILIP RUSSOM
- 47 The Confusing Future of BI and Data Warehousing
STEPHEN SWOYER

TDWI CHECKLIST REPORT SERIES

- 50 Predictive Analytics Project Design: A Nine-Step Guide
to Realizing Business Value
THOMAS A. "TONY" RATHBURN

55 TDWI WEBINAR SERIES

56 TDWI EDUCATION

Orlando World Conference Keynote

57 2013 TDWI EVENTS CALENDAR

58 BEST PRACTICES AWARDS 2012

64 BI SOLUTIONS

68 ABOUT TDWI

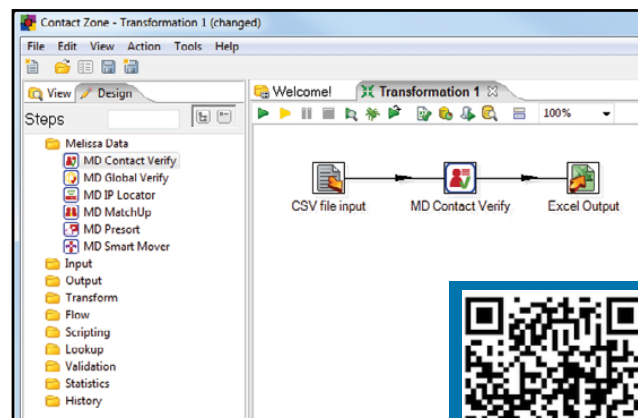
Your Data. Transformed.



Melissa Data's Contact Zone® – powerful data quality with advanced data integration capabilities – works by pulling in, cleaning, validating and enriching contact data, before quickly outputting the results to virtually any destination. Contact Zone offers an end-to-end solution to easily transform data into actionable intelligence.

Powerful Features:

- Global address, phone verification, and geocoding
- Create data flows with visual programming
- Flexibility for real-time or scripted batch jobs
- Built-in reporting
- Universal connectivity



Drag-and-Drop with a Palette of Powerful Data Transformations



SCAN TO WATCH A SHORT VIDEO

Try Contact Zone for Free at:
www.MelissaData.com/cz
Or call 1-800-MELISSA (635-4772)

MELISSA DATA®
Your Partner in Data Quality

tdwi.org

Editorial Director **Denelle Hanlon**
Managing Editor **Jennifer Agee**
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, 248.658.6365

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

Phone 212.221.9595; E-mail 1105reprints@parsintl.com;
Web www.magreprints.com/QuickQuote.asp

© Copyright 2013 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 2013, 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

We're proud to introduce TDWI's tenth annual *Best of Business Intelligence: A Year in Review*. Each year we select a few of our 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 "2012 in Review: The Big Data Explosion," he argues that data management and big data are changing as we know them—and the data warehouse could be next.

In this issue's "2013 Forecast," TDWI Research analysts Philip Russom and David Stodder share their predictions for the coming year, including the emerging technologies they expect to be adopted as well as trends shaping BI and analytics practices and technologies. And don't miss our first inclusion of a Checklist Report: *Predictive Analytics Project Design: A Nine-Step Guide to Realizing Business Value*.

To further represent TDWI Research, we've provided excerpts from some of the past year's Best Practices Reports. Stodder's "Applying Technologies for Social Media Data Analysis" covers customer analytics technologies and methods for social media data, and Russom's "Introduction to High-Performance Data Warehousing" defines "HiPer DW" and how it's perceived by organizations.

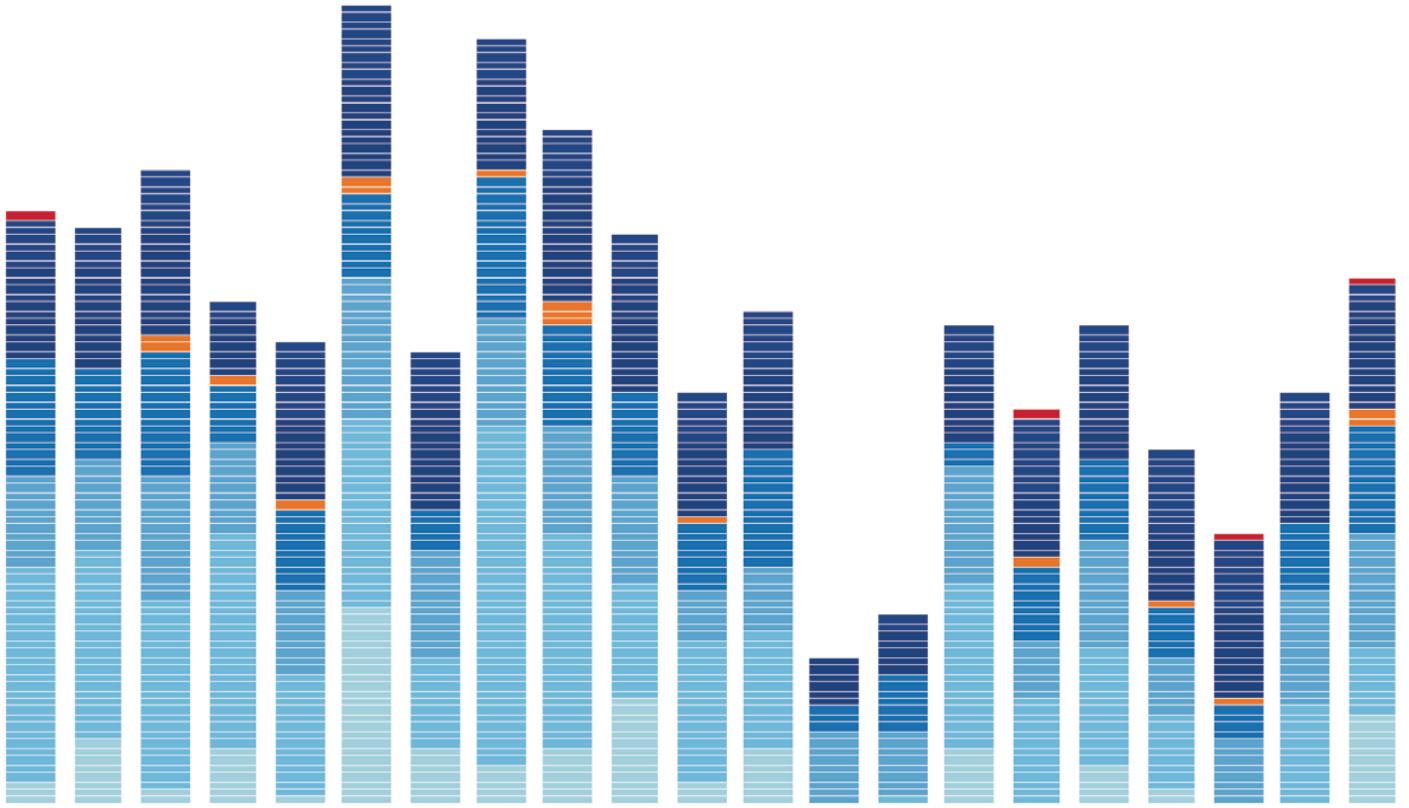
This volume's *Ten Mistakes to Avoid* will help you avoid some common pitfalls when validating your BI and data warehousing (DW) direction. And thanks to articles from TDWI's e-newsletters, you'll learn more about salary trends in the BI/DW industry, big data analytics, Hadoop myths, and the future of BI/DW.

In "The 2020 Workplace and the Evolution of Business Intelligence," one of our selections from the *Business Intelligence Journal*, you'll read about the evolving influence of BI in the workplace. Our second *Journal* piece, "The Requirements for Being an Analytics-Based Organization," helps organizations succeed with advanced analytics.

For the first time, *TDWI's Best of Business Intelligence* is including a selection of our informative, on-demand Webinars, as well as a peek inside a TDWI World Conference keynote address, given by popular speaker Cindi Howson.

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 TDWI's *Best of Business Intelligence*, Volume 10.

Denelle Hanlon
Editorial Director, *TDWI's Best of Business Intelligence*
The Data Warehousing Institute
ghanlon@tdwi.org



See and understand your data in seconds with Tableau.

The data you capture has massive potential to give you a competitive advantage. Turn that potential into reality with fast visual analysis and easy sharing of reports and dashboards from Tableau. By revealing patterns, outliers and insights, Tableau helps you find more insights in your data and get more value from it.



Tableau provides:

- Live, optimized connection to a variety of data sources
- The ability to analyze truly big data at interactive speeds
- An in-memory data engine that overcomes slow databases
- Drag & drop data visualization – no coding required

Tableau is changing the way companies are analyzing and sharing their data.
For a free trial visit, www.tableausoftware.com/tdwi

2012 IN REVIEW:

THE BIG DATA EXPLOSION

BY STEPHEN SWOYER

In 2012, several long-simmering forces emerged—and in the case of big data, exploded—to challenge the data management (DM) status quo. Collectively, the changes we saw in 2012 make for a perfect storm of challenges to data management's orthodoxy.

Data Management at a Crossroads

In 2012, the selection pressures acting on data management came into sharp relief.

For starters, the year served to confirm the business intelligence (BI) discovery trend, with new discovery-oriented offerings from SAS (Visual Analytics Explorer) and SAP AG (Visual Intelligence). At this point, all of the BI heavyweights—including Information Builders, IBM Cognos, Microsoft, MicroStrategy, and Oracle, along with SAP and SAS—market discovery-themed solutions aimed at blunting the success of the original discovery players such as TIBCO Spotfire and Tableau Software.

There's another wrinkle: in addition to adopting the visual discovery metaphor, the big BI players are building enhanced self-service and collaborative features into their products. In this respect, they've taken a page from another successful (and successfully disruptive) vendor: workgroup BI specialist QlikTech.

The upshot is that insurgent BI has forced the BI Powers That Be to adapt. Half a decade ago, this would've been unthinkable. "I'm an IT person who was an operational manager for 20 years. I bought those [BI] technologies, and I was a customer of Tableau's, too," says Dan Murray, director of business intelligence services and COO of InterWorks, an Atlanta-based integration and services firm that specializes in Tableau.

"I don't view Tableau as a replacement for any BI or [analytic] database company," Murray continued. "It's just going to make any existent [analytic] database or BI deployment better—make it more accessible [and] more useful to the average information consumer."



REGARDLESS OF HOW YOU POSITION IT, ONE THING IS CLEAR: DATA MANAGEMENT IS CHANGING.

Change isn't confined to the outer reaches of data management (i.e., BI tools). Analytic databases, for example, have been with us for a decade, but 2012 introduced a new variation on this theme: the analytic discovery platform, such as those marketed by ParAccel and Teradata. ParAccel's pitch is as an analytic discovery platform: an alternative to the overwhelmed enterprise data warehouse (EDW). Teradata's positioning is more complicated. The EDW is its bread and butter, and at its October Partners conference, Teradata unveiled its architecture (with supporting software and services) for Unity, a DM ecosystem that accords its Aster platform primacy of place as a *complementary* analytic discovery platform.

Regardless of how you position it, one thing is clear: DM is changing, and 2012 was the year the implications of this change first started coming into focus.

Big Data Explosion

If "discovery" gained critical mass in 2012, big data, by any objective standard, went supernova.

At the TDWI BI Executive Summit on big data in San Diego, for example, two TDWI regulars—Mark Madsen and colleague Marc Demarest—teamed up for an electrifying seminar on the pros and cons of big data. Both men are alike in thinking that big data is the Real Deal. Demarest, a principal with information management consultancy Noumenal, Inc., makes his case with characteristic bluntness. "The change is upon us and there is no way back," he argues.

Madsen, for his part, sees big data as a paradigm-shifting event. "We're in the midst of a paradigm shift, but the thing about paradigm shifts is that they take a long time. [T]he new paradigm is evident but not yet manifested," says the veteran data warehouse architect and a principal with BI and data management consultancy Third Nature, Inc. "We are in a market state where nobody has written the definitive architecture for the new world."

Marketers love a paradigm shift. In the case of big data, however, BI marketers at first failed to recognize what it is that makes a paradigm shift so special. Instead of promoting big data on the basis of its potential as a transformative force—as something that, fully realized, can radically reshape how we understand the world—the industry glommed on to the idea of promoting big data as a function of the volume, velocity, or variety of the information that comprises it.

Of course, the three Vs aren't new. They've always been with us. In fact, Gartner analyst Doug Laney first coined the volume-variety-velocity triptych more than a decade ago. If the term *big data* simply describes the volume, variety, and velocity of the information that constitutes it, our existing data management practices are still arguably up to the task. It also has the effect of pigeonholing big data as a data-management-specific event: paradigm shifts are sweeping, cutting across disciplines and domains.

If big data is as big a shift as Madsen, Demarest, and others believe, it must mean something more than volume, velocity, and variety. Savvy industry watcher and TDWI contributor Ted Cuzzillo, who blogs about BI at Datadoodle.com, uses the analogy of television to describe both what big data is and what it could ultimately be.

"Right now, big data's analogous to early TV. Skeptics called it 'radio with pictures,' and some of it was little more than that. But its new dimensions developed, with ever-higher resolution. Soon enough we had 'living color,' then HD, and now some of it's in 3-D," he explains.

That's the endgame of the big data paradigm shift: 3-D context. When, if ever, we'll achieve this is anybody's guess.

A New Role for the Data Warehouse?

Discovery and especially big data are putting pressure on DM practitioners to retest, re-evaluate, and (in some cases) *discard* core organizing or operating assumptions.

Even the data warehouse itself is coming under scrutiny.

We started to get a sense of this in 2012. Even after a year of big data hoopla, few data management practitioners can conceive of a world that doesn't have a data warehouse at its center. Try telling that to the average attendee at October's inaugural Strata + Hadoop World conference, however. Such folk aren't part of the data management mainstream; in fact, they're used to viewing the data warehouse as an obstacle—or as an archaism. It isn't that they can't conceive of a world that doesn't have a data warehouse at its center; it's that in many cases they're actively anticipating the emergence of just such a world.

That's the point: the era of the data warehouse as its own isolated fiefdom is ending; DM, and the data warehouse along with it, are being coaxed—or dragged—out into the open.

Big data is placing selection pressure on the DW in a number of ways. Some vendors inside and outside the data management industry envision the Hadoop framework—which by the end of 2012 had become virtually synonymous with big data—as an information management platform residing alongside (and possibly displacing) the traditional data warehouse.

There was David Inbar, senior director of big data products with data integration (DI) specialist Pervasive Software, who eloquently describes Hadoop as “a beautiful platform for all kinds of computation.” At this summer's Pacific Northwest BI Summit, held in Grants Pass, Oregon, Yves de Montcheuil, vice president of marketing with open source software (OSS) DI vendor Talend, outlined a vision of Hadoop as the central site of enterprise information integration. In 2012, Pervasive and Talend, along with competitors Informatica and Syncsort, all trumpeted Hadoop-centered, big-data-focused product or service announcements. Pervasive and Syncsort both announced dedicated ETL libraries for Hadoop, and Talend promoted the idea of Hadoop- and MapReduce-powered ETL. Informatica, for its part, announced a “Big Data Edition” of its PowerCenter ETL platform.

DI is a natural fit for Hadoop, which (when paired with MapReduce) has been described as the equivalent of a brute-force, massively parallel ETL platform. In 2012 we saw a new (more ambitious) variation on this theme thanks to a slew of BI-like offerings based on Hadoop.

THE ERA OF THE DATA WAREHOUSE AS ITS OWN ISOLATED FIEFDOM IS ENDING; DM, AND THE DW ALONG WITH IT, ARE BEING COAXED—OR DRAGGED—OUT INTO THE OPEN.



At Strata + Hadoop World, for example, Cloudera unveiled “Impala,” a real-time, interactive query engine that runs inside Hadoop. (One reason Hadoop is promoted for ETL is because it's a batch-centric processing environment.) Impala enables OLAP-driven discovery in a Hadoop environment, along with other BI-like use cases. Cloudera competitor MapR uses a different approach (basically mounting the Hadoop File System as an NFS share) to achieve a similar end; DataMeer, also at Strata + Hadoop World, touted a BI-like analytic discovery environment that it likewise implements on top of Hadoop. Other upstart players, such as Platfora, take similar Hadoop-centric, de-emphasized DW approaches.

From Redshift to Paradigm Shift

Perhaps the most intriguing news item of 2012 was Amazon.com's November announcement of Redshift, its data-warehouse-in-the-cloud offering for Amazon Web Services (AWS).

Redshift is based on a respected massively parallel processing (MPP) engine: the ParAccel analytic database. This technology would count for little, however, if Amazon hadn't also addressed the Achilles' heel of data warehousing in the cloud: unpredictable I/O performance. There's good reason to believe that AWS, which uses solid-state, disk-based (SSD) local storage, does just that. This is what makes Redshift so intriguing.

After all, the concept of the DW in the cloud isn't new. Analytic database stalwart Kognitio first announced a data-warehouse-as-a-service (DaaS) offering half a decade ago; shortly thereafter, the former Vertica (now part of HP) followed suit.

These offerings—and others—scale out by virtue of provisioning new "instances" of a database engine; this means running (and managing) additional copies of a database. In the public cloud, and—specifically in the case of I/O sensitive parallel processing—this invariably involves trade-offs in I/O performance and elasticity. That's because the traditional DW-as-a-service model constitutes the transplanting of technology that was designed and perfected for use in a well-defined paradigm (i.e., a distributed, physical client-server topology) into a kind of alien context: viz., that of an elastic, multi-tenanted, inescapably virtual topology.

This doesn't mean that Amazon has delivered a Redshift DW-as-a-service that rivals on-premises platforms from Actian, HP, EMC Greenplum, IBM, Kognitio, Oracle, Teradata, and SAP AG, or ParAccel, for that matter. "Amazon fixed this [I/O] problem way back when, but databases want to own the entire machine, not little bits of several [machines]. You can't run a query database with unpredictable I/O or unpredictable interconnects. This is why MPP databases in Amazon traditionally haven't been very successful: one node always trails for some reason," says Madsen, who argues that Redshift's target customer isn't a Teradata or Netezza shop; it's a company that's maxed out on an RDBMS-powered data warehouse. "To run SQL Server and other [data warehouse] platforms, you need a DBA to configure and run it in the cloud, just like a regular server. This is a service. You set it up and go."

A NoSQL Armistice

Love it or hate it, NoSQL is here to stay. If nothing else, it's *established*. This past year, for example, MongoDB turned five; this year, the Cassandra project will celebrate its fifth birthday. On top of this, NoSQL-like players such as MarkLogic and RainStor have been around even longer.

Something anticlimactic happened in 2012: the NoSQL wars came to an end without so much as a bang and with barely even a whimper. Whatever its problems or shortcomings from a DM perspective, NoSQL is now accepted as a legitimate product category. At all of TDWI's industry conferences in 2012, NoSQL exhibitors shared floor space with the likes of IBM, Oracle, and Teradata. There's evidence of acceptance in other contexts, too: SAP, for example, embeds NoSQL as one of three in-memory engines in its HANA platform.

"When I present on Hadoop and NoSQL, I always point out that it's [the product of a] programmer-centric world. It was built to solve a class of problems that these Web [application developers] were encountering for the first

time," says veteran data warehouse architect and TDWI presenter John O'Brien, a principal with information management consultancy Radiant Advisors. "Whatever we [data management practitioners] might think of it, NoSQL was really built to exploit the next generation of [Web] apps, not the SQL-driven [class of] business intelligence and analytic [tools]. It has a kind of expedient purpose."

Cloudpocalypse BI?

Although BI in the cloud has something of a checkered history, it's possible that Amazon Redshift, in combination with other upstart projects, might finally push BI and DW vendors to get more serious about taking on the cloud. Not on their own terms (by transplanting on-premises tools to a cloud context) but on terms appropriate to the multi-tenanted, virtual turf that's characteristic of the cloud, and which is integral to its promise.

THIS PAST YEAR WAS FULL OF
OTHER INNOVATIONS—OR
MODIFICATIONS—OF THE BI
STATUS QUO.



In addition to Amazon and Redshift, a pair of cloud-focused start-ups—Akiban Technologies and NuoDB—emerged in 2012 to tout their own, built-from-scratch takes on a DB in the cloud. Akiban in early February 2012 announced Akiban server, which it describes as a cloud-based, ACID-compliant ("NewSQL") DBMS platform. NuoDB iterated throughout 2012 on its tiered, SQL-compliant cloud DBMS platform that uses a mix of redundancy and *probabilism* to achieve ACID compliance. By December, NuoDB 1.0 was approaching general availability.

We also saw a big splash in the U.S. by Australian BI specialist Yellowfin, which—in spite of its cloud underpinnings—touts a retro (reporting-centric) take on business intelligence. Yellowfin positions its "mass-production" BI platform in the cloud as an alternative to kitchen-sink suites that, according to CEO Glen Rabie, "don't do anything particularly well."

Another BI player that made a series of major cloud moves in 2012 was Jaspersoft, which signed partnerships

with Red Hat and VMware to deliver platform-as-a-service (PaaS) cloud BI. (Prior to 2012, Jaspersoft had an existing arrangement with Amazon for SaaS BI.) "This data-driven world ... is increasingly going to be cloud hosted, [and] the analytics piece is going to be vital. We expect that [an] analytic reporting service is going to be a ... *de facto* component that PaaS providers are going to offer," Karl Van den Bergh, Jaspersoft's vice president of product and alliances, told TDWI's *BI This Week* e-newsletter in July.

Parting Thoughts

This past year was full of other innovations—or modifications—of the BI status quo. For example, a trio of vendors—Armanta, Cirro, and Quest—delivered what amounted to self-contained analytic platforms: products that combine a self-service analytic tool set with an underlying data virtualization (DV) layer and leave the DI and data preparation heavy lifting to us. Like DV players such as Composite Software and Denodo Technologies, they use DV to create canonical representations (or “views”) of data in source systems. Unlike Composite and Denodo, their respective DV technologies lack the refinement and hardening that accrue from extensive production use.

Acquisition-wise, 2012 was a strange year. Past years have been characterized by consolidation waves: there was 2003, for example, with its BI reporting consolidation; 2005, with its DI-focused consolidation wave; or 2007, with its BI suite extinction event. Things were relatively quiet on the acquisition front in 2012, however. There was acquisition activity, to be sure: QlikTech acquired veteran DI player Expressor Software in June, Dell bought veteran DM player Quest Software, and Oracle bought DataRaker (a machine-generated data specialist). The year nevertheless lacked significant *consolidation activity*, though a spate of related acquisitions (such as those in reporting, DI, or BI suites) are a sure sign of a red-hot market. Expect plenty of consolidation in 2013—with big data being a likely vector.

The truth is, 2012 was an extremely eventful year, and 2013 promises to be at least as eventful.

Stephen Swoyer is a technology writer based in Nashville, TN. Contact him at stephen.swoyer@spinkle.net.



2013 FORECAST

USERS CONTINUE TO ADOPT NEW BI TOOLS, DESIRE FLEXIBILITY AND SPEED WITH DATA

BY PHILIP RUSSOM AND DAVID STODDER

TRENDS IN EMERGING TECHNOLOGIES AND METHODS FOR BUSINESS INTELLIGENCE

PHILIP RUSSOM, RESEARCH DIRECTOR FOR DATA MANAGEMENT, TDWI

Part of the fun of being in business intelligence (BI), data warehousing (DW), data integration (DI), and analytics is the constant stream of new and exciting technologies, vendor tools, team structures, and user methodologies. TDWI refers to these collectively as emerging technologies and methods (ETMs).

Some ETMs are so new that they are truly just emerging. These include agile development methods for BI, BI tools and platforms on clouds, event processing, Hadoop, MapReduce, mashups, mobile BI, NoSQL, social media, solid-state drives, and streaming data. Other ETMs have been around for a few years, but are just now being adopted by appreciable numbers of user organizations. For example, consider data warehouse appliances, competency centers, collaborative BI, columnar databases, data

federation, open source, in-database analytics, in-memory databases, master data management (MDM), real-time operation, and unstructured data.

TDWI sees user organizations adopting all ETMs steadily into the near future. But which ones are being adopted the most aggressively? Which are ready for prime time today? Which are too new to be pressing? Why should users care about ETMs?

To answer these questions, TDWI circulated a Technology Survey at the Orlando World Conference in November 2012. The survey presented a list of 30 ETMs and asked survey respondents to identify those they have no plans for using, those they are already using, and those they'll adopt within three years. Survey responses reveal which ETMs are of little interest today (at least, to survey respondents) versus those that are already in use or will be soon.

This article presents a few observations about the available ETMs for BI and their adoption rates, plus when and why user organizations should consider adopting them.

Which of the following ETMs is your organization using for business intelligence (BI), data warehousing (DW), or data management (DM)?

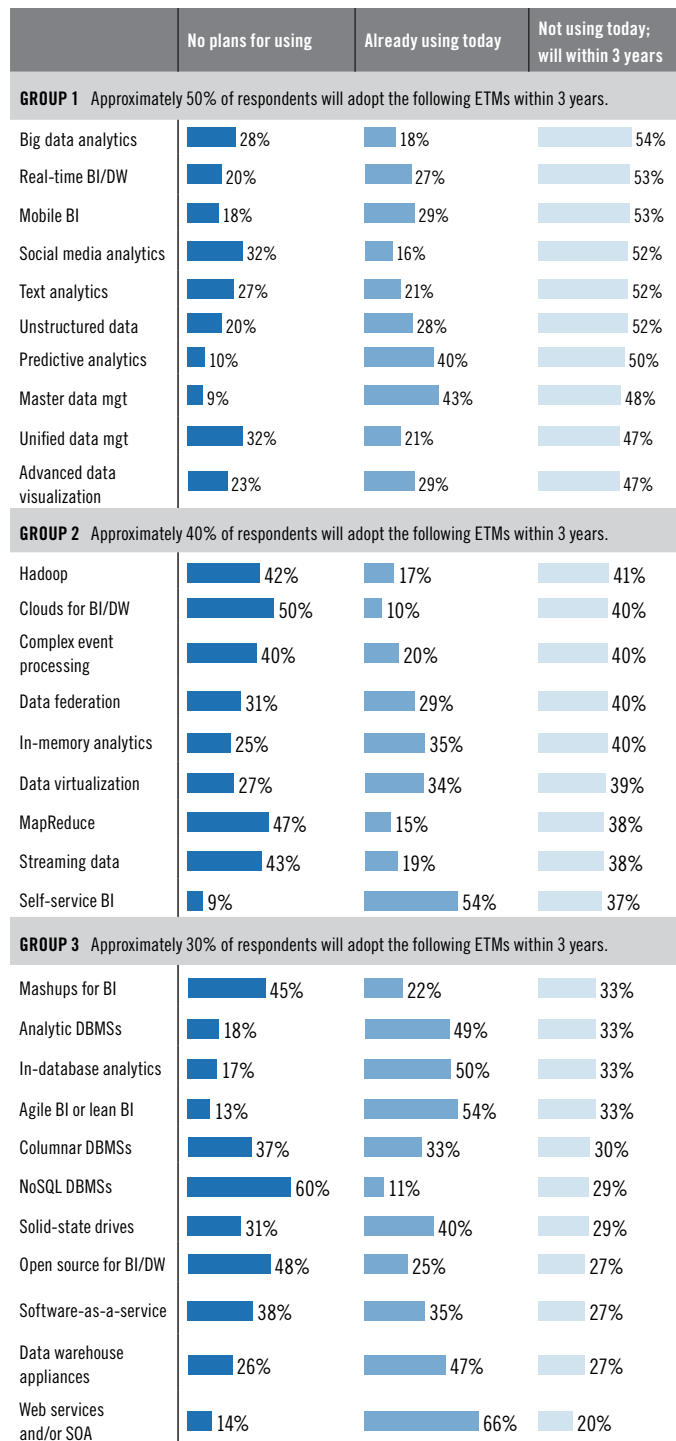


Figure 1. Based on 139 respondents in November 2012. Values in the table represent percentages of respondents. The table is sorted by the “Not using today; will within 3 years” column.

Top 10 Trends in ETMs for BI

- One-third of ETMs will see very aggressive adoption.**
 Let’s take a look into the near future by examining emerging technologies that users are not using today, but will soon. The ETMs in Group 1 in Figure 1 were each selected by approximately 50% of respondents as techniques they are not using today, but will be using within three years. The ETMs in Group 1 vary from very new techniques (big data analytics, text analytics, mobile BI, social BI) to techniques that have been with us for years, but are just now emerging in terms of brisk user adoption (real-time operation, MDM, and advanced data visualization).
- The newest ETMs are set for the most growth.**
 If you compare the percentage of survey respondents using an ETM today to the percentage for the same ETM in three years, the difference identifies a few ETMs that are poised for very dramatic growth. These are (in descending delta order) big data analytics (18% today; 54% in three years), social media analytics (16%; 52%), text analytics (21%; 52%), and clouds for BI/DW (10%; 40%).
- A few ETMs will be adopted by most organizations.**
 Very small percentages of survey respondents selected “no plans” for MDM (9%), self-service BI (9%), predictive analytics (10%), agile BI or lean BI (13%), and Web services or SOA (14%), which means that these are high priorities for most organizations.
- Not all ETMs are of interest to everyone.** In other words, some ETMs were selected by large percentages of respondents who have “no plans” for them, including NoSQL DBMSs (60%), clouds for BI/DW (50%), open source for BI/DW (48%), MapReduce (47%), and mashups for BI (45%).
- Some of the least used ETMs today will see appreciable adoption.** This is natural, given the newness of these technologies and the fact that it takes time for an ETM to move beyond its initial early adopters. In fact, the ETMs listed in the previous paragraph are about to make that move, as seen by comparing the percentage of survey respondents using each today to that in three years. For example, take another look at NoSQL DBMSs (11% today; 29% in three years), clouds for BI/DW (10%; 40%), MapReduce (15%; 38%), and mashups for BI (22%; 33%). These ETMs are poised to jump between 11 and

30 percentage points, according to survey results. Hence, if users' plans pan out, the ETMs just listed will go from rare to common in a mere three years.

6. **Some mature ETMs are already commonly used today.** A number of ETMs already have a large foot in the door, including Web services and SOA (66% in use today), agile BI and lean BI (54%), self-service BI (54%), in-database analytics (50%), analytic DBMSs (49%), and data warehouse appliances (47%). Since these mature ETMs are already somewhat saturated, they won't see as much growth as newer ones.
7. **It's not just technology; emerging methodologies are gaining, too.** This includes methods such as MDM, unified data management, self-service BI, and agile BI or lean BI.

USERS ARE EMBRACING EMERGING TECHNOLOGIES AND METHODS AGGRESSIVELY, AND THAT'S A GOOD THING.

8. **ETMs are helping organizations leverage more data types.** Users are digging deeper into big data analytics (54% in three years), text analytics (52%), unstructured data (52%), and master and reference data (48%).
9. **Many ETMs provide real-time operation for fast-paced business practices.** These include real-time BI/DW (53% in three years), event processing (40%), data federation (40%), and streaming data (38%).
10. **Users are planning aggressive moves into all things analytic.** This includes big data analytics (54% in three years), social media analytics (52%), text analytics (52%), predictive analytics (50%), in-memory analytics (40%), analytic DBMSs (33%), in-database analytics (33%), and so on. The focus on analytics among emerging technologies is not a surprise. User organizations have been deploying both new and old analytic technologies rather aggressively for about five years now.

Why Care about ETMs for BI?

Your peers in other organizations clearly see the importance of ETMs. The vast majority of survey respondents (82% in Figure 2) feel that emerging technologies and

methods (ETMs) are very important (59%) or somewhat important (33%).

ETMs have compelling benefits. They enable user organizations to address new business needs (73% in Figure 3), tap more options (56%), learn new skills (37%), and imagine new applications (32%).

ETMs have a few barriers, too. Embracing ETMs may be hindered by a lack of business value (73%), budget (58%), and skills (42%), as well as stodgy mindsets (26%).

How important do you think it is to embrace ETMs? Select only one.

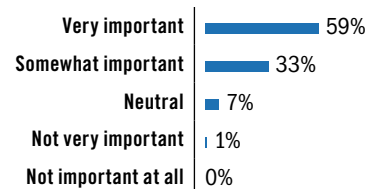


Figure 2. Based on 132 respondents in November 2012.

What are the benefits of embracing ETMs? Select all that apply.

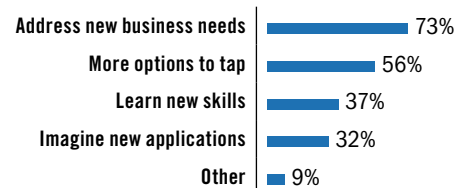


Figure 3. Based on 132 respondents in November 2012.

What are the barriers to embracing ETMs? Select all that apply.

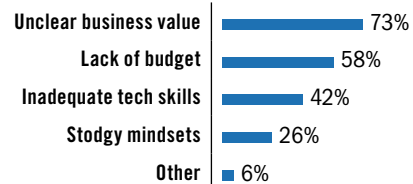


Figure 4. Based on 132 respondents in November 2012.

Conclusion

As seen in the results of the TDWI Technology Survey of November 2012, users are embracing emerging technologies and methods (ETMs) aggressively, and that's a good thing. After all, adopting ETMs is fundamental to gaining new insights via analytics (predictive analytics, data visualization, MapReduce), tapping new data sources (big data, social media, unstructured data, text analytics, Hadoop), learning new methodologies (agile BI, self-service BI, mobile BI, mashups), leveraging new platforms (clouds, analytic DBMSs, data warehouse appliances), and keeping pace with accelerating business operations (real-time BI/DW, streaming data, data federation, solid-state drives, event processing).

Philip Russom is director of TDWI Research for data management and oversees many of 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, @prussom on Twitter, and on LinkedIn at linkedin.com/in/philiprussom.

FIVE TRENDS FOR 2013: PEERING INTO THE FUTURE OF BUSINESS INTELLIGENCE

**DAVID STODDER, RESEARCH DIRECTOR FOR
BUSINESS INTELLIGENCE, TDWI**

Behind nearly every business intelligence (BI) tool implementation, there are strong ambitions to become a data-driven organization. Executives in business and IT want BI to support smarter decisions at all levels; they want to innovate by developing a thriving analytical culture that enables nontechnical users to discover insights that will help their organizations achieve competitive advantages.

To increase the business value of data and deliver that value to users sooner, many organizations are revamping traditional practices and asking more from their tools and technologies. BI, analytics, and data warehousing technologies must deliver more, but also be easier to use. In many cases, users need richer, more guided experiences that push IT to go beyond data access and fielding queries. Yet users also desire the freedom and flexibility to create their own experiences with data.

Here are five major trends that I see shaping BI and analytics practices and technologies as we head into 2013. Many build on innovations that I identified as emerging in 2012, but that are now maturing as organizations gain experience with BI, data discovery, and analytics technologies.

1. **The focus of self-service BI and analytics shifts to enabling better-managed and more context-sensitive experiences.** “Self service” remains one of the hottest trends today in BI and analytics. Personal, self-directed tools are available that give users more control over how they view and access data, create reports and data visualizations, and share insights to suit their roles and responsibilities. BI and analytic data discovery technologies are becoming easier to deploy and configure, and many take advantage of in-memory computing to enable robust work environments for performing serious analysis without involving IT at every step, as has been traditionally necessary. In addition, when users are able to do more for themselves, it can relieve pressure on IT's application backlog.

However, there is a fine line between user freedom with self-service technologies and a chaotic, expensive, and risky environment where users

are “going rogue” with their own systems. This sort of chaos is unfortunately the norm in many organizations. IT has historically had a difficult time putting constraints on users’ consumption of data in spreadsheets, including the loading of data from enterprise BI systems into “spreadmarts” that are not governed by IT. Business departments are known to set up shadow IT projects for running analytics on standalone data marts, or if cloud-based services are available, in analytic sandboxes that may be located on systems beyond IT’s management oversight, firewalls, and regulatory compliance procedures.

TDWI Research finds that this sort of chaos is one of the top three reasons organizations are implementing self-service BI and analytics technologies (see Figure 1). Thus, rather than trying to restrain users, many CIOs and IT managers today would prefer to let users choose their own BI and analytics tools but then work through organizations, such as BI competency centers, to achieve a better balance between user freedom and proper management. Self-service BI delivers greater benefits when the data supply chain is reliable, well organized, and expanding in a careful fashion to include new sources such as Hadoop files or external data services. Many tools in the market are maturing in ways that support more balanced management—certainly better than what exists with spreadmarts.

In 2013, along with more managed approaches to self-service, TDWI expects to see greater emphasis on placing self-directed reporting and analytics in the context of the users’ roles and responsibilities. What kind of visibility into data do users in specific

roles need? How do users with different types of responsibilities collaborate on decisions? Do they share insights over e-mail, in presentations, through contextual notes, or in social media? By understanding the context within which users perform BI reporting and analytics, organizations will be in a better position to create the right balance between data governance and user freedom.

FEW WOULD DISAGREE THAT IN NEARLY EVERY INDUSTRY, SPEED IS A COMPETITIVE ADVANTAGE.

- 2. **The pursuit of faster decision cycles begins to reshape deployment of BI and analytics systems.** Few would disagree that in nearly every industry, speed is a competitive advantage. Being first to market with products and services can be a sure path to success if organizations are able to back up their introduction with well-aligned business processes in customer service, marketing, product fulfillment, and more. To create this alignment, organizations need an excellent information flow that enables managers across enterprises to analyze data and share insights sooner and more frequently. Decision cycles, or the processes involved in getting from the beginning to the end of a strategic or tactical decision, must run faster to enable organizations to make speed a competitive advantage and take intelligent actions at the right moments.

TDWI Research finds that “shorter decision cycles”

What are your organization's main reasons for implementing self-service BI and analytics?

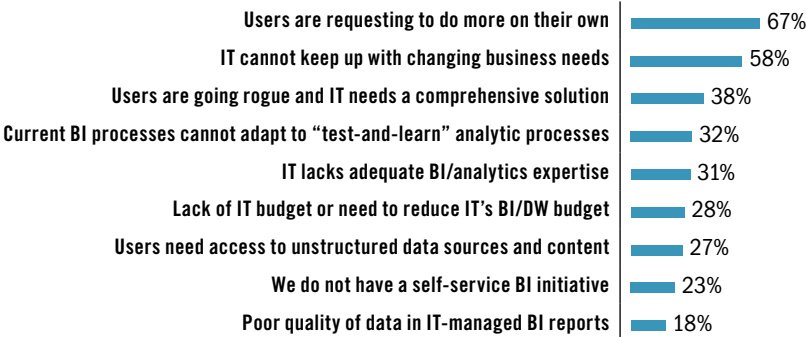


Figure 1. Source: TDWI, January 2013. Based on answers from 377 respondents; respondents could select more than one answer.

form one of the most disruptive factors facing organizations. It is unclear whether traditional BI and data warehousing development methods and systems can effectively serve the need for ever-faster decision making. These systems often came into being when decision cycles were slower, and were set up for batch historical reporting. Management decisions driven by traditional BI and data warehousing are largely human centered. Designers of most systems did not anticipate the growing use of algorithms to automate decisions. Increasingly, as Marc Demarest, CEO and principal of Noumenal, Inc., put it in his keynote at the TDWI Orlando World Conference last year, “If a decision is algorithmic, it belongs in code.”

Automating decisions often falls under the category of “decision management.” This field integrates elements of BI, predictive analytics, activity monitoring, business rules, and business process management with a focus on creating algorithms to automate decisions. Complex event processing and data sensors will “sniff” for trends, anomalies, or changes in the behavior of factors that are important to investments, credit approvals, Web-based marketing, manufacturing processes, supply chains, and more. Decision management systems speed up decision cycles through automation.

Organizations that are either testing or launching decision management systems are investigating how they can automate decisions that may currently be undertaken through traditional, human-centered reporting and analysis. They also need to establish ways of integrating the information flow coming to and from automated systems with their problem-solving analytics and reporting for decisions that must remain in human hands. In the next year, we will likely see vendors attempt to fill the need for data “supply-chain” products that can bridge the gap between human-centered and automated, real-time decision systems.

3. **Users grow to expect workspaces that integrate different types of data and methods of access and analysis.** As users gain experience with analytics and grow to depend on data to drive their decisions, many will seek to increase the depth and breadth of their view by reaching out to sources existing beyond the data warehouse. This data mix will include sources that are raw, structured, aggregate, dimensional, textual, semi-structured, and unstructured.

Spreadsheet data will continue to be part of the mix. In most organizations, however, the different types of data are encased in application, content, or database silos. Many of these sources are hard to access and integrate because tools generally are specialized for certain types of data or schema. Direct access to other sources such as operational systems may be restricted by IT because of performance degradation concerns. IT will be under increasing pressure from users to enable single views of information drawn from a widening array of sources. Business and IT users will seek to deploy tools and systems that can offer unified information access (UIA) to structured, semi-structured, and unstructured information by integrating search, query, and analytics specialized for different forms of data. Interest will also increase in creating master data management, global business glossaries, and other metadata classification systems to make it easier to gain a unified view of information and discover metadata and data relationships across sources. In addition, organizations will enable heterogeneous data access by deploying prebuilt adapters to different data sources.

IN 2013, WE WILL SEE WHICH WAY FORWARD FOR BI SYSTEMS GAINS THE MOST TRACTION.

Business intelligence systems are therefore at a crossroads: Do they remain focused on accessing structured data, but perhaps with the ability to tap a wider selection of these relational sources? Or do they evolve into UIA workspaces that mix access, analysis, and visualization of many different types of data? TDWI Research finds that most organizations report significant difficulty in adding or integrating unstructured data into their BI and data warehousing systems. Thus, interest in solving this problem is high, but solutions are not yet adequate. In 2013, we will see which way forward for BI systems gains the most traction.

4. **Data visualization moves beyond static images to provide improved data interaction and context inside narratives.** Data visualization is hot. Dashboards, already a common fixture on most BI tool users’ screens, are the stage for increasingly sophisticated

visualizations that go beyond simple charts and tables to include heat maps, scatter plots, and more. The rapid adoption of mobile devices is likely to increase the trend toward easy-to-use BI dashboards that users can access through either desktop or mobile platforms. Advanced data visualizations that can handle higher data volumes, frequent data updates, and changing views involving multiple dimensions are becoming critical to expert users in nearly every industry, including energy, financial services, gaming, fraud detection, law enforcement, and intelligence.

Providing more engaging and exciting ways of seeing information, however, is not enough to make data visualization a game changer. First, the charts, diagrams, and other representations must allow for interactivity. Early versions of dashboards were often static, providing only limited drill-down capabilities; more mature data portals have enabled users to dig deeper to perform what-if analysis and simulations to see how the visualizations change if they choose different variables or alter forecasts. With more advanced tools, users are able to fine-tune visual reporting and analysis on their own rather than having to seek out IT professionals who can do modifications for them. These tools guide users in selecting the best data visualizations rather than leaving them to choose on their own, like kids in a candy store.

Second, data visualizations need to be organized and shared within context. Advanced visualization tools can integrate the representations with business processes to provide real-time views of trends and conditions so that nontechnical personnel have actionable information. Some visualization systems exploit high-performance computing options, such as in-memory, in-database, and parallel computing to support deeper, more powerful visual analysis. Leading-edge personal BI and data discovery tools are introducing the concept of narratives and “storytelling” to help users assemble their visual analysis and reporting within larger business or analytical objectives to improve collaboration. In 2013, look for these improvements to push data visualization beyond “eye candy” to become essential to communicating with data and taking action.

5. **Organizations improve agility by revamping development and letting users determine the data's relevance.**

Impatient with the length of time it can take to gather users' data requirements, and the need to gather them all over again when requirements change, leading organizations are revamping how they develop BI and data warehousing systems. Although still in the early stages, an increasing number of organizations are adopting agile software development methods for BI and data warehousing projects (see Figure 2). These methods can enable delivery of business value sooner, in smaller increments, and with more flexibility in how systems are developed. Agile methods encourage greater involvement by users so that as requirements change they can be captured and brought into development cycles immediately.

AGILE METHOD ADOPTION AND THE USE OF BIG DATA TECHNOLOGIES SUCH AS HADOOP ARE LIKELY TO INCREASE IN 2013.

However, another key trend, sometimes running alongside agile method adoption, is to put data in front of users much earlier, before it is extracted, transformed, and loaded into a data warehouse. Users work with IT professionals to determine the value and relevance of this raw data for their needs. This practice is particularly useful for big data; organizations are making Hadoop files available to users, who can then dive into raw data, apply search or analytic algorithms to find important data or patterns, and then decide whether the data merits formal processes for loading it into a data warehouse, cleansing and integrating it there with data from other sources, and making it available for ongoing BI query and analysis.

As organizations seek to derive game-changing insights from bigger volumes of data, they need to shorten the path to business value. Traditional BI, ETL, and data warehousing practices are still critical, but organizations are changing them so that users can derive value sooner and IT can avoid wasting resources on working up irrelevant data. Agile method adoption and the use of big data technologies such as Hadoop for earlier access to data are likely to increase in 2013 as organizations seek more agile and flexible approaches to discovering and sharing big data insights.

How many agile development projects is your BI/DW organization currently running or has previously run?

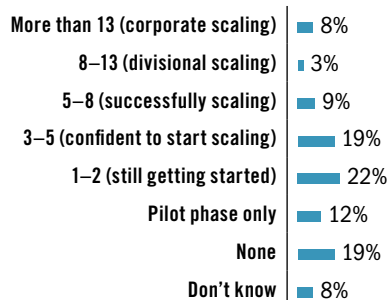


Figure 2. Source: TDWI, January 2013. Based on answers from 394 respondents. Research survey and analysis conducted jointly by TDWI and Ceregenics.

Speed and Flexibility: BI Imperatives

Taken together, these five trends reflect needs that organizations have for technologies that enable decision makers to respond effectively to adjust strategies and tactics as marketplaces change, customer priorities shift, and other external factors, such as new regulations and policies, take effect. Canned reports and limited, static views of data will be less and less acceptable—they do not help decision makers understand what is happening so that they can plan intelligently for the future.

User expectations for greater speed to insight and flexibility in BI, analytics, and data warehousing systems create challenging problems, but not ones without solutions. The new year will demand creativity and courage from BI/DW professionals. Exciting times lie ahead.

David Stodder is director of TDWI Research for business intelligence. He focuses on providing research-based insight and best practices for organizations implementing BI, analytics, performance management, data discovery, data visualization, 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, or follow him on Twitter: @dbstodder.



Experience the Power of AGILE BUSINESS ANALYTICS

Discover how you can quickly and easily analyze data across sales, financial and operational systems and build dashboards and reports to provide real-time, actionable insight to drive your business.

**TRY FOR
FREE**



Test drive Birst Express in a secure, cloud-based environment at no charge.

GO TO WWW.BIRST.COM/BEST

Visit us at www.birst.com and follow us on Twitter @BirstBI



TDWI RESEARCH Applying Technologies for Social Media Data Analysis

BY DAVID STODDER

TDWI Research found that “gaining deeper customer understanding” (56%) is the top objective that organizations seek to achieve from implementing customer analytics with social media data (see Figure 1). This is particularly the case for organizations that do not have good internal customer transaction or service record data sources and have little history of collecting customer satisfaction or other behavioral information. Social media listening can provide an unprecedented window on customer sentiment and the reception of an organization’s marketing, brands, and services. (As a reality check, 32% of organizations surveyed do not analyze social media data.)

Beneath the broad objective of gaining a deeper customer understanding, we can see in Figure 1 the relative importance of several tactical objectives. Nearly one-third (31%) seek to identify attribution, or paths to buying decisions, a topic that was discussed in the previous section. As they mature in their attribution inquiry and wish to examine much larger sets of detailed or unstructured data, organizations reach the limits of

Which of the following objectives does your organization seek to achieve by implementing customer analytics technologies and methods with social media data? (Please select all that apply.)

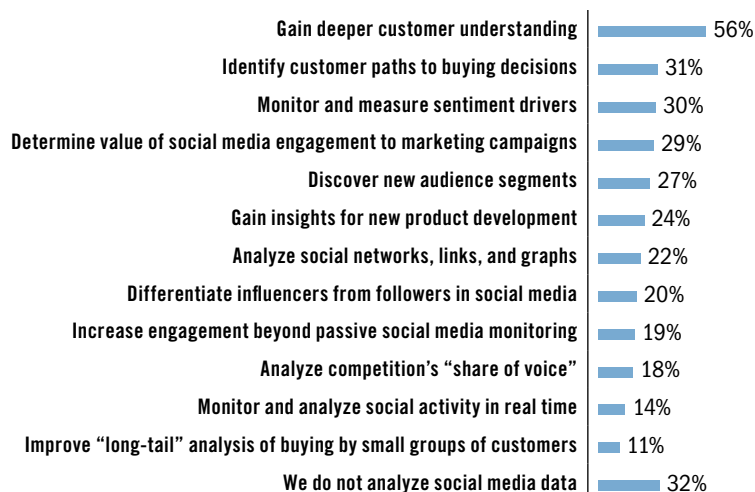


Figure 1. Based on 1,546 responses from 418 respondents; a bit more than three responses per respondent, on average.

available services such as Google Analytics and website analysis applications. These typically provide only an aggregated view of data. Organizations need detailed data to support analysis of how marketing and engagement processes can improve personalized interactions with smaller customer segments or even individuals. Organizations may want to examine purchase paths and attribution across the multiple online channels as well as offline sources. Robust database and customer analytics systems are frequently necessary to handle the big data needed for this more complex analysis.

Discovering Customer Sentiment. About the same percentage (30%) of respondents seeks to monitor and measure sentiment drivers. Sentiment analysis enables organizations to discover positive and negative comments in social media, customer comment and review sites, and similar sources. Sentiment analysis often focuses on monitoring and measuring the “buzz” value, usually through volume and frequency of comments around a topic. However, many organizations want more analytical depth so that they can discover what the buzz is about, where it originated, and who is benefiting or not benefiting from it.

For more sophisticated sentiment analysis, text analytics tools play a big role. These tools employ lexicons, word extraction, natural language processing, pattern matching, and other approaches to examine social media users’ expressions. Sentiment analysis can give organizations

early notice in real time of factors that may be affecting customer churn; in Figure 1, the research shows that 14% are interested in monitoring and analyzing social activity in real time. Sentiment analysis is also important to understanding competitors’ relative strengths and weaknesses in the social sphere. Our research found that 18% of respondents are examining social media data to analyze the competition’s “share of voice.”

Applying Analytics to Find and Influence the Influencers

One of the biggest challenges can be simply deciding which social media sites’ data to analyze. Organizations have to research where their customers are most likely to express themselves about brands and products. They need to spot influencers who have networks of contacts and take it upon themselves to play an advocacy role. About 20% of respondents are interested in differentiating influencers from followers in social media (see Figure 1). Link analytics tools and methods specialize in identifying relationships between users in social communities and enabling organizations to measure users’ influence. With some tools, data scientists and analysts can test variables to help identify social communities as “segments.” Then, as they implement segmentation models for other data sources, they can integrate these insights with social media network analysis to sharpen the models and test new variables.

Which of the following social media data sources is your organization currently accessing or planning to access for customer analytics, brand, or marketing management purposes?

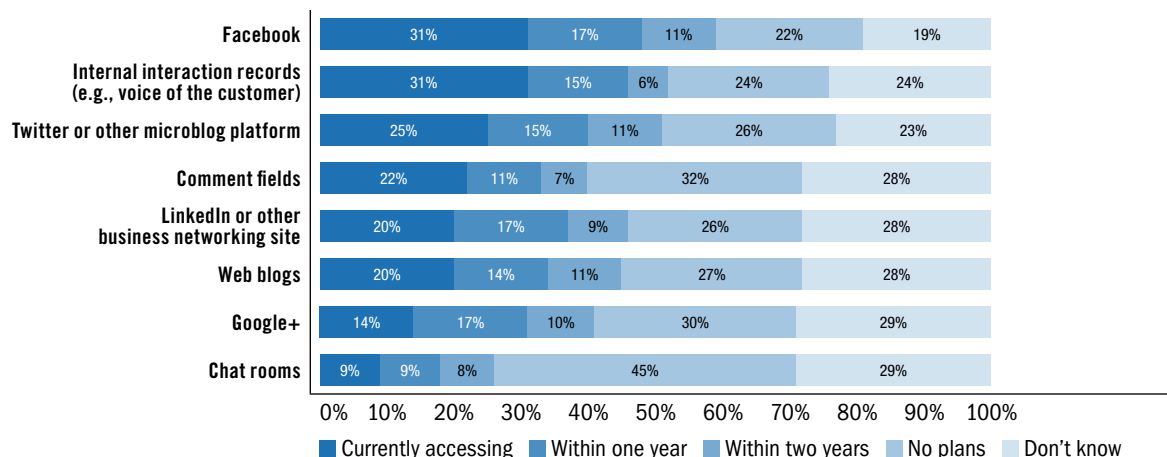


Figure 2. Based on 3,242 responses from 404 respondents; about eight responses per respondent, on average.

Analytics are critical for enabling organizations to make the right decisions about when, where, and how to participate in social media. It isn't enough to just listen; organizations must insert themselves and become part of the conversation. Leading companies will start viral campaigns, for example, using Twitter hashtags for a topic; the campaign could be a component of a larger marketing strategy. They can then monitor social media to see what people say and analyze how the campaign is playing among influencers and across networks. "My belief is that the sweet spot for social media is not conversion, but nurturing," said Brian Ellefritz, vice president of global social media at SAP. "Whether it's in your community, through Twitter, or through Facebook pages, you want to build an increasing conviction that your company is the one to do business with. It's about establishing a belief system that becomes robust with the support of fans and followers. The question is how you measure that and create value out of that investment."

Social media data analysis can also be a key component of risk mitigation in the marketplace. Pharmaceutical companies, for example, need to follow social media to watch for early signs of negative consequences with drugs or other treatments. How participatory these organizations and healthcare providers can be is a sensitive issue because of regulatory reporting requirements. Social media conversations fall into a gray area from a regulatory perspective, which has made many firms keep their distance other than for advertising. Overcoming

regulatory concerns in the healthcare industry could enable greater participation in social media by these firms.

Selecting and Accessing Internal and External Social Media Data

The top social media data sources that respondents in our research are currently accessing are internal interaction records such as voice of the customer (VOC) logs (31%; see Figure 2). Another 15% of respondents plan to access these data sources within one year. This suggests both the understandable immaturity of organizations' pursuit of social media data sources for analysis as well as a desire to apply advanced analytics tools and methods to internal customer information sources that they may view as more important.

Text analytics tools, discussed earlier in the context of sentiment analysis, are established for analysis of internal content such as VOC records. These records are often poorly exploited because they are left to either agents or overtaxed data analysts to examine manually. Text analytics can automate discovery and provide greater analytical consistency and depth. Some organizations view social media data analysis as a kind of extension of VOC analysis, except that customers' feedback and comments are less structured and are expressed outside the constraints of forms or questionnaires that require direct responses.

Overcoming Challenges to Accessing Social Media Network Data

TDWI Research found that Facebook (31%) and Twitter (25%) are the most common social media data sources that respondents currently access (referring again to Figure 2). LinkedIn (20%) and the relatively new Google+ (14%) are next. Another 17% plan to access Facebook within one year, and 15% plan to access Twitter data. Regarding LinkedIn and Google+ sources, an additional 17% of respondents plan to access each of these sources within one year.

Facebook and Twitter both present challenges for social media data analysis. The “big pipe” of data each site produces is no longer available to all analysts and developers. Facebook has many controls built into its application programming interface (API) to guard the privacy of its users and limit data access to its own proprietary use. Thus, external parties are essentially limited to what they can scrape from Facebook page descriptions. Except for fans’ “likes” and comments recorded on a company’s own Facebook pages, Facebook’s data richness is mostly kept inside for its own advertising and behavior analysis.

Twitter’s data fire hose consists of all public tweets, adding up by some estimates to more than 350 million tweets per day. This complete access is now available only through Twitter’s partners. One such partner is Gnip, which for a price provides full Twitter streams through monitoring, BI, and discovery analytics tools such as Tableau. Gnip is strictly for performing analytics rather than for displaying tweets. Through a partnership with the Web development

firm Automattic, Gnip can access content on Wordpress and Jetpack blogging and comment platforms, which enables Gnip to integrate these sources with Twitter data for social media analysis. DataSift is another Twitter partner with full access. Organizations can also work with specialized analytics services providers such as Dataminr that are Twitter partners. Organizations that are not partners of Twitter and wish to develop their own analytic applications are limited to the Twitter streaming API, which provides near-real-time access to small subsets drawn from the tweet fire hose.

David Stodder is director of TDWI Research for business intelligence. He focuses on providing research-based insight and best practices for organizations implementing BI, analytics, performance management, data discovery, data visualization, 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, or follow him on Twitter: @dbstodder.

This report was sponsored by Greenplum, a division of EMC; IBM; Informatica; SAP; SAS; Tableau Software; Teradata (including Teradata Aster and Aprimo); and Vertica, an HP Company.



TURNING BIG DATA INTO INTELLIGENT ACTION



ONLY VITRIA OPERATIONAL INTELLIGENCE (OI) PROVIDES:

- ▶ **Continuous Real-time Analytics™** on Big Data in motion & at rest
- ▶ **Elastically Scalable Operational Intelligence** on premise or in the cloud
- ▶ **Advanced Process Intelligence** for a 360° view of your business in real time

VITRIA OI PROVIDES REAL-TIME DATA & INTELLIGENCE FOR:

- ▶ Telecommunications
- ▶ Smart Energy
- ▶ Financial Services
- ▶ Manufacturing & Logistics
- ▶ Government & Cyber Security
- ▶ Healthcare & Pharmaceuticals

LEARN MORE ABOUT OI FOR BIG DATA ANALYTICS AT:
WWW.VITRIA.COM/BESTOFBI



TDWI RESEARCH

Introduction to High-Performance Data Warehousing

BY PHILIP RUSSOM

Performance Is about Speed and Scale, Complexity and Concurrency

High performance continues to intensify as a critical success factor for user implementations in business intelligence (BI), data warehousing (DW), data integration (DI), and advanced analytics. That's because just about everything we do in BI, DW, DI, and analytics nowadays has some kind of high-performance requirement, for both business and technology reasons. Yet, users are challenged by big data volumes, new and demanding analytic workloads, growing user communities, business requirements for real-time operation, and more.

In most user organizations, a DW and similar databases bear much of the burden of performance; yet, the quest for speed and scale also applies to every layer of the complex BI/DW/DI and analytics technology stack, as well as processes that unfold across multiple layers. Hence, in this report, the term *high-performance data warehousing* (HiPer DW) encompasses performance characteristics, issues, and enablers across the entire technology stack and associated practices.

High-performance data warehousing is primarily about achieving speed and scale while also coping with increasing complexity and concurrency. These are the four dimensions that define HiPer DW. Each dimension can be a goal unto itself; yet, the four are related. For example, scaling up may require speed, and complexity and concurrency tend to inhibit speed and scale.

HiPer DW's four dimensions are summarized in Figure 1. Here are a few examples of each:

Speed. The now-common practice of operational BI usually involves fetching and presenting operational data (typically from ERP and CRM applications) in real time or close to it. Just as operational BI has pushed many organizations closer and closer to real-time operation, the emerging practice of operational analytics will do the same for a variety of analytic methods. Many analytic methods are based on SQL, making the speed of query response more urgent than ever. Other analytic methods are even more challenging for performance due to iterative analytic operations for variable selection and reduction, binning, and neural net construction. Out on the leading edge, events and some forms of big data stream from Web servers, transactional systems, media feeds, robotics, and sensors; an increasing number of user organizations are

now capturing and analyzing these streams, then making decisions or taking actions within minutes or hours.

Scale. Upon hearing the term “scalability,” most of us immediately think of the burgeoning data volumes we’ve been experiencing since the 1990s. Data volumes have recently spiked in the phenomenon known as “big data,” which forces organizations to manage tens of terabytes—sometimes hundreds of terabytes, even petabytes—of detailed source data of varying types. But it’s not just data volumes and the databases that manage them. Scalability is also required of BI platforms that now support thousands of users, along with their thousands of reports that must be refreshed. Nor is it just a matter of scaling up; all kinds of platforms must scale out into ever larger grids, clusters, clouds, and other distributed architectures.

HiPer DW's Four Dimensions

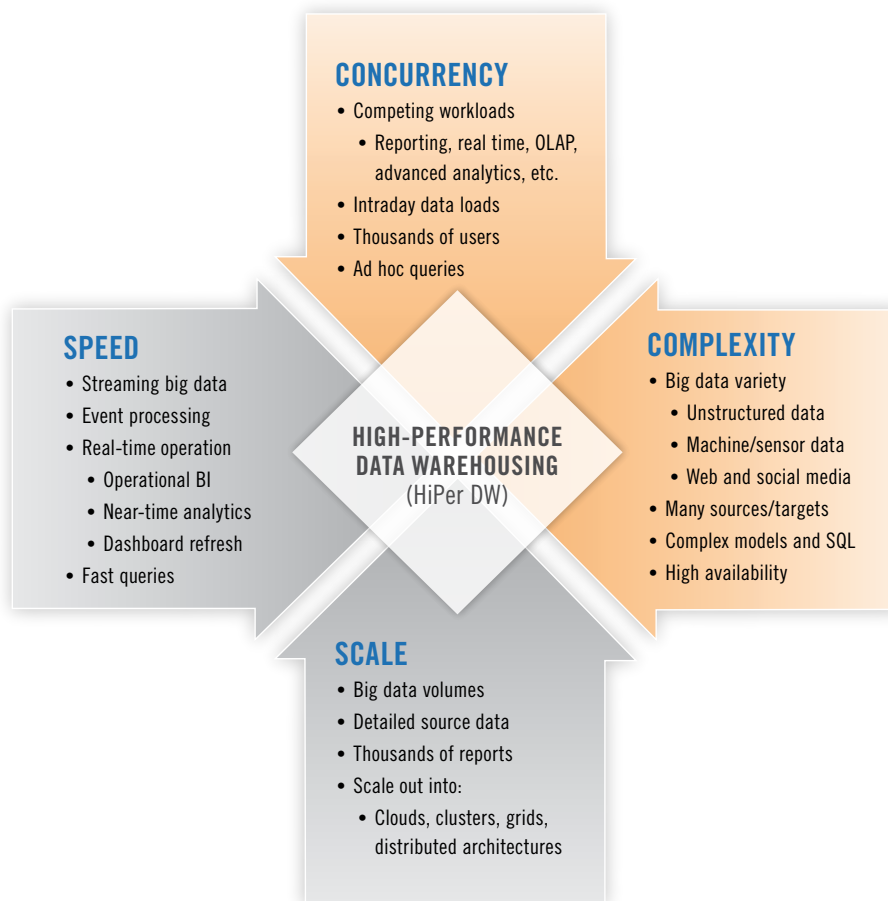


Figure 1. HiPer DW's four dimensions are speed and scale, plus complexity and concurrency.

Complexity. Complexity has increased steadily with the addition of more data sources and targets, not to mention more tables, dimensions, and hierarchies within DWs. Today, complexity is accelerating as more user organizations embrace the diversity of big data, with its unstructured data, semi-structured data, and machine data. As data's diversity increases, so does the complexity of its management and processing. Some organizations are ensuring high performance for some workloads (especially real time and advanced analytics) by deploying standalone systems; one of the trade-offs is that the resulting distributed DW architecture has complexity that makes it difficult to optimize the performance of processes that run across multiple platforms.

Concurrency. As we scale up to more analytic applications and more BI users, an increasing number of them are concurrent—that is, using the BI/DW/DI and analytics technology stack simultaneously. In a similar trend, the average EDW now supports more database workloads—more often running concurrently—than ever before, driven up by the growth of real-time operation, event processing, advanced analytics, and multi-structured data.

High Performance: Problem or Opportunity?

In recent years, TDWI has seen many user organizations adopt new vendor platforms and user best practices that helped them overcome some of the performance issues that have dogged them for years, especially data volume scalability and real-time data movement for operational BI. With that progress in mind, a TDWI survey asked: “Across your organization, is high performance for DW, BI,

and analytics considered mostly a problem or mostly an opportunity?” (See Figure 2.)

Two-thirds (64%) consider high performance an opportunity.

This positive assessment isn't surprising, given the success of real-time practices such as operational BI. Similarly, many user organizations have turned the corner on big data—no longer struggling to merely manage it, but instead leveraging its valuable information through exploratory or predictive analytics to discover new facts about customers, markets, partners, costs, and operations.

Only one-third (36%) consider high performance a problem.

Unfortunately, some organizations still struggle to meet user expectations and service-level agreements for queries, cubes, reports, and analytic workloads. Data volume alone is a showstopper for some organizations. Common performance bottlenecks center on loading large data volumes into a data warehouse, running reports that involve complex table joins, and presenting time-sensitive data to business managers.

HiPer DW Solutions Combine Vendor Functionality with User Optimizations

Speed, scale, complexity, and concurrency (in that priority order) are compelling goals for high performance, but they are challenging to achieve. Luckily, many of today's high-performance challenges are addressed by technical advancements in vendor tools and platforms.

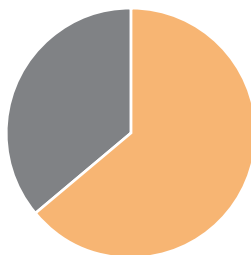
For example, there are now multiple high-performance platform architectures available for data warehouses, including massively parallel processing (MPP), grids, clusters, server virtualization, clouds, and SaaS. For real-

Across your organization, is high performance for DW, BI, and analytics considered mostly a problem or mostly an opportunity?

PROBLEM

because it's hard to achieve from a technical viewpoint

36%



64%

OPPORTUNITY

because it enables new, broader, and faster business practices

Figure 2. Based on 278 respondents.

time data, databases and data integration tools are now much better at handling streaming big data, service buses, SOA, Web services, data federation, virtualization, and event processing. Meanwhile, 64-bit computing has fueled an explosion of in-memory databases and in-memory analytic processing in user solutions; flash memory and solid-state drives will soon fuel even more innovative practices. Other performance enhancements have recently come from multi-core CPUs, appliances, columnar storage, high-availability features, Hadoop, MapReduce, and in-database analytics. Later sections of this report will discuss in detail how these and other innovations assist with high performance.

Vendor tools and platforms are indispensable, but HiPer DW still requires a fair amount of optimization by technical users. The best optimizations are those that are designed into the BI and analytic deliverables that users produce, such as queries, reports, data models, analytic models, interfaces, and jobs for extract, transform, and load (ETL). As we'll see later in this report, successful user organizations have predetermined standards, styles sheets, architectures, and designs that foster high performance and other desirable characteristics. Vendor tools and user standards together solve many performance problems up front, but there is still a need for the tactical tweaking and tuning of user-built BI deliverables and analytic applications. Hence, team members with skills in SQL tuning and model tweaking remain valuable.

USER STORY ACHIEVING HIGH PERFORMANCE CAN INVOLVE A SERIES OF RELATED OPTIMIZATIONS.

"A lot of my data warehouse performance problems stem from the fact that the warehouse is an operational system, as well as a decision-making platform," said the BI director at a home improvement retailer. "For example, managers at headquarters and store managers alike need to run reports with complex queries against low-level detailed data that identifies products by SKU [stock-keeping unit]. One reason is so that store managers can look into a truck manifest—in a real-time, self-service way—to see exactly what's heading to them today. That can be useful when they have an in-store customer who wants to buy a product that's out of stock. A similar reason is to run the Lost Sale Report. That shows which products were out of stock in which stores recently, thereby resulting in a lost sale. Based on the report, merchandising managers can move inventory around to keep sales brisk in the stores.

"We've put a lot of effort into optimizing the data models of consolidated detailed source data, and now all those

complicated SKU-based reports perform very well. We've licked that performance problem, but we still have others to work on.

"Most of the SKU data comes from ERP systems that—because of their obtuse data models—require the data to be staged in the ERP environment before extraction. Believe it or not, ERP staging takes more time in our nightly window than data extraction, warehouse loading, and report refresh combined. We're putting the final touches on a new high-performance ETL implementation that will go around the staging requirement and extract ERP data more directly. That same implementation also speeds up data warehouse loading. When the ETL implementation is in place, the new speed of ERP extraction and warehouse load will open up two-thirds of our nightly window. We'll then fill that opening with new extraction jobs that satisfy new business requirements."

Philip Russom is director of TDWI Research for data management and oversees many of 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, [@prussom](https://twitter.com/prussom) on Twitter, and on LinkedIn at [linkedin.com/in/philiprussom](https://www.linkedin.com/in/philiprussom).

This report was sponsored by Cloudera, IBM, Oracle, ParAccel, SAP, SAS, Teradata, and Vertica.



Analytics for the Agile Business

Lavastorm Analytics provides a new agile way to analyze, optimize and control data and processes, enabling you to discover new business value in your data with visual tools that provide you with more power, agility, insight and speed than your current approach, with less dependence on IT. Test and debug your analytic model, explain it to non-technical managers and executives, easily publish your results, and when appropriate, create your own set of persistent business controls and continuous audits that deliver on-going value, putting you firmly in control.



“Lavastorm is a Top Company to Watch in 2013”

Robin Bloor, Chief Analyst, Bloor Group

Get Started Today - Download the Free Lavastorm Public
<http://www.lavastorm.com/tdwi>



TEN MISTAKES TO AVOID

When Validating Your BI/DW Direction

BY JONATHAN G. GEIGER

FOREWORD

It is often said that “If it ain’t broke, don’t fix it,” but how do you know if your business intelligence/data warehouse (BI/DW) program isn’t broken? Even if it meets some business needs and provides business users with data to support analyses and decision making, there are always opportunities for improvement. Companies should periodically review where they’ve been with their BI/DW program, and more important, where they’re heading. However, exploring your BI/DW direction without a well-thought-out approach and ignoring critical areas may lead to inaccurate conclusions about the environment—and suboptimal improvements.



ONLY FOCUSING ON MEETING CURRENT REQUIREMENTS

A natural approach to validating your BI/DW direction is to examine what is currently being done about your requirements. It is more important to know what is *not* being done.

BI/DW products should be reviewed and compared to expectations that were set as the environment evolved. However, even if the products meet the requirements, business expectations *may not be met*. Business analysts often uncover additional needs and reset their expectations. Both the original and new expectations should be evaluated, as should the process by which expectations and requirements are captured.

A potentially more important area is to assess ways in which the BI/DW environment can be leveraged beyond the original and evolving requirements. It’s extremely important to determine if the existing data and capabilities could be applied in other ways to provide significant value. For example, are there strategic enterprise goals not being addressed by the environment even though the data has been captured?

By validating both the current and potential use of their BI/DW environment, enterprises should be able to increase their business value and return on investment.



ASSUMING THE ARCHITECTURE AND METHODOLOGY ARE SOUND

As companies embark on their BI/DW journeys, they fret over which architecture should be used. The two leading architectures are hub-and-spoke and bus. There are two major reasons to validate your architecture for an ongoing program:

1. Assuming the architectural decision is still valid, organizations should examine the degree to which their existing architecture complies with best practices for that architecture. It is not uncommon to have an architecture morph over time to the point that it is inconsistent with the originally conceived direction. If that is the case, then it's important to recognize whether there are any deficiencies that can be traced back to departures from the original architecture.
2. The original architecture may not be appropriate for evolving needs. For example, the architecture may have been based on satisfying strategic needs and now the volumes have exploded, or the company is moving to an operational BI environment. The shift may dictate incorporating a messaging bus, EII technologies, an appliance with data marts deployed as views, columnar databases, and so on.

The methodology needs to be program oriented and consistent with the architecture. As the program evolves, the methodology must evolve as well. The validation effort should examine the extent to which the methodology continues to meet the BI/DW demands, roles of the business and IT participants, and duration of incremental enhancements to the environment.

By validating both the degree to which the environment may have departed from the original architecture and methodology and the degree to which these can meet current and anticipated needs, you can pursue appropriate steps to ensure that the environment is built on a strong foundation.



IGNORING THE BUSINESS PERSPECTIVE

The data warehousing team is often located within the IT department,

and when it decides to evaluate the company's BI/DW direction, it may only focus on the technical aspects. Although it is important to evaluate these aspects (see Mistake Four), an efficient environment does not necessarily provide business value. To gain a complete picture of your BI/DW direction, it is critical to evaluate it from a business perspective.

Ask these questions:

- What business value is being delivered?
- What additional business value could be delivered?
- How satisfied are the business users?
- How are people using the environment?
- What other sources are business people using to obtain analysis and decision support data?

Answers to questions such as these provide valuable information that can be used to determine what's working and what's not. In addition, involving the business in the process improves the credibility of the findings and recommendations.

By validating your BI/DW direction from a business perspective, both strengths and weaknesses will be identified and can be leveraged to determine the best direction for moving forward.



IGNORING THE TECHNICAL PERSPECTIVE

Focusing on just the business perspective is not enough. There are many elements within a sophisticated BI/DW environment, so it is equally important to validate the direction from a technical perspective. The technical perspective includes the overarching architecture (previously discussed), tools, technology, service-level satisfaction, reliability, methodology, roles and responsibilities, and so on.

Ask these questions:

- To what extent are documented and undocumented service-level objectives being satisfied?
- What is the mix of technologies being used? Will these technologies meet future needs?
- What tools are needed? Which ones are being used? How well are they being supported?

- Will the infrastructure support a mobile decision-making workforce?
- How well staffed is the development and support group? What is their skill level relative to what is needed?

Answers to questions such as these provide valuable information that can be used to determine what's working and what's not from a technical perspective.

Validating your BI/DW direction from a technical perspective allows both strengths and weaknesses to be identified. These can be leveraged to ensure appropriate technical support for moving forward.



IGNORING OTHER INITIATIVES

The BI/DW environment does not exist by itself. It's part of a greater ecosystem that includes many systems and data stores, and at any point in time there are several projects (e.g., ERP or MDM) planned and under way, including some related to the data warehouse. Often, some of these other projects entail system replacements or upgrades, and the affected systems are ones that interact with the data warehouse. The BI/DW evaluation needs to consider these changes to the environment because inputs to the data warehouse may change.

There are other factors to keep in mind. The migration to a replacement system, for example, includes a data conversion as well as outputs in the form of interfaces to other systems and reports. A complete validation of your BI/DW direction needs to include an understanding of the systems development plans and consideration of capabilities that could be better served with the BI/DW environment. In the cited example, the historical data from the old system may not be needed for operational purposes. Off-loading that data into the data warehouse makes it available for query, reporting, and analytics while simultaneously reducing the data volume within the operational environment.

By including an understanding of the systems development plans when validating your BI/DW direction, you gain a more holistic perspective, and you can place capabilities within the application best suited to providing them.



ONLY FOCUSING ON QUERY AND REPORTING NEEDS

The data warehouse coupled with business intelligence tools provides an excellent source of information for generating reports and satisfying regular and ad hoc query needs. These are basic capabilities, so validation of your BI/DW direction needs to examine the need for other capabilities and how well they are being satisfied.

The basic capabilities, supported with online analytical processing (OLAP), help you understand the existing and historical environment, which is certainly important. The goal of the environment, however, is to enable action. Predictive analytics enables the business community to go beyond knowing what happened to better understand why it happened and better evaluate options for meeting business goals.

Predictive analytics entails a collaborative discovery process. It includes appropriate statistical tools and techniques, displays that encompass data visualization, and a working environment that encourages people to exchange ideas and work together with data. This means that the BI/DW validation needs to address three important characteristics. First, it needs to review data granularity and quality to ensure that it can support predictive analytics. Next, it needs to ensure that the necessary tools and technologies exist and that people know how to use them. Finally, and possibly most important, it needs to assess the organization in terms of its ability to perform predictive analytics.

By validating your BI/DW direction in terms of the capabilities provided, your evaluation has a more futuristic viewpoint and ensures that the environment can move beyond meeting basic needs into one that provides for the application of more advanced capabilities.

Jonathan G. Geiger is an executive vice president at Intelligent Solutions, Inc. Jonathan has been involved in many data warehouse, business intelligence, and data management projects within most industries. In his 40 years as a practitioner and consultant, Jonathan has managed or performed work in virtually every aspect of information management. During his tenure at a major utility company, Jonathan was one of the leaders in the company's successful quest for the Deming Prize. He has authored or co-authored books and numerous articles, presents frequently at national and international conferences, and teaches several public seminars. Mr. Geiger can be reached at jgeiger@intelsols.com.



TDWI SALARY SURVEY: WAGES, BONUSES ON THE RISE

BY MARK HAMMOND

With organizations increasing investments in data management and analysis in the era of big data, it's a lucrative time to work in the BI/DW industry.

Average salaries for BI/DW professionals rose 2.7 percent in 2011 to an average of \$106,095 across 19 roles, according to the popular *2012 TDWI Salary, Roles, and Responsibilities Report*, published in March 2012. Viewed as a median, salaries rose 3.5 percent to \$102,000 in 2011.



Meanwhile, the survey found a stunning 51 percent increase in average bonuses, from a record low of \$10,866 in 2010 to a record high of \$16,444 in 2011. This reverses three consecutive years of recession-related bonus declines; the median bonus increased as well, up 19 percent to \$9,500. The percentage of BI/DW professionals receiving bonuses grew from 58 percent to 64 percent.

Now in its ninth year, the TDWI salary report provides a comprehensive, data-based guide valuable to both employees and managers to compare compensation at their organizations. Based on a survey of 1,210 BI/DW professionals, the 2012 report breaks down compensation by key role, industry, organizational BI/DW maturity, practitioner experience, age, gender, geographic region, certification, and other attributes.

MEDIA AND ENTERTAINMENT IS THE MOST LUCRATIVE INDUSTRY FOR BI/DW, WITH AVERAGE SALARIES OF \$116,363 IN 2011.

The survey has traditionally found, for instance, that professional certification pays off in higher wages. In 2011, BI/DW practitioners with three certifications averaged \$115,466 in wages—14 percent more than those with one certification. Other highlights include:

- **Gender.** Women are compensated less than men by a substantial degree. The gender gap widened in the 2012 study to a 15.3 percent margin, with women averaging \$95,390 versus \$110,010 for men. Twenty-seven percent of our respondents were female.
- **Industry.** Media and entertainment is the most lucrative industry for BI/DW, with average salaries of \$116,363 in 2011. Consulting and professional services was second, at \$115,237. State and local government, as well as education, ranked lowest at \$89,201 and \$85,113, respectively.
- **Role.** The role of BI director commands the highest average salary, at \$132,317. Factor in an average bonus of \$17,021 (received by 79 percent of BI directors) and compensation for this role is nearly \$150,000. Coming in last, business requirements analysts earn \$88,419 in salary.
- **Job satisfaction.** After hitting a record low in 2010, job satisfaction rebounded to 50 percent of respondents

characterizing their satisfaction as “high” or “very high” versus 11 percent as “low” or “very low.” Forty-five percent feel they are fairly compensated, 35 percent think they are paid too little, and 19 percent are unsure.

- **BI maturity.** Organizations with “advanced” BI environments pay 9 percent more than “beginner” organizations—\$109,646 versus \$100,198.
- **Years of experience.** Unsurprisingly, experience translates to higher wages—\$118,512 for BI practitioners with 10 or more years of experience versus \$83,941 for those in their first year.

The salary report also takes an in-depth look at 10 key roles, with breakdowns of salaries and bonuses, average age and years of experience, certifications, professional background, and more. It also benchmarks BI maturity, with 38 percent calling their BI environments “advanced,” 45 percent “intermediate,” and 17 percent “beginner.”

Given that the majority of organizations are still on the road to BI maturity and need skilled resources to handle increases in data volume, variety, and velocity, we can expect fairly strong increases in BI/DW compensation for the foreseeable future.

[Download a copy](#) of the *2012 TDWI Salary, Roles, and Responsibilities Report*. This report is exclusively available to TDWI Premium Members.

Mark Hammond is a veteran contributor to TDWI, including a number of research reports, the Business Intelligence Journal, What Works, and more. You can reach him at mjhammond@comcast.net.



BY COLIN WHITE AND CLAUDIA IMHOFF

The current industry obsession with big data and big analytics suggests that these are radical new approaches for managing and analyzing large amounts of data. Well, not really! Organizations have always struggled with efficiently and cost-effectively handling data workloads that push the boundaries of existing hardware and software technologies. Today's big data and big analytics solutions do, however, enable us to deploy applications that were not previously possible because the required information was not available, the costs were prohibitive, or the technology couldn't support the extreme workloads involved.

The Evolution of Extreme Workloads

The struggle to support extreme workloads can be traced back to the early 1960s (see Figure 1). In these early days of high-volume processing, performance issues were solved with custom-built systems that were optimized to support the extreme workloads involved. One of the first online transaction processing (OLTP) systems, for example, was

the SABRE airline reservation system. This system was custom built by American Airlines, in conjunction with IBM, to automate flight booking, which had hitherto been a manual process. When introduced in the early 1960s, SABRE cost some \$40 million (the equivalent of about \$400 million today) to develop and install, and handled roughly 83,000 reservations each day. Today, SABRE operates 24 hours a day and handles in excess of 60,000 transactions per second during peak periods, using more than 8,000 servers.

The introduction of retail point-of-sale terminals and bank automated teller machines during the '60s and '70s continued to challenge IT in supporting OLTP workloads. As with SABRE, custom-built and optimized systems often had to be built to manage these workloads and provide the required performance.

During the 1980s, transaction processing and database management technologies improved dramatically, and by the time relational database systems matured toward

The Evolution of Electronic Data

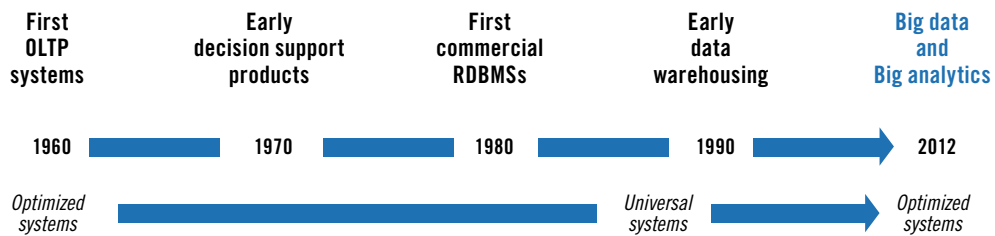


Figure 1. Source: BI Research, 2012.

the end of the decade, the need to build custom systems decreased. Organizations were able to use relational database technology to support most OLTP workloads. This helped reduce both development and maintenance costs as well as improve time to value for IT investments.

The early '90s saw the introduction of data warehousing, which enabled organizations to move beyond basic operational reporting by providing detailed analytics about business performance. Relational database technology had improved to support both transaction and analytical processing, albeit in separate systems. In 1992, Walmart was one of the first companies to deploy a terabyte data warehouse environment. Today, the Walmart data warehouse stores several petabytes of data, and multi-terabyte data warehouses are now commonplace.

As the new century approached, Internet growth continued to add to the data mountain. More recently, increasing use of sensors and sensor networks has increased data volumes to unprecedented levels. This data growth involves not only traditional structured data, but also increasing amounts of multi-structured data from a variety of new internal and external data sources.

Today, the ability of organizations to integrate, manage, and analyze growing data volumes is a major issue, and once again IT has resorted to using customized and optimized solutions to support extreme workloads. The difference this time around is that vendors provide a number of different technologies and tools that reduce the amount of custom coding required. This is what big data and big analytics are about—enabling the implementation of analytic applications that are difficult to support using traditional “one-size-fits-all” solutions.

An important distinction this time around is that these new optimized solutions don't replace the systems we

have today; they extend them. These solutions not only enable companies to improve the information content of existing decision-making applications, but also offer the possibility of using this information to identify potential new business opportunities.

The Next Generation of Innovation

We can now support today's extreme workloads by removing many of the technology limitations of the past. At the same time, we can now perform big data analytics more cost effectively and on entire sets of data as well as new data sources. These innovations have given new life to our BI environments by broadening their analytical capabilities with minimal expense and disruption. Now all enterprises can perform highly complex and sophisticated analytics to become the much-desired fact-based organization.

Colin White is the founder of BI Research (www.bi-research.com) and DataBase Associates. As an analyst, educator, and writer, he is well known for his in-depth knowledge of data management, information integration, and business intelligence technologies. He can be reached at cwhite@bi-research.com.

Claudia Imhoff, Ph.D., is the president of Intelligent Solutions (www.IntelSols.com) and founder of the Boulder BI Brain Trust (www.BoulderBIBrainTrust.org). She is a popular speaker and internationally recognized expert on business intelligence and its technologies and architectures. She can be reached at isiclaudia@aol.com.



The 2020 Workplace and the Evolution of Business Intelligence

BY HOLLIS HENRY AND TROY HILTBRAND

Abstract

The workplace is changing rapidly and will continue to change as the decade wears away and 2020 approaches. Business intelligence (BI) now has many of the tools that will ensure employees in this future workplace can be successful and deal with these environmental changes. Ideas such as “swarming,” “hyperconnectedness,” “de-routinization of work,” and “the collective” will morph employee engagement, and those organizations with mature BI practices will be prepared to support the success of their employees.

Introduction

Are “swarming,” “hyperconnectedness,” “de-routinization of work,” and “the collective” familiar terms in your organization? In the 2010 article “Watchlist: Continuing Changes in the Nature of Work, 2010–2020,” by Tom Austin of Gartner, these terms were used to describe the workplace of the coming decade. BI will play a crucial role as we move forward and as the information age matures. The role of information will become central to the way businesses function—and one of the most valued assets that enterprises possess.

Swarming

Swarming refers to small groups of people who come together for short periods of time to execute a task, take advantage of an opportunity, or solve a problem. These teams are formed to leverage a disparate set of skills that are needed to address the task at hand. Many organizations today are very structured or hierarchical, so transforming into a dynamic, organic structure will require cultural changes and will depend heavily on the effective sharing of



information across organizations. Information democratization will be a fundamental requirement to ensure teams can be successful in swarming.

The Pickup Group (PUG) and the Armory

We have already seen a manifestation of swarming in the gaming world. In the popular game *World of Warcraft*, groups of players are faced with different challenges that require unique skills each time they play. The challenge might require the brute strength possessed by a warrior, the cunning agility of a night elf, or the stamina of a dwarf; the best team composition depends on the nature of the quest at hand. Players have termed these dynamic teams of individuals who possess the right set of skills *pickup groups*, or PUGs for short.

MANY OF TODAY'S CHALLENGES HAVE BEEN CALLED "WICKED" PROBLEMS AND REQUIRE NEW WAYS OF THINKING.

For PUGs to be successful, they must be based on sound judgments about the skills and abilities other players possess and how those skills and abilities can be leveraged to tackle the next challenge. This is where BI plays a role. Blizzard, the maker of *World of Warcraft*, recognized the need for a repository of information that captures the skills and strengths of each player, and set up a site called "the Armory." The Armory pulls player stats directly from game servers and allows players to shop for team members with whom they can establish successful partnerships and PUGs. Players can investigate the skills the other players possess and view statistics about other players' past accomplishments, including their progression and maturity. With this information, players can engage others who possess the right skill mix and maturity level to form a strong PUG to take on the next epic challenge.

PUGs and the idea behind the Armory can have a direct tie to business organizations. To be successful with swarming, organizations must gain a firm grasp of the competencies and skills associated with each of their employees. This information will help teams to better decide with whom they should work to address the task at hand.

In the past, when organizations were less dispersed across continents and time zones, employees tended to be more aware of their colleagues' skills and competencies, which made it easier to select teams. Today, however, it has become woefully inadequate to rely on interpersonal relationships and face-to-face social interaction to effectively match skills to tasks. BI repositories similar to the Armory could help provide key information to organizations.

"Wicked" Problems in the Enterprise

The challenges of the past tended to be more routine, and solutions could often be found within the confines of a single department. In contrast, today's challenges require a much broader set of skills. Many of these challenges have been called "wicked" problems and require new ways of thinking.

A wicked problem is one that is extremely complex and crosses a number of knowledge domains. As groups attempt to solve wicked problems, additional challenges surface or are created. Wicked problems are always rooted in people, and the solutions are highly dependent on the changing attitudes of those people. With a wicked problem, the goal is not necessarily to solve the problem but to optimize the outcome. Because wicked problems are complex and diverse, it is necessary to attack them from multiple perspectives, and the diversity of these perspectives will typically require many different people with different backgrounds and competencies to come together. The concept of swarming enables these groups to form quickly, optimize the challenge, and then disband to form other swarms and address other organizational challenges.

Work "Dating"

BI can provide great insight into choosing the right individuals to come together and address pressing problems. We can best see the evolution in this pairing of individuals in dating sites that exist across the Internet. In the early days of mass adoption of the Internet, dating sites were perceived as electronic classified-ad systems where singles could post their profiles in the hopes of finding a love match. As these sites matured, however, companies such as eHarmony.com and chemistry.com applied complex statistical models to members' profiles to improve recommendations for potential connections.

As swarming takes hold in the workplace, organizations will leverage BI to become "matchmakers" capable of pairing individuals based on complex statistical models.

Rather than providing a mere repository of individuals with their associated skills and competencies, BI will allow an organization to move to the matchmaking stage, where statistical modeling is applied on the back end to match swarms with those individuals who possess the skills and personalities needed to make the swarm optimally effective.

Social Network Analysis

One of the factors that makes social media services such as Facebook, LinkedIn, Twitter, and most recently Google+ successful is their ability to recommend individuals to include in your circle of friends. These services use social network analysis to map connections and identify similarities in order to recommend other individuals who share similarities within *their* connections. As each connection is added to the overall network, the recommendation engine becomes increasingly accurate.

In addition to individual-to-individual connections, other connections are formed by analyzing aspects of a user's profile. As members define their profiles and specify attributes such as location, hobbies, interests, and work history, these indirect connections become part of the overall social network. This combination of direct and indirect links forms a basis for evaluating similarities between individuals in the network.

ORGANIZATIONS WILL LEVERAGE BI TO BECOME "MATCHMAKERS" CAPABLE OF PAIRING INDIVIDUALS BASED ON COMPLEX STATISTICAL MODELS.

When social network analysis is applied to the organization, there is an opportunity to infer linkages in the network through a plentitude of unstructured data. At Idaho National Laboratory (INL), one of the challenges we face is identifying potential partnerships to advance scientific research. Attending scientific conferences or being published in scientific journals are effective ways to communicate past research; by extracting key terms from this research and identifying connections with other researchers who have similar patterns (both internal and external to INL), we can gain deeper insight into linkages that exist. This added insight augments the overall social

network and can help identify researchers who may wish to join together to achieve optimal results.

Through effective social network analysis, BI can optimize participation in impromptu swarms, thereby increasing their efficiency and ensuring individuals with requisite competencies can be identified and brought together to achieve common goals.

Hyperconnectedness and "the Collective"

Hyperconnectedness refers to the "always-on" generation, where individuals are perpetually connected, often in multiple ways, via a vast network of communication channels. This hyperconnectedness means individuals are constantly fed a stream of information that keeps them up to date with the status of both their physical and virtual worlds.

Hyperconnectedness has led to an aura of "the collective." No longer are individuals alone in their activities; they are constantly attached to others through this communication network and act together to perform work and accomplish tasks.

With a user base that is constantly connected, delivery mechanisms for communicating information will continue to evolve. The paradigm of sitting in front of a desktop or laptop to receive information will give way to information being delivered in smaller chunks and to devices with smaller screens and more constricted bandwidth. Mobile BI will become the norm to ensure individuals can consume personalized content in formats optimized for their connection methods. This personal customization will force organizations to move toward ensuring BI activities are architected to fit the "develop once, deliver anywhere" model. In other words, content will be enhanced with sufficient metadata to allow the end interface environment to interpret the intent and render it in an optimal fashion. Languages such as XML allow content and metadata to be packaged together and parsed by the destination to render the output as appropriate for the target platform.

Embedded Collaboration

In the past, collaboration applications were separate and distinct from business solutions. As we move forward, this segregation between solutions will vanish and collaboration will become an integral part of business solutions. This shift will have a significant impact on the way BI is delivered.

Information provided through BI will be delivered to the collective, and, working together, individuals will assess and augment it to enhance the overall experience for all participants. The collective includes all who participate, both actively and passively. Active participants interact with the data and add value to it, while passive participants consume it along with all of the enhancements produced by the active participants. Social commenting, social tagging (“folksonomies”), and voting will all play a role in the collective’s interaction with and enhancement of information delivered through BI platforms.

Recommendation Engines

The effectiveness of today’s recommendation engines is directly related to the collective. Companies such as Netflix and Amazon take pride in their ability to help end users find the items they are looking for as well as offer suggestions about what else might interest them. Recommendation engines are a competitive advantage for these companies. Shoppers come to their sites to find items similar to those with which they are already familiar; they end up staying because their experience is incredibly seamless and enjoyable. To be successful, these recommendation engines analyze what similar shoppers are looking at and purchasing and then use this information to optimize the recommendations they generate through their engines. Input provided by the collective is the power that drives the recommendation engine to function optimally.

BI platforms can leverage the knowledge of the collective to provide guidance to business users about other areas of interest that are highly likely to attract their attention. By effectively using recommendation engines, BI can expand the visibility of key information across the organization.

Context-Aware Computing

One of the benefits associated with hyperconnectedness is also one of its greatest challenges: the reality that users cannot disconnect. This reality of the future workforce will require an increase in context-aware computing. Context-aware computing uses elements of a user’s context—where they are, what time it is, and what they are currently doing—to determine the configuration of the user experience. With the exponential increase in the amount of information flowing to end users every day, plus the “always-on” connectedness, BI must apply context awareness to its filtering model.

Examples of this filtering include (1) categorizing information based on importance and setting thresholds of time

when end users will receive different levels of importance, and (2) setting up work hours for each profile and delivering only vital information during nonwork hours.

Location-based BI will also ensure that users only receive information that is of the greatest importance to their current context. As Global Positioning System (GPS) capabilities become standard on mobile devices and as more mobile devices are used for BI, a mashup of capabilities will produce information relevant to a user’s location. For example, if a user is in the office, the flow of information might be unrestricted; however, if that user goes from the office to a supplier’s site, the flow of information would be filtered to include only what is relevant to the relationship with that supplier. If the user then travels toward home, the flow of information would be halted altogether—with the exception of only the most vital updates—to ensure the mobile device does not become a distraction and impede safe driving.

The experience of information delivery will mature to a model of context-sensitive personalization. This enhancement will enable users to continue in a hyperconnected manner without the burnout that is so often the result of this new reality of work.

De-Routinization of Work

As technology matures, it becomes increasingly capable of consuming the routine aspects of work. This automation frees employees to take on the more complex challenges that work has to offer. As these complex aspects of the job become routine, technology evolves to meet those challenges as well. This cycle will increasingly push the workforce to more challenging tasks, and the rate at which this cycle matures will become more rapid as the decade progresses. The Gartner article predicts that within five years, nonroutine work will increase from 25 percent of an employee’s day to 40 percent. Nonroutine work includes activities such as innovation, teaming, selling, and leading, which do not lend themselves to automation.

Predictive Analytics and Data Mining

BI as we know it today has four main domains: operational reporting and dashboards, data mining, real-time analytics, and predictive analytics. Both operational reporting/dashboards and real-time analytics are routine in nature. Operational reporting and dashboarding speak to what has already happened, whereas real-time analytics speaks to what is happening now. Both of these aspects of BI are highly important for organizations, but much of it is

developed once and moved into production to form part of ongoing operations.

WITHOUT THE RULES ASSOCIATED WITH NATURAL LANGUAGE PROCESSING, UNSTRUCTURED DATA IS NOTHING MORE THAN A BAG OF WORDS.

This leaves data mining and predictive analytics as the more nonroutine areas of BI. Data mining looks for hidden patterns in the sea of information generated in the past, with the intention of explaining why things happened. This function complements operational reporting and dashboards, but it is nonroutine because data mining can be used only *after* events have happened. Real-time analytics is similar to operational reporting, but it allows users to visualize what is happening right now. Once these events occur, data mining is useful for identifying *why*. Because *what happened* and *what is happening* are constantly changing, *why* they happened will continually evolve as well. Predictive analytics depends on many factors, which depend on the environment, which is always changing. New environmental factors will continue to appear and must be factored into prediction models to ensure reliable predictions.

Machine Learning

Advances in machine learning allow for continually evolving predictive models. Concepts of clustering and classification allow incoming input to be evaluated against a set of predefined data that trains the model to make intelligent decisions. As the model generates results and as those results are evaluated and fed back in as training data, the model refines itself to become more intelligent. These models are already used in many organizations to assess events in real time, but they will become a more critical component of the de-routinization of work. Rules-based engines of the past, which had to be defined to perform work tasks, will now be driven by organically changing models.

The Department of Energy (DOE) currently has projects that leverage these advances in machine learning technology. These projects allow robots to dynamically adapt to their environments and perform tasks that previously had to be performed by humans. For instance, they can operate in hazardous nuclear environments with limited or no human

interaction. Because of thick shielding in these areas, communication with the robots is virtually nonexistent, so they must be autonomous (Bruemmer et al, 2002).

Such examples of consuming situational data and making interpretations will become increasingly prevalent as the decade advances. In addition, routine activities will continue to be automated, freeing up the workforce to perform nonroutine work.

Recognizing Weak Links

Information is a valuable asset to any organization and can make the difference between success and failure. In a hyperconnected world, the window of opportunity associated with competitive advantage is short. Companies must be quick to recognize patterns relating to their core competencies and be agile enough to execute rapidly to seize the moment. *Carpe diem* will have ever-increasing importance in the survival of companies. To succeed in this environment, organizations will not have the luxury of waiting until signals become strong before taking action. They will need to monitor the influx of information and parse relevant patterns while they are still weak. The ability to identify these patterns while they are still forming will be a competitive advantage in a fast-paced and information-driven world.

In the past, BI has focused mostly on structured data; its primary task has been integrating data from a plethora of disparate transactional systems and associating meaning to that consolidated data. The conundrum, however, is that data associated with cutting-edge patterns is rarely packaged in a highly structured fashion. Unlike the data that organizations control, public data is almost always unstructured. Public data sets that *are* highly structured are typically not useful because they aren't timely enough. Patterns tend to exist in news feeds, Twitter feeds, Facebook statuses, and discussion boards, but these data sets do not have form or format and are highly dependent on the providers of the information. Such information not only lacks structure, it lacks a common language.

Applying structure to unstructured data requires natural language processing, which helps to extract key terms and concepts from the unstructured data. Language is complex, and without the rules associated with natural language processing, unstructured data is nothing more than a bag of words. With natural language processing, algorithms can separate parts of speech, remove stop words—those words that are common in almost all text but that add little value to its meaning (e.g., *a*, *the*, and *and*)—

and better evaluate the context by removing ambiguity. Thus, BI applies structure to text and increases organizations' ability to consume the information and derive value from it.

Extracting important concepts from the vast influx of unstructured data is fundamental to identifying weak patterns in near real time, which is often crucial to achieving competitive advantage. Time is of the essence.

Social Mining

As social networking becomes a fundamental interaction, the ability to capitalize on information embedded in this stream of social collaboration becomes critical. Unlike the structured capture of the collective conscience that recommendation engines harness, Twitter and Facebook provide information in the form of conversations. They hold priceless nuggets of insight that parallel or even dwarf what comes from structured sources.

Twitter allows end users to enhance the semantic nature of their information through the use of hash (#) and @ symbols. However, it remains a challenge for organizations to parse the information and identify the true meaning of what is being said, and to do so in an automated fashion that allows real-time analysis of its content.

To successfully capture the social experience, BI groups must master capturing the collective conscience, extracting critical patterns, and matching those relevant patterns. Many social networks provide access to feeds of this information through their publicly exposed APIs; however, they also recognize that information is a strategic asset within their control, so they frequently place limits on information that provides a competitive advantage.

Simulation and Experimentation

The application of visualization models to "what-if" scenarios to identify the optimal outcome is referred to as *simulation and experimentation*. As gaming engines—such as the one that powers the game *Second Life*—become available, organizations will leverage these capabilities to enhance the information coming from BI. No longer will it be sufficient to see a listing of potential outcomes in the form of charts and graphs. Organizations will want to take this information, feed it into visualization tools, and provide a richer experience to the end-user community. Scenario-based planning will take on another dimension of visualization, which will allow end users to see the nature of the outcomes and observe details from additional facets that were previously unrecognized.

At laboratories such as Oak Ridge National Laboratory, Pacific Northwest National Laboratory, and INL, scientists already have the capability to run large data sets through complex models and produce highly visual simulations. At INL specifically, scientists can test multiple scenarios against real-world nuclear power plants to visualize what would happen under a myriad of conditional variables. This simulation allows them to visually inspect the inner workings of the reactor under multiple scenarios and experiment to optimize their scientific research. One of INL's projects takes years of historical observations and simulates how nuclear reactors will respond to different variables to determine how to extend the life of nuclear reactors.

Processing power that was at one time possible only through the use of supercomputers in the scientific realm will now be available with all types of business data to all users at their desktops. This proliferation will provide BI that is *experienced* by end users—not merely *displayed* to them.

A Final Word

As the workplace evolves, so, too, will the impact and influence of BI on the organization and its success. The next decade will bring many changes to the face of the workplace as we know it, but organizations that master the art of BI will be poised to meet those challenges head on and provide a foundation for organizational success.

References

Austin, Tom [2010]. "Watchlist: Continuing Changes in the Nature of Work, 2010–2020," Gartner, March.

Bruemmer, David J., Julie L. Marble, Donald D. Dudenhoeffer, Matthew O. Anderson, and Mark D. McKay [2002]. "Intelligent Robots for Use in Hazardous DOE Environments," *PerMIS 2002*, August.

Hollis Henry is the organizational change management lead for information management at the Idaho National Laboratory. You can reach her at hollis.henry@inl.gov.

Troy Hiltbrand is the enterprise architect and IT strategist for Idaho National Laboratory. You can reach him at troy.hiltbrand@inl.gov.



The Requirements for Being an Analytics-Based Organization

BY HUGH J. WATSON

What is the status of analytics in your organization? If we are talking about *descriptive analytics* that describe what has already occurred (e.g., dashboards and scorecards), you are probably pretty far along. If, on the other hand, we are discussing *predictive analytics* that forecast what will occur (e.g., regression analysis, neural networks) or *prescriptive analytics* that help determine what *should* occur (e.g., mathematical programming), you may have pockets of use, but overall your enterprise may not be far along the maturity curve. We will refer to predictive and prescriptive analytics as *advanced analytics*.

In some enterprises, advanced analytics is moving from being a “nice-to-have” feature to a requirement for competing in the marketplace. Such enterprises are *analytics-based organizations*. For example, think of large online retailers such as Amazon.com and Overstock.com that depend on advanced analytics for demand forecasting, pricing, dynamic display of product recommendations, customer segmentation analysis, campaign management, customer lifetime value analysis, and more.

Other companies turn to advanced analytics to seize market opportunities. For example, Harrah's (now a part of Caesars Entertainment) became an industry leader by using advanced analytics to better understand its customers' gaming preferences and to offer them attractive incentives to play at its properties.

There is growing evidence that companies prosper from using advanced analytics. A 2010 IBM/MIT *Sloan Management Review* study found that top-performing companies in their industries are much more likely to use analytics rather than intuition across the widest range of possible decisions (LaValle et al, 2010). A 2011 academic study revealed that firms that adopt data-driven decision making have output and productivity that is 5 to 6 percent higher than what would be expected given their other investments and IT usage; in addition, there is a positive relationship of these measures with other performance measures such as asset utilization, return on equity, and market value (Brynjolfsson et al, 2011).

The evidence is clear that many firms will and should employ advanced analytics. The question is—what does it take to become a successful analytics-based organization?

Requirements for Success

The following factors are important for success in advanced analytics:

- A clear business need
- Strong, committed sponsorship
- Alignment between the business and IT strategy
- A fact-based, decision-making culture
- A strong data infrastructure
- The right analytical tools
- Personnel with advanced analytical skills

At first glance, this list is similar to one for business intelligence (BI) in general. However, as with many things, the devil is in the details. Our discussion will concentrate on the final four factors above, skipping the first three simply because any differences from conventional understandings are not as interesting as the others.

A Fact-Based, Decision-Making Culture

In many organizations, decisions have traditionally been based on experience, intuition, and some information. For example, at Harrah's they talk about the "Harrahisms" (things that were assumed to be true) that used to guide decision making before the company turned to advanced analytics as the basis for its decisions. In contrast, I've heard that now three things will get you fired at Harrah's—steal-

ing, sexual harassment, and a failure to use fact-based decision making.

It's not always easy to change the decision-making culture. In some cases personnel will have to change. I am familiar with one bank that changed to a very analytics-based customer intimacy strategy and, as a result, the entire marketing staff either left the bank or moved to other positions. As the CEO told me, "The old marketing staff thought that handing out balloons and suckers along the teller line and running focus groups was marketing. Marketing in our bank is now very analytical."

In addition to turnover, Deloitte Consulting identifies several things that senior management can do to create a fact-based, decision-making culture. One is to be a vocal supporter and stress that outdated methods must be discontinued. Managers can also ask to see what analytics went into decisions and link incentives and compensation to desired behaviors. BI managers can also contribute by providing the training and resources needed to benefit from advanced analytics.

A Strong Data Infrastructure

Many companies have a data warehouse whose existence facilitates advanced analytics. However, there are limitations. The warehouse may not provide all of the data needed, there may be performance-related issues, and big data may require additions to the architecture.

A warehouse contains data that has been cleansed. Sometimes modelers want data that is raw and detailed. They may need to include other data, perhaps from third-party data providers. Advanced analytics requires an iterative process of initial analysis, model development, model testing, and model maintenance. It is important to be able to partition or divide data for model development and testing purposes. Models need to be tested constantly using incoming data to detect changes in the environment. This allows them to be updated as needed to keep them relevant.

Advanced analytics tends to be CPU intensive and can negatively impact the running of other BI applications, so alternative approaches and technology may be required. For example, modelers may be given an "analytical sandbox" on a separate server. In-database data mining (such as SAS within Teradata or Oracle) is quickly taking hold. Some companies are placing advanced analytics on analytical appliances (e.g., ParAccel or Aster Data).

Then there is big data from social media, machines, and location-based systems, which can overwhelm companies' current data infrastructures. This can bring new technologies such as Hadoop and MapReduce into play to help filter and analyze the data.

In general, advanced analytics requires companies to rethink their data infrastructures.

The Right Analytical Tools

Most organizations' BI environments support descriptive BI. Although traditional BI vendors may claim their tools support data mining and predictive analytics, this is not always the case. Slicing/dicing and data visualization are *not* data mining. Having said this, traditional BI tools and predictive analytics are highly synergistic. For example, BI tools are useful in understanding the data and thinking about relationships before using predictive analytics, and data visualization tools are useful for interpreting the output from models.

Data mining requires tools that incorporate algorithms and processes designed specifically to find hidden relationships in data. SAS and SPSS are two of the traditional leaders in this space. R is a programming language and software environment for statistical computing and graphics and is now the most popular tool used by data miners. It is also at the core of many open source products. Many modelers like to work with open source products or experiment with new ones. Although you may want to standardize on a single product or a few products, modelers will often have preferences for specific tools (perhaps ones they learned in school) or those that are well suited for specific tasks.

Personnel with Advanced Analytical Skills

Advanced analytics requires three skills: knowledge of the business domain, the ability to work with large amounts of data, and modeling skills. The typical business analyst is strong in the first two areas but doesn't have advanced modeling skills (e.g., multivariate analysis). With proper training, some business analysts may be able to accept the advanced analytics challenge, at least for more structured analytics (such as customer segmentation analysis) when supported by appropriate software.

For the "rocket-science" work, new personnel usually need to be hired. These people are increasingly referred to as *data scientists* and often have Ph.D. degrees in such areas as mathematics, statistics, operations research/management science, artificial intelligence, or econometrics; they may have master's degrees in analytics, for which a number of

programs have emerged recently. Their business-domain knowledge is often limited, so they often need to work closely with people who understand the business.

Conclusion

There are many requirements for using advanced analytics or becoming an analytics-based organization. If you are not ready to commit yet but have a business need, you might consider outsourcing some of the work. For example, companies such as Revenue Analytics, Mu Sigma, and MethodCare (formerly Apollo Data Technologies) provide predictive and prescriptive modeling services.

Companies that choose to take this approach don't have to invest as much time and money in developing in-house capabilities. Of course, some may not feel comfortable turning their data over to a third-party provider, and it can get expensive if there is considerable analytics work to be done.

References

- Brynjolfsson, Erik, Lorin M. Hitt, and Heekyung Hellen Kim [2011]. "Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?" Social Science Research Network (SSRN), April 22.
- LaValle, Steve, Michael S. Hopkins, Eric Lesser, Rebecca Shockley, and Nina Kruschwitz [2010]. "Analytics: The New Path to Value," IBM Institute for Business Value /MIT Sloan Management Review, October.

Hugh J. Watson is a C. Herman and Mary Virginia Terry Chair of Business Administration in the Terry College of Business at the University of Georgia. He is a Fellow of TDWI and senior editor of the Business Intelligence Journal. You can reach him at hwatson@uga.edu.



Busting 10 Myths about Hadoop

BY PHILIP RUSSOM

Although Hadoop and related technologies have been with us for over five years now, most BI professionals and their business counterparts still harbor a few misconceptions that need to be corrected about Hadoop and related technologies such as MapReduce. I hope that the following list of 10 facts will clarify what Hadoop is and does relative to BI, as well as in which business and technology situations Hadoop-based BI, data warehousing, and analytics can be useful.

Fact #1. Hadoop consists of multiple products.

We talk about Hadoop as if it's one monolithic thing, whereas it's actually a family of open-source products and technologies overseen by the Apache Software Foundation (ASF). (Some Hadoop products are also available via vendor distributions; more on that later.)

The Apache Hadoop library includes (in BI priority order): the Hadoop Distributed File System (HDFS), MapReduce, Hive, Hbase, Pig, Zookeeper, Flume, Sqoop, Oozie, Hue, and so on. You can combine these in various ways, but HDFS and MapReduce (perhaps with Hbase and Hive) constitute a useful technology stack for applications in BI, DW, and analytics.

Fact #2. Hadoop is open source but available from vendors, too.

Apache Hadoop's open-source software library is available from ASF at www.apache.org. For users desiring a more enterprise-ready package, a few vendors now offer Hadoop distributions that include additional administrative tools and technical support.

Fact #3. Hadoop is an ecosystem, not a single product.

In addition to products from Apache, the extended Hadoop ecosystem includes a growing list of vendor products that integrate with or expand Hadoop technologies. One minute on your favorite search engine will reveal these.

Fact #4. HDFS is a file system, not a database management system (DBMS).

Hadoop is primarily a distributed file system and lacks capabilities we'd associate with a DBMS, such as indexing, random access to data, and support for SQL. That's okay, because HDFS does things DBMSs cannot do.

Fact #5. Hive resembles SQL but is not standard SQL.

Many of us are handcuffed to SQL because we know it well and our tools demand it. People who know SQL can quickly learn to hand code Hive, but that doesn't solve compatibility issues with SQL-based tools. TDWI feels that over time, Hadoop products will support standard SQL, so this issue will soon be moot.

HADOOP GETS A LOT OF PRESS ABOUT HOW INTERNET COMPANIES USE IT FOR ANALYZING WEB LOGS AND OTHER WEB DATA.

Fact #6. Hadoop and MapReduce are related but don't require each other.

Developers at Google developed MapReduce before HDFS existed, and some variations of MapReduce work with a variety of storage technologies, including HDFS, other file systems, and some DBMSs.

Fact #7. MapReduce provides control for analytics, not analytics per se.

MapReduce is a general-purpose execution engine that handles the complexities of network communication, parallel programming, and fault-tolerance for any kind of application that you can hand code—not just analytics.

Fact #8. Hadoop is about data diversity, not just data volume.

Theoretically, HDFS can manage the storage and access of any data type as long as you can put the data in a file and copy that file into HDFS. As outrageously simplistic as that sounds, it's largely true, and it's exactly what brings many users to Apache HDFS.

Fact #9. Hadoop complements a DW; it's rarely a replacement.

Most organizations have designed their DW for structured, relational data, which makes it difficult to wring BI value from unstructured and semistructured data. Hadoop promises to complement DWs by handling the multi-structured data types most DWs can't.

Fact #10. Hadoop enables many types of analytics, not just Web analytics.

Hadoop gets a lot of press about how Internet companies use it for analyzing Web logs and other Web data. But other use cases exist. For example, consider the big data coming from sensory devices, such as robotics in manufacturing, RFID in retail, or grid monitoring in utilities. Older analytic applications that need large data samples—such as customer-base segmentation, fraud detection, and risk analysis—can benefit from the additional big data managed by Hadoop. Likewise, Hadoop's additional data can expand 360-degree views to create a more complete and granular view.

Philip Russom is director of TDWI Research for data management and oversees many of 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, @prussom on Twitter, and on LinkedIn at [linkedin.com/in/philiprussom](https://www.linkedin.com/in/philiprussom).



The Confusing Future of BI and Data Warehousing

BY STEPHEN SWOYER

At the February 2012 TDWI World Conference in Las Vegas, folks were buzzing about change. Transformation. Disruption. Uncertainty.

Things are certainly changing fast; the status quo is quickly being upended and the familiar fixtures of business-as-usual no longer seem quite so fixed. Recently, for example, Apple disclosed that sales of its iOS-based products—the iPod Touch, iPhone, and iPad—had surpassed those of its Mac desktops. Five years of iOS sales against 28 years of Mac sales.

The upshot was a conference experience in which newly coined terms such as “gamification” were combined with familiar acronyms such as ACID to describe BI in the not-too-distant future.



Thanks to some combination of mobility, cloud, big data, CEP, NoSQL, and a matrix of other technical and social forces, business intelligence and data warehousing are in the midst of profound transformations. The caveat, of course, is that even though things are changing and will never again be the same, no one can quite agree about what's going to change and how. This is particularly true when it came to DW futures.

Take John Cruppi, CTO of JackBe, an upstart BI and analytics vendor that specializes in connectivity to nontraditional information sources. Cruppi says most enterprise data warehouse (EDW) practices are grounded in a misguided premise.

“Consider all of] the data warehousing guys, all of the data architects and [data] modelers, the MDM guys, [and] the data quality guys. They've worked for years—decades—getting the data [into the warehouse] and making sure it's right,” he argues. “But say you're actually able to achieve that. What happens then? Isn't it the case that [the EDW] becomes somewhat of a silo again? If the only thing you're really going to be able to get from your data warehouse is ad hoc [reporting] and dashboards, and if the self-service side is very limited, in a sense you've gone ahead and you've created another silo.”

Cruppi's point concerns the inertia of the traditional data warehouse environment. To the extent that traditional DW architectures—or, more precisely, the data management processes and disciplines that evolved to support the data warehouse—impose limitations on the freshness of data or the practical ability of DW architects and BI pros to respond to the needs of business users, Cruppi claims the traditional data warehouse has itself become a barrier.

Most people accept this argument to some degree. It explains not only the persistence of the operational data store (ODS)—which has been given a new lease on life thanks to real-time/right-time BI—but the emergence of so many new (or nontraditional) data management tools, from MPP appliances to technologies such as MapReduce (implemented natively or via Hadoop) to NoSQL repositories.

Cruppi takes it one step further, arguing that the traditional EDW shouldn't so much be abolished as set aside. That's what's happened over the last decade, he claims, as frustrated business users went out-of-band around IT, approaching vendors directly to redress their BI or analytic needs.

It's the Warehouse, Stupid

Few others put the matter quite so starkly, but many vendor attendees at the TDWI World Conference were happy to talk to *BI This Week* about the shortcomings or failures of the (once sacrosanct) EDW, or, for that matter, of ETL-centric approaches to data integration—to say nothing of the decline in the importance or prestige of relational data.

The TDWI exhibit hall, for example, played host to several consulting and services vendors that specialize in tackling emerging data integration and data warehousing problems. Companies such as Impetus Technologies, Intellicus Technologies, and Lunexa LLC joined Hexaware Technologies in promoting their specialty services to show attendees.

All of these vendors had some kind of big data pitch, too. This was fitting, as the show floor positively bristled with big data-oriented vendors. In addition to first-wave players such as Teradata, seminal competitors such as IBM (now the steward of DW appliance pioneer Netezza), appliance mainstays such as Kognitio, ParAccel, and Vertica (an HP company), and a bevy of NoSQL players were also present.

In this respect, vendors such as MarkLogic, RainStor, and SAND Technology spoke their version of truth (NoSQL) to DW professionals, championing new—typically Hadoop-oriented—ways to solve old and new problems, all while maintaining the atomicity, consistency, isolation, and durability (ACID) that are sacred to so many in the data management Old Guard.

Plenty of established vendors got into the act. Take Kalido, which—in the spirit of NoSQL—recently took up what might be called the “NoETL” banner. (Its booth featured a sign with the unicode “No” symbol superimposed over the letters “ETL.”)

“In our latest release, we built in a lot of functionality to eliminate the need for a lot of ETL,” says Kalido CEO Bill Hewitt. “We aren't eliminating the need to extract and transform data, we're eliminating the need for a separate tool to do that,” he says, adding that “we can take data directly from the source, determine what kind of data it is, put it in the right staging tables, and load it directly.”

Kalido has long championed an iterative design philosophy that incorporates what are now recognized as agile concepts and methods. Its Kalido Information Engine

likewise uses a “generic” data model that the company says helps to insulate it from business change.

AFTER A DECADE OF DISRUPTION, THINGS AREN'T IMPROVING FAST ENOUGH.

Hewitt, too, offered a vision of where things are heading, invoking the Wild-West-like atmosphere of the networking segment prior to Bob Metcalfe's invention of Ethernet to illustrate the problem of data governance.

“Network management today is virtual. If something breaks, it repairs itself or it finds an alternate route. It hasn't always been that way. In my early days at IBM, we had different data centers with multiple networks. When something went wrong, it was difficult to determine who was at fault: was it the carrier or the hardware vendor?” he explains, noting that both parties tended to blame one another.

“With governance, the same thing needs to happen. At a minimum, I need to be able to go to those underlying systems and figure out which systems are more problematic [i.e., from a governance perspective] than others. I can't do that now, because the tools [or] the standards [for connectivity, for interoperability] don't exist.”

Goodbye Excel, Hello Avatar

A distinct Futurist bent was on display, at least among a few of the vendor representatives.

Composite VP of marketing Bob Eve, for example, wasn't alone in describing a future in which BI and analytics are consumed primarily through an App-Store-like delivery channel. What's more, Eve predicted, the BI user experience will itself be less explicitly application like and altogether more real: immersive, interactive—the stuff of virtual reality.

Eve wasn't the only one talking immersion, either. Qlik-Tech Inc.'s Donald Farmer—the kind of Big Thinker who can turn a discussion involving the anesthetic properties of peat moss (Farmer is a Scot) into a compelling object lesson in BI—also championed immersion, describing it as characteristic of the user experience of the not-too-distant future. (Farmer likewise had a lot to say about the concept of “gamification.” We'll hear more from him in a future *BI This Week* article.)

If Composite Software's Eve sounds optimistic about the future of BI and DW, his assessment of the BI and DW status quo—or its IT aspect, at least—sounds considerably more pessimistic, if not downright acerbic. Business dissatisfaction with IT is nothing new, Eve points out, citing a 2011 TDWI survey in which more than half (57 percent) of respondents expressed dissatisfaction with IT's inability to respond to requests in a timely manner.

After a decade of disruption—from software-as-a-service (SaaS) to data warehouse appliances, from agile practices to “self-service” BI—things aren't improving fast enough.

“IT's full-steam-ahead mentality ... is not a course for long-term survival,” he observes, comparing the imminence of significant disruption in the near term to the “proverbial iceberg” that sunk the RMS Titanic. IT, too, will founder on the iceberg of change, warns Eve, who—in addition to the rise of an App-Store-like delivery channel—predicts a massive expansion in cloud computing capacity.

Eve likewise foresees the development of a culture in which “mavens” or “super users” have more freedom to lead by influence or example, without being hobbled by IT constraints or “standards.” “What will become of IT as we know it?” he wonders, rhetorically. “At their retirement clubs, someone needs to rearrange the deck chairs.”

Stephen Swoyer is a technology writer based in Nashville, TN. Contact him at stephen.swoyer@spinkle.net.



TDWI RESEARCH

Predictive Analytics Project Design: A Nine-Step Guide to Realizing Business Value

BY THOMAS A. "TONY" RATHBURN

Foreword

Predictive analytics continues to develop as a key decision technology across virtually every type of business. The maturation of data warehousing and the exponential growth of complex data repositories is pushing organizations to develop competencies in predictive analytics to enhance business performance and derive significant competitive advantage.

The business performance enhancements achieved are generally based on more effective allocation of the organization's resources. The most common business applications of predictive analytics allow decision makers to enhance resource allocation decisions in functional areas such as marketing, attrition/retention, fraud, risk management, and other areas where relatively low-incidence behaviors have a significant impact on business performance. The identification of groups that consistently and reliably display a behavior that impacts performance, at a rate that is different from other groups, allows organizations to adjust their resource allocation strategies.

Organizations that are successfully implementing predictive analytics solutions have learned that effective use of technology is a must. Although the proliferation of business intelligence software continues to provide ever-increasing capabilities, a project's success is primarily determined by the strategic implementation of the technology—not by the technologies themselves. No piece of software, and no algorithm, understands the domain of the decision process or the project team's unique performance metrics.

Ultimately, predictive analytics projects are evaluated based on their return on investment and by the contribution they make to the business objectives of the

sponsoring organization. That contribution is always based on each organization's unique project performance metrics.

Predictive analytics development efforts continue to be pushed to the functional decision makers to ensure that the project design captures the nuances of business problems. IT personnel and quantitative specialists generally provide specialized support to these decision makers to enhance the capabilities of the project team.

Ultimately, the effective utilization of predictive analytics is the goal-driven analysis of data to enhance business performance metrics. This TDWI Checklist Report examines the strategic steps that predictive analytics project teams must take in order to define, design, and implement successful projects.



NUMBER ONE

DEFINE THE BUSINESS OPPORTUNITY.

Predictive analytics is not a cure-all. It is one approach to enhancing organizational decision making. Successful projects focus on opportunities that arise from the following steps.

Define the business objective. Start each project with a complete, concise statement of the business's objectives. This becomes the project mission statement.

Determine performance metrics. Predictive analytics requires specific business performance metrics. As project teams proceed with development efforts, they generate many competing alternatives. These alternatives should be evaluated based on relative performance against the predetermined metrics.

Specify resource allocation decisions. The single most effective way to enhance business performance is to utilize organizational resources more effectively. Predictive analytics projects that support enhanced resource allocation decisions generally achieve a far higher return on investment than projects that attempt to model why people behave in a particular way.

Identify the behavior of interest. Business relationships include many different types of behaviors. A predictive analytics project should be based on a particular behavior that affects performance metrics. The goal is to develop a scoring system that distinguishes between those people who display the specific behavior and those who do not.

Design an evaluation process. Algorithms and software do not understand a given domain or business

performance metrics. They simply generate formulas that must be evaluated against alternative strategies, including the current decision strategy.

Plan the implementation strategy. A decision strategy has no value unless it can actually be employed in the organizational environment. This implementation must be planned from the project's start to ensure it meets delivery requirements. In addition, many development decisions—including the timing of data access and processing—need to take the implementation strategy into consideration.

Ensure the availability of historical data. Most predictive analytics projects use supervised learning techniques. These require sufficient historical data related to the behavior of interest, as well as data about the characteristics of individuals who displayed the behavior.



NUMBER TWO

PLAN FOR INCREMENTAL IMPROVEMENTS.

Incrementally developing enhancements to decision processes allows for more rapid implementation and design flexibility. It also generally results in a higher return on investment. Because human behavior is inconsistent, incremental improvements allow for the imprecision of initial modeling efforts. The following phased approach offers many advantages over large-scale, comprehensive designs.

Single-tailed models. Predictive analytics excels at identifying low-incidence, high-impact occurrences in the tails of the distributions of human behavior. Devoting the first two phases of a project to these areas of the solution space often provides significant performance enhancements with minimal effort.

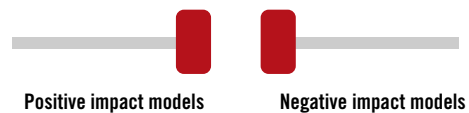


Figure 1. The two distribution tails: positive and negative.

EXAMPLE

INCREMENTAL MODEL DESIGN FOR ATTRITION/RETENTION

Suppose an organization wishes to model customer loyalty. The positive impact tail would include customers with high loyalty who are likely to be retained. The negative impact tail would consist of customers with low loyalty who are likely to be lost to attrition.

For most organizations, one of these tails has a higher priority than the other and should be modeled first. Developing each tail individually is generally simpler than trying to develop a single loyalty scoring system, especially because the variables for each behavior may be different. In general, the initial development of a scoring system for the priority behavior tail can be completed in less time, at a lower cost, and may even achieve the primary objectives of the model development effort. Such models are simple classifications with a binary output variable.

Once the priority tail model is complete, the project team may revisit the project definition and determine whether the effort of modeling the opposite tail is justified.

If both tails of the loyalty distribution are modeled, customers are typically scored on both. It is generally helpful to score the low-priority tail first. This allows high-scoring individuals to be eliminated from further consideration, which significantly reduces false positives on the priority tail.

Two-tailed models. The development of the first two phases produces two models that occasionally reach conflicting conclusions. It is often useful to then extend the analysis to a single model that simultaneously identifies behaviors in both tails of the distribution.



Figure 2. Modeling both tails simultaneously.

EXAMPLE (CONTINUED)

INCREMENTAL MODEL DESIGN FOR ATTRITION/RETENTION

After successful completion of single-tail models for both tails of the loyalty distribution, some individuals may appear to have a high expectation of both retention *and* attrition. To eliminate this conflict, it is often desirable to develop a single model that forces these conflicting instances into one tail or the other—or leaves them as an unknown.

It is important that this model be completed *after* the development of the two single-tail models. This model is also a classification with a bipolar output variable. Typically, the negative impact behavior is represented with a score of -1 and the positive behavior is represented with a score of 1, resulting in a scoring scale of -1 to 1.

Continuum models. Following the successful completion of a two-tailed model, project teams may want to rank behavior across a continuous range between tails. Again, because of the inconsistencies of human behavior, such efforts at additional precision should be completed incrementally. A quintile system (Figure 3) is a good first step that may be followed by decile and percentile ranking models until the desired precision is achieved.



Figure 3. A quintile continuum model.

EXAMPLE (CONTINUED)

INCREMENTAL MODEL DESIGN FOR ATTRITION/RETENTION

An organization may find that there are no appropriate actionable strategies for individuals in the extremes of the loyalty distribution. However, there may be significant advantages to adjusting resource allocation strategies for individuals in the low and high categories. Alternatively, some organizations find that they achieve the most significant enhancements by taking preemptive action when individuals move from one category to another.

Thus, the business objectives of the project may require additional precision in the modeling effort—ranking loyalty across a continuum from “very low” to “very high.”

To retain the categorical nature of the modeling effort, the output variable should retain an extended bipolar structure, generally on a scale of -2 to 2, with appropriate boundary thresholds established for each category.

Note that each step of this model development effort should be completed successfully before moving incrementally to the next step (e.g., decile or percentile ranking systems).

Forecasting. The inconsistency and unreliability of human behavior makes it extremely difficult to achieve the level of precision available in physical systems projects. Therefore, traditional forecasting approaches often do not succeed. Many project teams find that the best way to implement a continuous valued outcome is to develop a continuum model as just described and process the value as a separate computation, treating all observations consistently within a category.



NUMBER THREE

SHOW YOUR MODEL WHAT TO LEARN.

For decades, predictive analytics focused on the manipulation and processing of candidate input variables. Today, practitioners would be wise to invest more time and energy in defining and representing their models' output variables.

As the majority of current predictive analytics projects utilize supervised learning techniques, historical data sets associate a historical outcome with each record of candidate input variables. It is imperative that these output variables represent exactly what you want your algorithm to learn.

Single-tailed models. The incremental development approach discussed in Number Two views the early phases of a project as typically classification problems focused on identifying behaviors in one of the tails of the behavior distribution. These models are scoring systems that rank who is most likely to display the behavior of interest. The output variable in the historical data set takes on a binary (1/0) representation, where those individuals displaying the behavior of interest have a value of 1, and those who did not display the behavior have a value of 0.

USER CASE

FRAUD DETECTION

A project team for an insurance company developed a single-tail model to determine which cases to assign to auditors for review. The initial coding assigned a value of 1 to all claims containing errors or any recovery value and a value of 0 to all other claims. A large number of cases were identified that contained little or no recovery value. The team then redefined the historical output variable such that cases with a recovery value of \$5,000 or more scored a 1, and all other cases were assigned a value of 0. The net recovery value of identified audits increased over 1,000 percent.

Two-tailed models. When identifying behaviors in both tails of the distribution simultaneously, a bipolar representation is generally most useful. In a bipolar representation, we assign a value of 1 to historical cases in the upper tail of the behavior distribution and a value of -1 to cases in the lower tail of the behavior distribution.

Continuum models/forecasting models. When we develop ranking systems that model across the continuum, we encounter the increased complexity of developing forecasting models as opposed to classification models. For quintile models, practitioners have had success extending the bipolar representation to a -2 to 2 range. However, it is often necessary to convert to a continuous valued output at

this point, especially in models that are developing decile levels of precision or higher.



NUMBER FOUR

PREPARE DATA TO ENHANCE INFORMATION CONTENT.

Data is rarely stored in a manner that is consistent with the business question posed. In contrast to physical systems data, there is no underlying structure to human behavior data. It is never normally distributed; it lacks consistency between candidate input variables and the output variable, both within and between observations; and it is often incomplete. Project teams commonly devote 75 percent or more of their time on a project to manipulating and testing candidate input variables to enhance context-specific information.

Collect initial data. List candidate data considered for the project, together with any selection requirements for more detailed data. Define the relative importance of attributes and costs associated with the collection, manipulation, and integration of data sources. Special note should be given to desired data roll-ups and as-was requirements.

Describe the data. Complete volumetric analysis of the data. Identify attribute types and values. Identify data sources and key fields. Complete fundamental statistical analysis.

Explore the data. Identify the data representation strategies appropriate to the data types and project design, domain-specific relationships, interaction effects, grouping strategies, and issues related to non-linear relationships and distribution transformations.

Verify the data quality. Develop strategies for addressing missing data, data errors, query integrity, and context verification.

Select data. Algorithms interpolate between known points from the training data, so comprehensive sets of examples are required. Additional records can increase precision, but also contribute additional noise due to the inconsistency of human behavior. Model confidence is greatly enhanced with up to about 2,000 records. Huge increments are required for small gains in confidence above 4,000 records (see Figure 4).

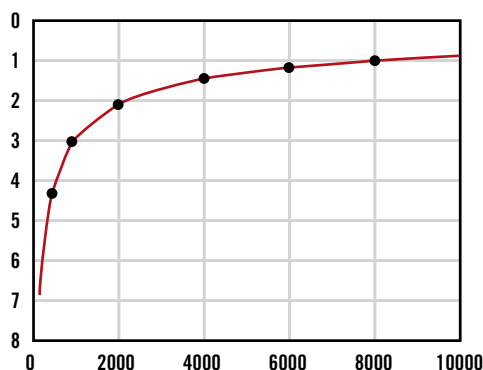


Figure 4. Gains in model precision or confidence level off after 4,000 records.

Clean data. Implement the data quality strategies on the selected data.

Construct data. Construct a data sandbox to implement and document data handling issues. The sandbox will require definitions of record structure on both a row and column orientation as well as integration of data from multiple sources. Sandboxes are typically built in a flat-file format. The construction of training, test, and validation data sets is covered in Number Seven.



NUMBER FIVE

SELECT CHARACTERISTICS THAT DISCRIMINATE BETWEEN OUTCOMES.

The next goal is to identify groups within the population that consistently and reliably display the behavior of interest at a rate that is significantly different from the population as a whole.

The search for groups. Input variables in models do not explain why people display a particular behavior. Instead, they identify who is likely to display the behavior based on a relative ranking system. Determining whether a given variable is important in the model is an empirical decision that must be tested and validated in the context of specific performance metrics.

Changing the distribution. Predictive analytics techniques are often derived from statistics, and many perform better when the input variables are approximately normal in their distribution. It is generally desirable to apply a mathematical transformation to highly skewed distributions.

Linear relationships. Most techniques from statistics assume a linear relationship between the input variable and the output variable. A mathematical transformation of non-linear relationships into linear transformations generally enhances the strength of a variable's contribution.

Reduction of power: why it works. Many variables are continuous in their natural form, yet because human behavior is inconsistent and unreliable, this level of precision often detracts from performance. Reducing continuous data fields to interval representations, and often to ordinal representations, typically enhances the reliability of a model as well as its performance.

Pruning candidate attributes. Select input variables to be retained in your models by including all candidates initially, developing a number of preliminary models, and determining which variables are used by the models that perform best. An iterative selection process allows other variables to be removed from the candidate pool on subsequent iterations until all remaining variables are used by at least one of the best-performing models.

Correlation techniques. Finally, it is often useful to conduct a correlation and covariance analysis to both confirm the relationships and to potentially further reduce the number of input variables in the models.

Thomas A. "Tony" Rathburn has more than 25 years of experience in the business utilization of predictive analytics technologies. He taught MIS and statistics as an instructor at the College of Business at Kent State University. His applied orientation to predictive analytics is grounded in his experience as vice president of applied technologies for NeuralWare, Inc., director of research at Lake Shore Trading, and his extensive international consulting experience with major corporations and governmental agencies in Europe, Asia, Australia, and Latin America, as well as in the United States and Canada. He has been a regular presenter at TDWI World Conferences for eight years and is a presenter of the popular monthly Webinar, "Data Mining: Failure to Launch." Tony is currently a senior consultant and training director with The Modeling Agency, a Pennsylvania company that provides guidance and results for those who are data rich, yet information poor. He can be contacted at tony@the-modeling-agency.com or 281.667.4200.

This report was sponsored by SAP.



TDWI Webinars deliver unbiased information on many BI/DW topics. Each live Webinar runs one hour in length and includes an interactive Q&A session with TDWI's expert presenters.

Here are some of the most popular Webinars TDWI broadcast in 2012.



VIEW MORE ON-DEMAND AND UPCOMING WEBINARS:
TDWI.ORG/WEBINARS

NEXT GENERATION MASTER DATA MANAGEMENT

watch now ▶

Master data management is one of the most widely adopted data management disciplines of recent years. That's because the consensus-driven definitions of business entities and the consistent application of them across an enterprise are critical success factors for important cross-functional business activities. The current challenge is to move on to the next generation. This Webinar presents many of the findings of the new 2012 TDWI Best Practices Report of the same title.

Original Webcast: April 10, 2012



Speaker:
Philip Russom
Sponsors:
DataFlux, IBM,
Oracle, SAP,
Talend

MYTHS AND MANDATES FOR SELF-SERVICE BI

watch now ▶

Many companies have self-service BI as their vision for empowering users. But empowerment can come at the price of overwhelming business users accustomed to receiving only a fixed report. This Webinar highlights types of self-service BI, the vision for self-service BI, best practices in deploying these types of tools, and pitfalls to avoid.

Original Webcast: June 13, 2012



Speaker:
Cindi Howson
Sponsor:
Information Builders

DATA GOVERNANCE: A BUSINESS OPPORTUNITY THE BUSINESS CAN NO LONGER IGNORE

watch now ▶

Every business leader wants to have trusted, secure, consistent, and usable information. We're now at the brink of a perfect storm of unprecedented IT megatrends. The convergence of cloud, social, mobile, and big data foreshadows the upcoming tsunami of data ripe with potential business value.

Original Webcast: August 7, 2012



Speaker:
Claudia Imhoff
Sponsor:
Informatica

TDWI World Conferences provide the leading forum for business and technology professionals looking to gain in-depth, vendor-neutral education on business intelligence and data warehousing. TDWI World Conferences feature basic to advanced courses, peer networking, one-on-one consulting, certification, and more.

The TDWI World Conference in Orlando, which took place November 11–16, 2012, focused on emerging technologies in BI/DW. The Thursday keynote, presented by BI Scorecard founder Cindi Howson, covered the trends most disruptive to BI, including mobile, social, and visual data discovery.

CLICK BELOW TO WATCH THE KEYNOTE NOW OR [VIEW ON YOUTUBE](#)



EMBRACING EMERGING TRENDS IN BI: BRINGING COOL INTO YOUR BI PORTFOLIO

Thursday, November 15, 2012

Cindi Howson, Founder, BI Scorecard

It's hard to be innovative when your BI team is deluged with fixes, fighting fires, and basic data requests. Yet transforming from reactive, report-focused development to breakthrough BI demands innovation from BI teams and technologies. In this keynote, Cindi Howson, founder of BI Scorecard and author of *Successful Business Intelligence: Secrets to Making BI a Killer App*, highlights:

- Being proactive when there's no time or budget for innovation
- Evangelizing BI in a culture resistant to change
- Prioritizing innovations that will provide the biggest value
- The trends most disruptive to BI, including mobile, social, and visual data discovery



VIEW MORE VIDEOS FROM THE ORLANDO WORLD CONFERENCE: TDWI.ORG/ORLANDO2012VIDEOS



VIEW UPCOMING TDWI WORLD CONFERENCES: TDWI.ORG/CONFERENCES

2013 TDWI EVENTS CALENDAR

Each year, TDWI offers five major educational conferences, executive summits, educational seminars and symposiums, and more. Here's a selection of some of our upcoming events.



VIEW ALL TDWI EDUCATION EVENTS: TDWI.ORG/EDUCATION



MARCH 11-18
BI SYMPOSIUM:
BUILDING AN ANALYTICS-
DRIVEN ORGANISATION
London, UK

APRIL 22-25
SEMINAR:
DIMENSIONAL MODELING
Washington, DC

MAY 5-10
WORLD CONFERENCE:
BIG DATA TIPPING POINT
Chicago, IL

JUNE 2-4
SOLUTION SUMMIT:
MASTER DATA
MANAGEMENT, QUALITY,
AND GOVERNANCE
San Diego, CA

AUGUST 19-21
BI EXECUTIVE SUMMIT
San Diego, CA

SEPTEMBER 15-17
SOLUTION SUMMIT:
BIG DATA ANALYTICS
Austin, TX

OCTOBER 20-25
WORLD CONFERENCE:
BUSINESS-DRIVEN BI
Boston, MA

NOVEMBER 4-7
SEMINAR:
BUSINESS ANALYTICS
Vancouver, BC

DECEMBER 8-13
WORLD CONFERENCE:
EMERGING TECHNOLOGIES
2014
Orlando, FL

Find out more about the many education events TDWI has to offer!

[TDWI WORLD CONFERENCES](#)

[TDWI BI EXECUTIVE SUMMITS](#)

[TDWI FORUMS](#)

[TDWI SEMINARS](#)

[TDWI SYMPOSIUMS](#)

[TDWI SOLUTION SUMMITS](#)

Best Practices Awards 2012

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 2012.

For more information, visit tdwi.org/bpawards.



Advanced Analytics Seminole Gaming

SOLUTION SPONSOR: BUSINESS INTELLIGENCE SYSTEMS SOLUTIONS

Seminole Gaming operates seven casinos throughout Florida. The multibillion-dollar company prides itself on controlling costs while achieving ultra-high profit margins. Credited with pioneering Native American gaming in the United States in 1979, the company has long been recognized for its innovation in gaming and technology.

Seminole Gaming has developed internal advanced analytics capabilities, combined them with award-winning data visualization software from BIS2 and predictive modeling from SAS, and added highly analytical human resources to create a powerful strategic analytics team. This team has helped the company operators leverage analytics in a way that drives tens of millions of dollars to the company's bottom line.

The foundation of the company's advanced analytics tool is a "big data" data warehouse that takes in 14 source systems to provide 360-degree views of both the customer and the business. Multiple methods are used to attack this big data set, including customer analytics, slot and table analytics, enterprise decision management, bundled tools, dashboard reporting, scorecards, data visualization, and predictive analytics.

Seminole's department of strategic analytics has developed a five-step process to maximize the company's benefit from this advanced analytics tool. At its heart is the concept of a "war room" in which 4-by-6-foot pages of analytics and data visualizations are posted on the walls of a meeting room and company operators work with the strategic analytics team to make the best data-driven decisions possible. The total benefit to date in terms of measured and realized incremental profit is \$26.8 million.

Across the company, new "war rooms" have sprung up. It is clear that stepping away from traditional analytical techniques and into this approach to visual analytics offers a level of teamwork, collaboration, and insight that would otherwise be unachievable.

► BI on a Limited Budget Emory University

Emory University, recognized internationally for its outstanding liberal arts college, superb professional schools, and one of the Southeast's leading healthcare systems, is located in Atlanta, Georgia. Several of Emory's graduate schools and programs are ranked among the best in the nation in *U.S. News & World Report's* "2013 Best Graduate Schools" guide. Its schools of business, medicine, and law are ranked among the top 25 programs. Emory's joint department of biomedical engineering program with Georgia Tech ranks second in the nation.

Emory's implementation of the PeopleSoft Financial Management system in September 2009 was widely anticipated across campus. A mainframe financial system had been used for decades and an upgrade was necessary for a variety of reasons. It became apparent that users' reporting needs were not being met through the new financial system. Executives ultimately approved the purchase of Oracle's Enterprise Performance Management and Business Intelligence Enterprise Edition (OBIEE) products. The BI system selection, approval, and deployment processes would take some time, so the university needed to deploy an interim system that would provide immediate relief to users.

The BI/DW team had been using WebFOCUS, Business Objects, and Microsoft Reporting Services (SSRS) for a variety of BI projects on campus. Several key individuals came together to brainstorm interim solutions for financial reporting. Emory determined that it could meet a large percentage of users' needs by deploying several targeted and flexible reports with SSRS. The university met with stakeholders in November 2010 and signed governance documents in early December. The project, known as Financial Online Reporting Services (FORS), went live across campus at the end of March 2011.

► Emerging Technologies and Methods Novation

Founded in 1998, Novation is the leading healthcare supply chain and contracting company for more than 65,000 members of VHA Inc. and UHC, two national healthcare alliances, Children's Hospital Association, an alliance of the nation's leading pediatric facilities, and Provista, LLC.

Based in Irving, Texas, Novation develops and manages competitive contracts with more than 600 suppliers. Novation recently earned the coveted Ethics Inside Certification from Ethisphere Institute, a leading international think tank dedicated to the research and promotion of best practices in corporate ethics and compliance, as well as its 2012 World's Most Ethical Companies list.

Novation's database includes data from more than 1,600 member organizations representing more than \$65 billion in annual spend. Benchmark data is updated weekly, including current pricing information. To allow member hospitals instant, on-the-go decision support, the company made its key analytic applications mobile.

Novation uses the Apple App Store to offer customer-facing iPad and iPhone native applications and has additional field-facing iPad applications. The applications allow procurement resources, pharmacists, physicians, and field teams to make data-driven decisions on price, spend, and other data in a mobile setting. MicroStrategy is the BI platform and the data marts are based on SQL Server. Collectively, these applications represent approximately 125 analytics dashboards.

With these mobile applications, Novation aimed to equip its customers and staff with mobile access to its healthcare supply chain analytics, integrate its tools into customers' daily workflow, and provide an exceptional mobile user experience that would increase adoption and engagement. Novation has gained additional value from completing several customer-facing BI applications for member hospitals: thanks to lessons learned, the company now uses a few key methodologies and design principles that keep costs low and improve the mobile user experience. This has in turn increased the adoption or "stickiness" of the apps and helped support a positive ROI.

► Enterprise BI
► Right-Time BI
Sabre Holdings

SOLUTION SPONSOR: TERADATA CORPORATION

Sabre Holdings is a global travel technology company that provides solutions to travel agencies, corporations, travelers, airlines, hotels, and car, rail, cruise, and tour operator companies through its four businesses: Sabre Travel Network, Sabre Airline Solutions, Sabre Hospitality Solutions, and Travelocity.

As a world leader in travel technology, Sabre's distribution and solutions businesses have served travel agencies, travel suppliers, and consumers for more than 30 years. Combined, Sabre's global distribution and airline reservation systems process more than 350 million travel reservations each year.

Data associated with travel reservations processed through the Sabre system, including three years of trend data and up to 331 days of future bookings and pre-trip data, is captured and stored within Sabre's enterprise travel data warehouse (ETDW). Using the Teradata Enterprise Travel Data Warehouse and the IBM Cognos Business Intelligence platform, Sabre created a powerful enterprise BI/DW platform that is the foundation for business intelligence products utilized by Sabre and its customers.

These products help Sabre solve its biggest business challenges—primarily, helping its customers generate revenue, reduce costs, and provide better customer service. The previous time-consuming processes involved collaboration among multiple business units and manual information gathering and review, with reports on data that was days (and often weeks) old.

The solutions developed and implemented provide a significant business impact every day to Sabre's customers as well as its internal users. These solutions provide valuable offerings beyond Sabre's traditional reservation products, allowing Sabre to retain and grow its customer base and enable customers to improve the services offered. Sabre has moved beyond traditional reporting and now offers dashboards for its business units of airline solutions, travel network, and customer delivery, as well as a customer management product with integrated functions and content for traveler risk management.

► Enterprise Data Management Strategies
Cooper Lighting (co-winner)

SOLUTION SPONSORS: STIBO SYSTEMS, INC. AND

ACTIONTEK INCORPORATED

Cooper Lighting, a subsidiary of Cooper Industries plc (NYSE: CBE), is a leading provider of world-class lighting fixtures and controls to commercial, industrial, retail, institutional, residential, and utility markets. The company has manufacturing facilities throughout the U.S., Canada, and Mexico.

Cooper embarked on its master data management (MDM) initiative to consolidate product data for its 23 brands. Leveraging Stibo Systems' STEP MDM application, Cooper met its goals of reducing data variability, increasing maintainability, and making data available to other applications.

Cooper Lighting's parent company has grown primarily by acquisition, resulting in inconsistent data among its brands. For example, order information tables (OITs) were maintained at a single level in a single table field, sometimes with the effect of duplicating data or creating errors.

Updating product ordering data was difficult and time-consuming. A new five-level OIT architecture was designed so each element exists only once in the STEP application. The number of table headers was consolidated from 400 to 150, making it easier for employees and customers to learn and understand product data.

All product data was reviewed and updated as required using a workflow process tracked in real time. The workflow process was dynamically adjustable to reflect real-world review sequencing requirements and took six months to develop and program. The resulting first uniform review of all Cooper Lighting product data was carried out in only three months by 120 participants in six departments and three countries.

Previously, the company's website was updated on an ad hoc basis, and product updates often took weeks. The site is now linked to the Stibo Systems STEP platform, so data updates in STEP are automatically reflected on the website after appropriate quality approvals. In addition, the project provided product data to newly updated print and virtual Product Selection Guides, as well as iPad and Android mobile deployments.

► Enterprise Data Management Strategies

Fiserv (co-winner)

SOLUTION SPONSOR: SAS DATAFLUX

Fiserv, a publicly traded financial services company based in Brookfield, Wisconsin (Nasdaq: FISV), specializes in payments, processing services, risk and compliance, customer and channel management, and insights and optimization. Fiserv serves more than 16,000 clients worldwide. One of the company's flagship products is CheckFree RXP, which supports electronic billing and payment for tens of millions of consumers via the online banking sites of more than 3,600 financial institutions across the U.S.

It is important for Fiserv to be able to extract accurate, high-quality data from CheckFree RXP to ensure that client needs and compliance requirements are being met, as well as to understand how the service is being used and how it can be continuously improved. Previously, this was a highly manual process that required a specialized skill set.

In order to streamline the process and increase accuracy, Fiserv engaged DataFlux to help establish a data quality framework that would ensure any data extracted from CheckFree RXP was reliable, consistent, and accurate.

The project established metrics to gauge data quality, which is now measured using a standard set of dimensions including uniqueness, completeness, accuracy, validity, consistency, and integrity. Data quality metrics are continually monitored and refined as the system scans millions of transactions daily.

One of the major business impacts of the solution is that it helps the company avoid unnecessary costs. For example, Fiserv created a centralized repository for performance-related statistics to track data-focused SLA measures. This helps the company proactively spot atypical performance and address issues to maintain or exceed client expectations and avoid possible penalties.

The Fiserv DataFlux deployment repaid its initial investment in two years. In addition, the impact of the data quality initiative reached further across the business than the Fiserv team had initially envisioned, benefiting areas ranging from marketing to compliance.

► Enterprise Data Warehousing

Government and Non-Profit

Michigan Departments of Technology, Management & Budget (DTMB), Community Health (DCH), and Human Services (DHS)

SOLUTION SPONSOR: TERADATA CORPORATION

Through customer service, resource optimization, and the innovative use of information and technology, Michigan Department of Technology, Management & Budget (DTMB) impacts every area of government. Nearly 10,000 users in five major departments, 20 agencies, and more than 100 bureaus rely on the EDW to do their jobs more effectively and better serve Michigan residents. The EDW achieves \$1 million per business day in financial benefits.

The EDW helped Michigan achieve \$200 million in annual financial benefits within the Department of Community Health alone, plus another \$75 million per year within the Department of Human Services (DHS). These savings include program integrity benefits, cost avoidance due to improved outcomes, sanction avoidance, operational efficiencies, and the recovery of inappropriate payments within its Medicaid program.

The Michigan DHS data warehouse (DW) provides unique and innovative information critical to the efficient operation of the agency from both a strategic and tactical level. Over the last 10 years, the DW has yielded a 15:1 cost effectiveness ratio. Consolidated information from the DW now contributes to nearly every function of DHS, including accurate delivery of and accounting for benefits delivered to almost 2.5 million DHS public assistance clients.

Michigan has been ambitious in its attempts to solve real-life problems through the innovative sharing and comprehensive analyses of data. Its approach to BI/DW has always been "enterprise" (statewide) in nature, rather than having separate BI/DW platforms for each business area or state agency. By removing barriers to sharing enterprise data across business units, Michigan has leveraged massive amounts of data to create innovative approaches to the use of BI/DW, delivering efficient, reliable, enterprise solutions using multiple channels.

► Organizational Structures

Telenor Group—Business Intelligence Groupwide (BIG), BI Competence Center

The Telenor Group, the world's sixth largest mobile communications provider, operates in 14 countries and serves over 180 million customers with more than 40,000 employees. The Telenor Group is structured into separate legal entities in each country (referred to as business units); each has its own profit and loss responsibility.

BI has long been a focus for the Telenor Group because it has access to massive quantities of data, although BI implementations success has been inconsistent. Some units are recognized as world-class in BI; others struggle with the basics.

BIG, the Telenor Group's BI competence center, has a mandate from the highest decision-making body to bring all business units to a world-class level of BI and ensure that BI is exploited in transforming business. It is dedicated to building a community in BI, spreading best practices, setting a direction for group units, monitoring progress, and supporting group companies through projects.

For community building, BIG has organized yearly conferences, developed and maintained an intranet site with over 20 GB of materials, publishes monthly BI news bulletins, and provides on-demand expertise. BIG develops best practices in collaboration with business units, collects and shares business unit development guidelines and documentation, and actively promotes best- and worst-practice sharing in its communications.

BIG sets the strategic direction for business units by creating and sharing a BI strategy development framework with processes, templates, and tools. The competence center benchmarks BI in the Telenor Group every year, conducting on-site assessments of each unit every two years, and holding monthly BI status calls with each business unit.

BIG has put BI on the agenda of every company's senior management during its short tenure. Perhaps the biggest measure of BIG's impact is the rising demand by business units for its services.

► Performance Management

Con-way Freight

Con-way Freight is North America's leading less-than-truckload (LTL) freight transportation company, providing 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. (NYSE: CNW), a \$5 billion diversified freight transportation and logistics services company.

In the past, Con-way could not easily measure operational performance or how performance measurements were tracking to strategic objectives. Each service center had an isolated scorecard, and manual coordination was required to gather the data. Even then, there were concerns about its reliability and completeness.

To address these challenges, business executives translated strategic goals into key performance metrics to manage operational performance. These metrics are calculated using data sets from the EDW that are sourced from a wide variety of application and systems. The Con-way EDW team leveraged Informatica features such as near-real-time data replication, Web services, and reusable transformation to quickly add new data sources, integrate data, and respond to changing business needs.

The operations scorecard had to be flexible in terms of time to develop, change, and deploy. Because it is used by about 20,000 employees, the scorecard also needed to respond quickly and handle thousands of user requests concurrently.

To achieve its vision, the Con-way Freight BI team used innovative techniques such as third normal form, asynchronous ETL load stream, and unconventional design patterns for high-performance ETL, among others. The final product is delivered using MicroStrategy, whose features enable it to be delivered with a dynamic, user-friendly interface across a variety of channels and platforms. The resulting low-cost solution is highly automated and is also flexible enough to adapt to future problems.

Thanks to this BI solution, Con-way personnel are able to make timely, data-driven decisions at the point of operations process execution. As a result, the company is able to realize a number of key performance improvements, including an optimized number of trucks and drivers on the roads, increased safety for employees and customers, and reduced fuel consumption, carbon emissions, and traffic congestion.

TDWI thanks this year's panel of expert judges:

Chris Adamson

DW Specialist, Oakton Software LLC

Steve Dine

President, Datasource Consulting, LLC

Jill Dyché

CBIP, Partner, Baseline Consulting

Daniel R. Evans

CBIP, Sr. Group Manager/Solution Lead, Avanade, Inc.

Aaron Fuller

Principal, Superior Data Strategies, LLC

Jonathan G. Geiger

CBIP, Executive Vice President, Intelligent Solutions, Inc.

Patty Haines

President, Chimney Rock Information Solutions

Cindi Howson

Founder, BI Scorecard

Ralph Hughes

Chief Systems Architect, Ceregenics, Inc.

Krish Krishnan

Industry Analyst and Data Warehouse Expert, Sixth Sense Advisors, Inc.

Mike Lampa

BI Business & Solution Development, Dell Services

Evan Levy

Partner, Baseline Consulting

Mark Madsen

President, Third Nature

John Myers

Senior Analyst for Business Intelligence, Enterprise Management Associates

Mark Peco

CBIP, Partner, InQvis

Shawn Rogers

Vice President Research, BI and DW, Enterprise Management Associates

Philip Russom

TDWI Research Director

David Stodder

TDWI Research Director

Jed Summerton

Business and IT Consultant, Summerton Consulting, LLC

John Thompson

CEO, Marketing Sciences, LLC

Hugh Watson

Professor of MIS, University of Georgia

Colin White

President & Founder, BI Research

Nancy Williams

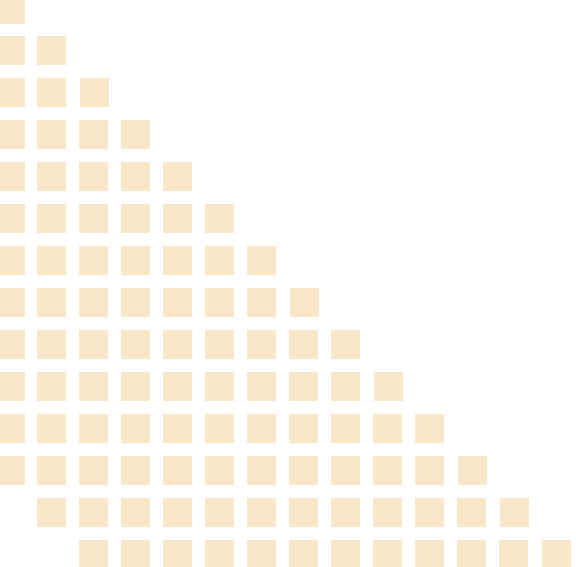
CBIP, Vice President, DecisionPath Consulting

Steve Williams

President, DecisionPath Consulting

Barb Wixom

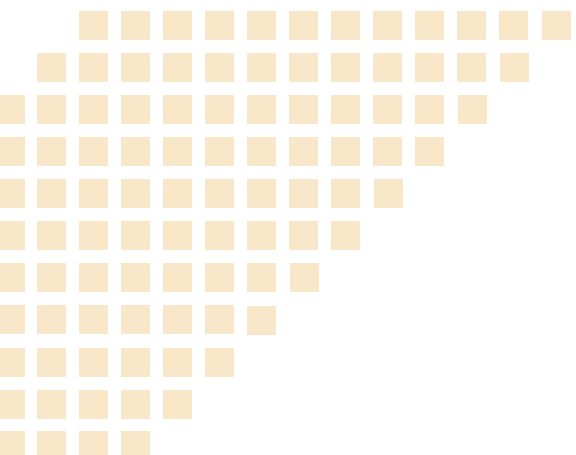
Associate Professor, University of Virginia



BI SOLUTIONS

Transforming Technologies

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

- Analytics and Reporting
 - Analytics for Everyone
 - Business Intelligence for SAP
 - Customer Data Integration
 - Dashboards, Scorecards, and Visualization
 - Data Integration
 - Data Quality
 - Data Warehousing
 - Enterprise Business Intelligence
 - Operational Intelligence
 - Performance Management
 - Predictive Analytics
 - Search Business Intelligence
- 



Birst

www.birst.com

BI CATEGORIES: Analytics and Reporting; Dashboards, Scorecards, and Visualization; Enterprise Business Analytics; Enterprise Business Intelligence; Predictive Analytics

Birst agile business intelligence delivers accurate, actionable content in an intuitive, self-service environment. It allows users to combine data from different source systems to get answers to their most pressing business concerns in real time. And, when the questions change, it adapts quickly to the new request. 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.

Unified Business Analytics

With Birst, users can create pixel-perfect banded reporting, ad hoc queries, data discovery, and in-memory analysis—all from a single platform. Every interface has an extensive library of interactive visualizations.

Advanced Analytics

All data elements in our analytical data store, including modeling results and all logical measures, are accessible for OLAP-style point-and-click analysis and report generation.

Collaborate and Distribute Analysis

Distribute results via e-mail, file servers, or on mobile devices. Every report and dashboard can be scheduled; exported to PDF, Excel, PowerPoint, or CSV; and distributed via e-mail throughout the enterprise.

Birst provides data extract options for on-premises databases, flat and structured files, as well as operational and cloud applications.

Deploy BI in Weeks, Not Months

Our customized solutions are deployed in weeks and subsequent features are released iteratively. Birst's BI solution ensures data accuracy and automation, easy integration and aggregation, support for multiple information consumption styles, flexible deployment options, and enterprisewide scalability.



IBM

www.ibm.com

BI CATEGORY: Analytics for Everyone

The Cognos family delivers analytics into the hands of everyone to drive better business outcomes. Each family member offers reporting, analysis, modeling, planning, and collaboration to address the needs of BI, performance management, and business analytics initiatives. Right-sized to meet the needs of your organization, the Cognos family offers products for:

- Individuals, who need freedom and flexibility yet want to access corporate BI and share insights with wider communities
- Workgroups or midsize businesses, who need to be up and running fast with a BI solution that is easy to install and manage
- Organizations, who need BI to meet strategic objectives spanning multiple functions, levels, and business units

Cognos Family

Cognos Insight for personal, individual users

Cognos Express for workgroups, business units, and midsize businesses

Cognos Enterprise for organizations with hundreds or thousands of users

Cognos family information to help you choose which products are best for you

Download the personal edition of Cognos Insight at no charge



Lavastorm Analytics

www.lavastorm.com

BI CATEGORIES: Analytics and Reporting, Enterprise Business Intelligence

Lavastorm—Analytics for the Agile Business

Lavastorm is a global analytics software company that enables a new, agile way to analyze, optimize, and control data and associated business processes. According to the 451 Research Group, Lavastorm Analytics was “in big data before it was cool to be in big data.” The company's products provide analysts with programming powers previously limited to IT to rapidly unify disparate data, improve data quality across diverse data sets, easily construct complex analytics, and effectively deliver actionable insight and results. Lavastorm's products enable organizations to analyze data without creating duplicate instances of the data. And analytic applications built in Lavastorm can easily be deployed as persistent solutions, providing continuous analytic monitoring across multiple systems and complex, data-intensive processes. Lavastorm's solutions have identified business improvements worth billions of dollars for some of the largest corporations in the world and have recently received industry recognition by winning the Best Revenue Assurance and the Best Business Solution awards.

Download free Lavastorm Desktop Public for easier file-based analytics, use our low-cost desktop version to create a robust data mart on the fly and build complex and conditional analytics, or use our server-based products for big volume and continuous analytic monitoring.



Melissa Data

www.melissadata.com

BI CATEGORIES: Customer Data Integration, Data Integration, Data Quality, Data Warehousing

Melissa Data is a global leader in contact data quality whose solutions are used by thousands of business and IT professionals across a wide range of industries. The company's flagship product, Contact Zone, is a comprehensive, yet easy to deploy data quality software system with powerful data integration capabilities. The product offers the following advantages:

- **Contact data quality:** Correct, verify, update, and geocode addresses in 240 countries, find and eliminate duplicates, and update and append contact name, address, e-mail, and phone.
- **No programming:** Drag-and-drop interface provides business users with access to a robust palette of data transformations to cleanse and integrate data.
- **Broad connectivity:** Access all of your sources of data, no matter where they lie.
- **Scalable:** Ideal for individuals and businesses of all sizes.

Melissa Data also offers professional services and learning programs to help support your program, and get you up and running fast. If you are looking for an affordable solution to create a single, accurate, and unified view of your customer data—turn to Melissa Data.



Tableau Software

www.tableausoftware.com

BI CATEGORIES: Analytics and Reporting; Dashboards, Scorecards, and Visualization; Enterprise Business Intelligence; Business Intelligence for SAP

About Tableau

Tableau Software helps people see and understand data. Ranked by Gartner and IDC in 2011 as the world's fastest growing business intelligence company, Tableau helps anyone quickly and easily analyze, visualize, and share information. More than 10,000 companies get rapid results with Tableau in the office and on the go. And tens of thousands of people use Tableau Public to share data in their blogs and websites.

See how Tableau can help you by downloading the free trial at www.tableausoftware.com/tdwi.



Vitria

www.vitria.com

BI CATEGORIES: Analytics and Reporting; Dashboards, Scorecards, and Visualization; Enterprise Business Intelligence; Performance Management; Predictive Analytics; Operational Intelligence; Search Business Intelligence

Operational intelligence (OI) is a form of real-time, dynamic business analytics that delivers visibility and insight into business operations, enabling smarter decisions. OI helps organizations gain real-time visibility into big data, in context, through live dashboards, and delivers continuous insight into business performance so that action can be taken.

Vitria Operational Intelligence is the only unified platform to combine continuous, real-time analytics on big data, streaming events, and business processes to enable immediate action. Vitria OI combines and correlates analytic insights from big data in motion and big data at rest into a single, actionable view. With Vitria OI, organizations can proactively address market opportunities, maintain compliance, manage SLAs, monitor network events, and anticipate the needs of customers.

Vitria OI can be used for a broad spectrum of use cases, including:

- Big data analytics
- Continuous monitoring of network events
- Customer care
- Cybersecurity
- Demand response management
- Information security
- Advanced process intelligence
- Service assurance
- Situational awareness

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, on-site courses, certification, solution provider partnerships, an awards program for best practices, live Webinars, resourceful publications, an in-depth research program, and a comprehensive website, tdwi.org.

PREMIUM MEMBERSHIP

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

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

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.

TDWI Education has even more to offer. Visit tdwi.org/education for a full lineup of Solution Summits, BI Symposiums, Forums, and BI Executive Summits.

CHAPTERS

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 website to find a local chapter in your area.

ONSITE EDUCATION

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

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

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.