THIRD QUARTER 2007
TDWI BEST PRACTICES REPORT

BEST PRACTICES IN OPERATIONAL BI

Converging Analytical and Operational Processes

By Wayne W. Eckerson

tdwi
THE DATA WAREHOUSING INSTITUTE

Research Sponsors

Business Objects

HP

Hyperion Solutions Corporation

InetSoft

SAS

Sybase

Syncsort Incorporated

Teradata, a division of NCR

BEST PRACTICES IN OPERATIONAL BI

Converging Analytical and Operational Processes

By Wayne W. Eckerson

Table of Contents

About the Author

WAYNE W. ECKERSON is director of TDWI Research for The Data Warehousing Institute (TDWI), a worldwide association of business intelligence and data warehousing professionals that provides education, training, research, and certification. Eckerson has 19 years of industry experience and has covered data warehousing and business intelligence since 1995. Eckerson is the author of many in-depth reports, a columnist for several business and technology magazines, and a noted speaker and consultant. His book, *Performance Dashboards: Measuring, Monitoring, and Managing Your Business,* was published by John Wiley & Sons in October 2005. He can be reached at weckerson@tdwi.org.

About TDWI

The Data Warehousing Institute™ (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 strategies, techniques, and tools required to successfully design, build, and maintain business intelligence and data warehousing solutions. It also fosters the advancement of business intelligence and data warehousing research and contributes to knowledge transfer and professional development of its Members. TDWI sponsors and promotes a worldwide Membership program, quarterly educational conferences, regional educational seminars, role-based training, onsite courses, certification, solution provider partnerships, an awards program for best practices, resourceful publications, an in-depth research program, and a comprehensive Web site (www.tdwi.org).

About TDWI Research

TDWI Research provides research and advice for BI professionals worldwide. TDWI Research focuses exclusively on BI/DW issues and teams up with industry practitioners to deliver both broad and deep understanding of the business and technical issues surrounding the deployment of business intelligence and data warehousing solutions. TDWI Research offers reports, commentary, and inquiry services via a worldwide Membership program and provides custom research, benchmarking, and strategic planning services to user and vendor organizations.

Acknowledgments

TDWI would like to thank many people who contributed to this report. First, we appreciate the many users who responded to our survey, as well as those who responded to our requests for phone interviews. Second, our report sponsors, who diligently reviewed outlines, survey questions, and report drafts. Finally, we would like to recognize TDWI's account and production team: Jennifer Agee, Denelle Hanlon, Deirdre Hoffman, and Bill Grimmer.

Sponsors

Business Objects, HP, Hyperion Solutions Corporation, InetSoft, SAS, Sybase, Syncsort Incorporated, and Teradata, a division of NCR, sponsored the research for this report.

©2007 by 1105 Media, Inc. All rights reserved. Printed in the United States. 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. TDWI is a division of 1105 Media, Inc., based in Chatsworth, CA.

Research Methodology

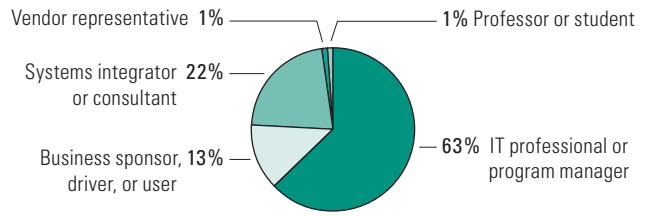
Focus. This report is designed for the business or technical manager who oversees a BI environment and wishes to learn the best practices and pitfalls of implementing an operational BI capability.

Methodology. The research for this report is based on a survey that TDWI conducted in February 2007, as well as interviews with BI and analytics practitioners, consultants, and solution providers. The survey was sent to TDWI's database of active BI professionals.

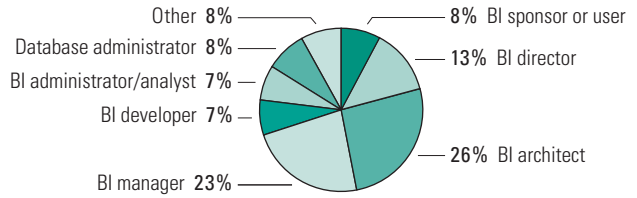
Respondent profile. A majority of the 423 survey respondents (63%) are corporate IT professionals who serve as mid-level managers in the U.S. and who work for large organizations. The bulk of the survey data presented in this report is based on 225 respondents whose groups have implemented operational BI according to TDWI's definition of the term (see page 6). The remaining respondents said their groups were exploring or planning to implement operational BI, but had not done so yet.

Demographics

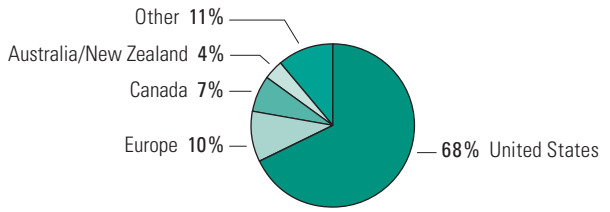
Position



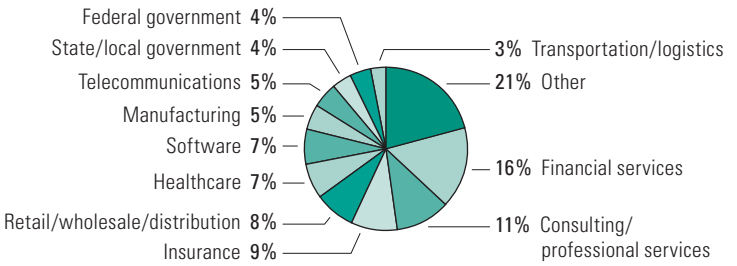
Role



Location



Industry



Based on 423 respondents, February 2007.

Executive Summary

Operational business intelligence (BI) represents a turning point in the evolution of BI. Traditionally, BI has been the province of technically savvy business analysts who spend many hours with sophisticated tools analyzing trends and patterns in large volumes of historical data to improve the effectiveness of strategic and tactical decisions. But operational BI changes this equation: it moves BI out of the back room and embeds it into the fabric of the business, intertwining it with operational processes and applications that drive thousands of daily decisions. In essence, operational BI merges analytical and operational processes into a unified whole.

In addition, operational BI increases the value of BI by delivering information and insights on demand to all workers—from the shipping clerk to the CEO—so they can work smarter and faster to achieve critical business objectives. In essence, operational BI delivers the right information to the right people at the right time so they can take action. In its extreme form, operational BI encapsulates business insights into rules and models that organizations can use to automate decisions and responses, eliminating the need for human intervention. Automating decisions not only streamlines processes and reduces costs, but also improves service and gives organizations a competitive advantage in the marketplace.

Beyond operational reporting. There are many flavors of operational BI, ranging from operational reporting and process monitoring to composite applications and decision automation. While most organizations already support some form of operational reporting, many have yet to embrace more complex types of operational BI that generate greater business value. Thus, operational BI opens up a new field of endeavor for BI and gives organizations a chance to reap greater dividends from their BI investments.

However, operational BI poses several challenges. It stretches the architectural boundaries of current BI solutions, forcing BI professionals to rethink the way they design and build systems. Queries must return in seconds rather than minutes or hours, and reports must update dynamically. Operational BI systems must capture large volumes of data in near real time without degrading the performance of existing processes and jobs on source or target systems. There is also less time to recover from a server outage, making it imperative for BI professionals to build resilient, highly available systems with sufficient backup and recovery.

This report describes the promise of operational BI and provides suggestions about how to surmount the challenges involved in converging operational and analytical processes.

Context for Operational BI

From strategic to operational BI. Operational business intelligence (BI) represents a significant departure from other forms of BI. In the past, organizations used BI primarily to support strategic and tactical decision making. *Strategic* BI helps executives and managers make a few, broad-based, strategic decisions, such as whether to introduce a new product line, acquire a competitor, change a pricing model, or expand the sales force. *Tactical* BI enables departmental managers and staff to make weekly and monthly tactical decisions, such as how to allocate resources for the new budget, how to craft a new marketing promotion to optimize sales, or how to analyze the impact of a new system or project. (See Figure 1.)

Operational BI versus strategic and tactical BI

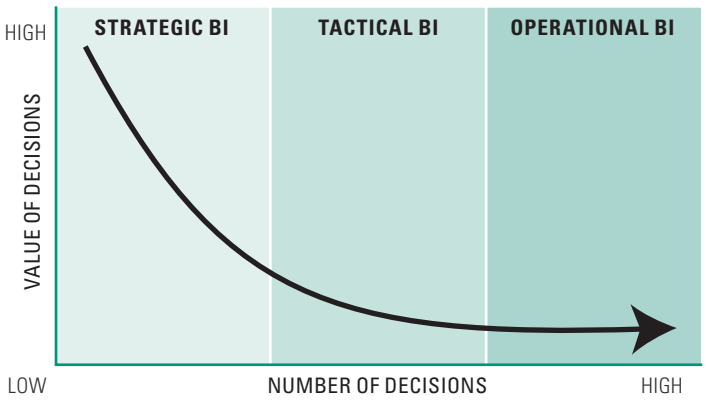


Figure 1. Operational BI supports more lower-value decisions than strategic or tactical BI, but the collective impact of those decisions is significant.

Operational BI, on the other hand, helps a much wider range of business users—from dispatchers and field technicians to hands-on managers and executives—make thousands of operational decisions each day, such as how to treat an angry customer, accommodate a late-arriving flight, or replace a defective part that threatens to shut down an assembly line. Unlike strategic and tactical decisions, operational decisions must be made quickly before a problem escalates into a crisis or a fleeting opportunity disappears. This means that operational BI systems must deliver information and insights to business users as soon as a significant event occurs so they can take immediate action. In other words, operational BI requires a just-in-time information delivery system that provides the right information to the right people at the right time so they can make a positive impact on business outcomes.

But operational BI doesn't always require human intervention. Using predictive analytics and rules engines embedded within a real-time data delivery environment, companies can streamline processes and automate responses to various types of business events. For example, an e-commerce company might dynamically generate customized recommendations for other products that Web shoppers might want to purchase. Financial services companies may approve many credit and loan applications online to speed service and lower costs. And credit card and insurance companies can detect fraud without having to manually inspect each transaction.

Operational BI helps run the core processes that drive the business on a daily, hourly, or minute-by-minute basis.

The bottom line is that operational BI helps run the core processes that drive the business on a daily, hourly, or minute-by-minute basis. As such, it supports more users, more decisions, and lower-latency data than either strategic or tactical BI and requires a fundamentally different design and architecture. And, while each operational decision may not have the same impact on business results as a strategic or tactical decision, collectively they determine whether an organization will meet its financial goals and business objectives. Thus, ironically, operational BI is very strategic.

"Companies want to offer better, faster, and cheaper services and use information to make more significant operational decisions so they can compete more effectively in the marketplace," says Craig Bliss, senior principal at Knightsbridge Solutions, a BI consultancy now owned by HP. "Clients now ask us to build an entire architecture around low-latency data, not just individual systems. These are strategic rather than departmental projects."

TDWI's Definition of Operational BI

Operational BI delivers information and insights to a broad range of users within hours or minutes for the purpose of managing or optimizing operational or time-sensitive business processes.

Trends in Operational BI

Status. Recognizing the value of operational BI, a majority of organizations surveyed by TDWI have taken the plunge into operational BI. Using TDWI's definition of operational BI (see sidebar), 53% say they currently support operational BI, while 40% do not, and 7% are not sure. (See Figure 2.)

Interestingly, among those respondents that support operational BI, only a small percentage report having mature environments. Less than one-fifth (16%) claim to have a "fully" or "fairly" mature operational BI environment, while almost half (45%) are "exploring" or "selling" the idea, and another 27% are "under development." So, while a majority of organizations have implemented some form of operational BI, few have mature or sophisticated deployments. (See Figure 3.)

We also asked respondents to select a term that their group uses to represent the functionality described in TDWI's definition of operational BI. The lion's share of the respondents selected "operational reporting" (31%), with "near-real-time BI" (13%) and "operational BI" (13%) trailing far behind, among others. (See Figure 4.) Since operational reporting has been around longer than BI (remember mainframe-generated greenbar reports?), it's clear that most organizations are still in the nascent stages of operational BI as defined by TDWI and have yet to implement more value-added capabilities, which we will discuss later.

There is an interesting duality to operational BI. There is a part of operational BI—reporting directly against transaction systems—that predates the data warehouse. But there is also a bleeding-edge side that involves blending historical and real-time data to fuel agile decision making and just-in-time responses. The challenge for many BI professionals is figuring out how to design a unified

Does your group support operational BI as defined by TDWI?

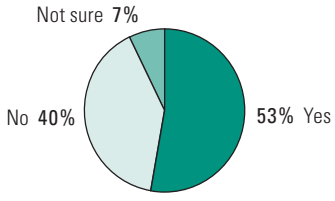


Figure 2. Based on 423 respondents.

What is the status of your operational BI environment?

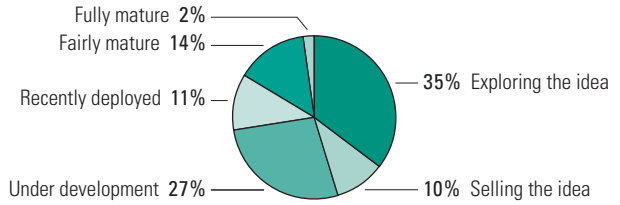


Figure 3. Based on 423 respondents.

architecture that supports both types of operational BI without impinging on traditional tactical and strategic BI applications.

Drivers. There are many reasons why organizations implement operational BI (or at least "operational reports"). Chief among them are: "improves operational efficiency," "enables workers to be more proactive," and "provides better customer service." (See Figure 5.)

Deployment. Like most BI applications, finance, sales, service, and marketing departments are leading the way in the deployment of operational BI. (See Figure 6.)

Which term does your group use?

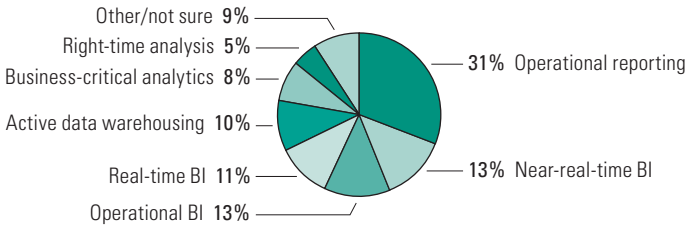


Figure 4. Based on 423 respondents.

Ranked benefits of operational BI

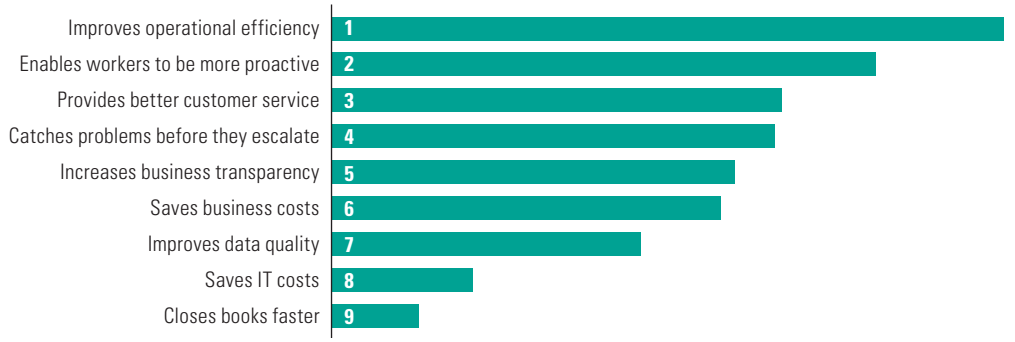


Figure 5. Respondents were asked to rank the top five benefits from the list above. The bars represent the proportional weighted rankings.

What functional areas does your operational BI environment support?

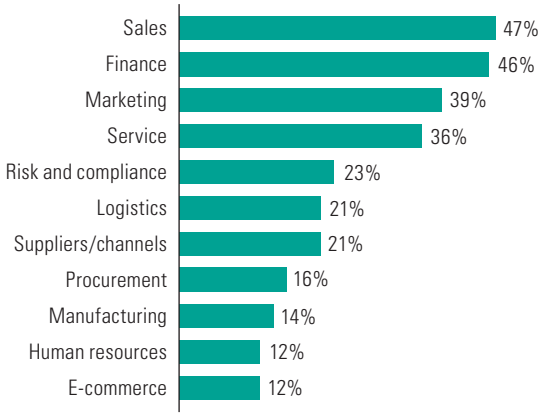


Figure 6. Based on 225 respondents who said their group has implemented operational BI based on TDWI's definition.

Operational BI opens up a whole new class of users to BI. Our survey shows that operational BI supports a higher percentage of customer-facing workers (51%) and operations workers (48%) than strategic or tactical BI applications. Interestingly, managers (77%), business analysts (65%), and executives (48%) lead the list of operational BI users, indicating that operational BI users are not restricted to customer-facing employees. Many executives check operational dashboards on an hourly basis to track performance and ensure their organization is executing against predefined objectives. (See Figure 7.)

Figure 7. Operational BI applications are used by many different types of users, but much larger numbers of operational and customer-facing workers than with other BI applications. Based on 223 respondents who have implemented an operational BI application of some kind.

Architecture. Delivering just-in-time data is a radical departure for most BI professionals, who have been schooled to update a data warehouse in a batch process during off hours, usually at night or during the weekend. BI professionals must now rethink how to architect key data acquisition and delivery mechanisms—either within or outside a data warehousing environment—to deliver information and insights on an hourly or even second-by-second basis and ensure high availability to support mission-critical business processes.

Low-latency data delivery. I use the terms "just in time" or "right time" to describe a low-latency data delivery system that supports operational processes. What is "just in time" varies among organizations and applications. Some operational BI applications—such as Wall Street trading systems and e-commerce recommendation engines—apply analytics to real-time data where latencies of a fraction of a second make a huge difference. But most operational BI applications don't have such stringent data delivery or execution requirements, and can be supported with data latencies between 15 minutes to several hours before users need to take action or workflows are triggered.

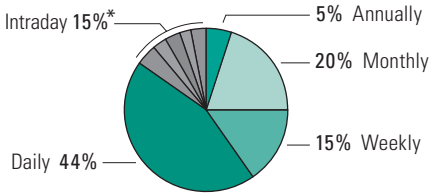
For example, 1-800 CONTACTS, a provider of replacement contact lenses, recently implemented an operational dashboard, updated every 15 minutes, that tracks call volumes and orders in its call center operation. The new dashboard, which replaces weekly reports, gives sales agents "the information they need to change behavior now instead of at the end of the week," says Jim Hill, data warehousing manager at 1-800 CONTACTS.

"Just in time" puts the emphasis on the business value of information, not its latency.

Ultimately, business executives don't care about the degree of latency in a BI system. They simply want these systems to deliver information when they need it to make a decision, fix a problem, or capitalize on an opportunity. *Right time* and *just in time* put the emphasis on the business value of information, not its latency.

Refresh rates. To support operational BI, many organizations update a portion of the data elements in their data warehouses on a just-in-time basis. On average, organizations update 15% of the data elements in their data warehouses more frequently than every 24 hours, ranging from every several hours to every several seconds. (See Figure 8.) Many organizations use a variety of techniques besides a data warehouse to capture and deliver just-in-time data, including enterprise information integration (EII) tools, operational data stores, Web services, and composite applications (we'll discuss these in a moment).

Refresh intervals in the average data warehouse



*Intraday processing includes: every several hours (4%), hourly (3%), every 15–30 minutes (3%), every several minutes (2%), and every several seconds (3%).

Figure 8. The percentage of data elements in an average DW that are updated at various intervals. Based on 223 respondents who have implemented operational BI.

Delivery environment. Most organizations deliver operational data via static reports (61%), followed by ad hoc queries (55%), parameterized reports (51%), and dashboards (47%). (See Figure 9.) The preponderance of static reports makes sense given that operational processes do not change often. Also, many operational reports deliver data at the lowest level of granularity, so there is no need for users to drill into more detail.

How do your users view operational BI data?

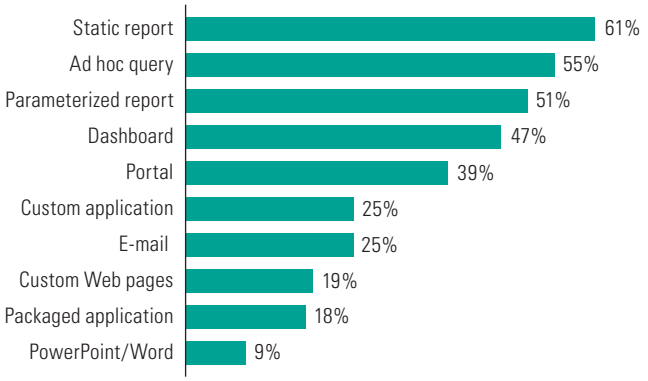


Figure 9. Based on 225 respondents who have implemented operational BI according to TDWI's definition.

Report refresh rates. We also asked respondents how operational reports are updated. The majority (64%) update operational reports at predefined intervals. (See Figure 10.) For example, Harley-Davidson has set the refresh interval at 15 minutes for the operational reports displayed on 56-inch monitors in its distribution center, says Jim Keene, systems manager of global information services at Harley-Davidson Motor Co. He adds that they also deliver parameterized operational reports that give users a choice of filters to apply to existing reports.

How are operational BI reports updated?

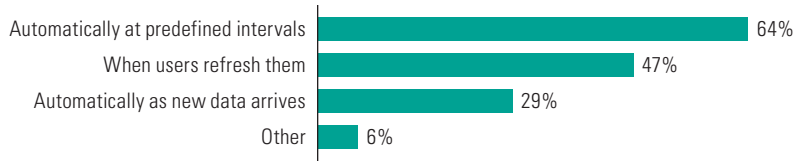


Figure 10. Based on 225 respondents who have implemented operational BI according to TDWI's definition.

KPIs and alerts. Our survey respondents rated the ability to access and analyze key performance indicator (KPI) or metric data in near real time—a key capability of a performance dashboard—as the most attractive operational BI reporting feature. Right behind is the ability to access, combine, and analyze data from two or more sources in near real time, another key dashboard attribute. (See Figure 11.)

Rate the importance to operational BI users of the following reporting features.

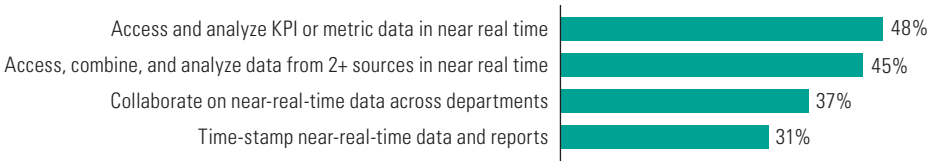


Figure 11. Based on user ratings equal to "high" from 223 respondents who have implemented operational BI.

Actions and triggers. The most common actions invoked by an operational BI system are alerts (41%) and e-mail messages (34%), with database updates, queries, and triggers following. This data suggests that most actions still require human intervention and analysis, while more automated responses (such as queries, updates, and triggers) have been integrated in about a quarter of cases. However, another quarter said they do not trigger any action using operational BI (see Figure 12).

What actions does your operational BI system automatically trigger?

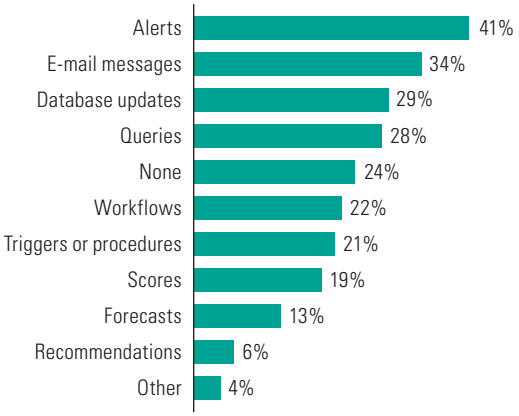


Figure 12. Based on 225 respondents who have implemented operational BI according to TDWI's definition.

A Framework for Understanding Operational BI

Operational BI is a big tent. It encompasses many different approaches, architectures, and technologies, which can be used to support a variety of decisions and accomplish a multiplicity of tasks. I have created a framework to help identify the different levels of operational BI; see Figure 13.

Operational BI framework

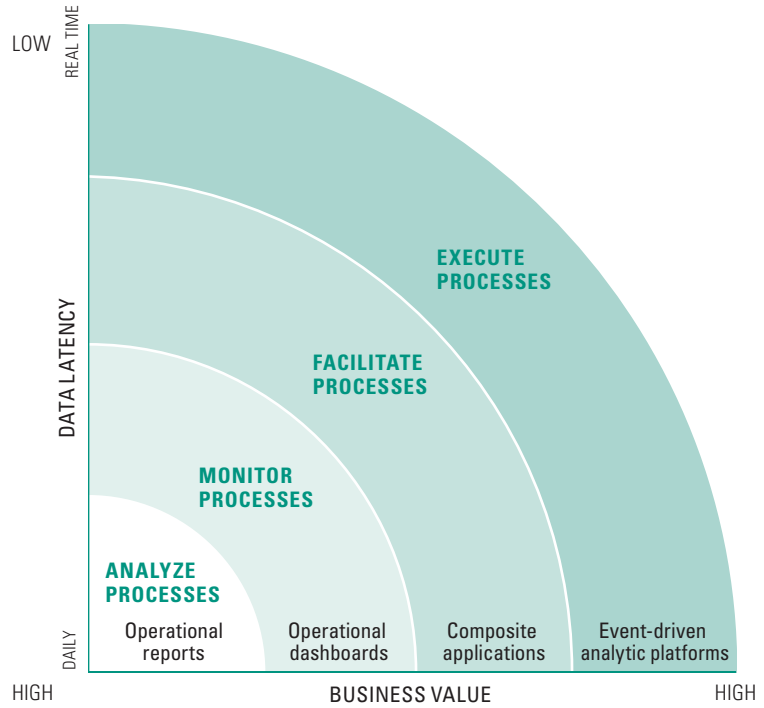


Figure 13. The progression of operational BI from lower to higher-level solutions.

Levels of operational BI. In the first level of operational BI, users *analyze* operational processes using traditional reports—something most organizations have done for a long time, even before they implemented data warehouses. The next level occurs when users *monitor* processes on a just-in-time basis using graphical key performance indicators. In the next level, IT developers *facilitate* processes by embedding BI into operational applications using service-oriented architecture (SOA) to merge operational and analytical processes into a single application. Finally, the culmination of operational BI is when organizations *execute* processes using event-driven analytic engines, predictive models, and other techniques that monitor events and trigger rules to automate or guide actions.

Most firms are stuck in level one of operational BI.

As organizations ascend these levels of operational BI, they gain greater value from their BI investments. At the same time, the data latency or time between an event and when data about that event is presented to a user or application decreases, sometimes to zero in event-driven systems that automate decisions and actions. Today, most organizations are stuck in level one—operational reporting—and are just beginning to explore options for monitoring processes using dashboards. At the same time, they are listening to thought leaders who proselytize the value of composite applications and SOA and observing a few leading-edge practitioners who are automating decisions.

Let's step through each of these levels in more detail to understand the range of possibilities within operational BI.

Level One: Analyze Processes

Purpose: Obtain current view of activity in one or more applications to help workers make decisions and optimize processes.

In level one, organizations use operational reports to analyze processes. These reports, which generally consist of master-detail reports or tabular lists of transactions drawn from a single system, help front-line workers and their managers see what's going on in specific processes and make spot decisions to ensure that processes run as efficiently and optimally as possible. Most organizations already use operational reports of some kind; it's hard to run a business without them.

There are many ways to implement operational reports, and companies have tried them all.

There are many ways to implement operational reports, and companies have tried them all. There are two major approaches: either companies can build operational reports by querying transaction systems directly, or they can off-load the transaction data and query the data separately. (See Table 1.) Most companies prefer to query transaction data directly, if possible.

APPROACH	COMMENTS
Query application directly	**Degrades OLTP performance**
1. Use packaged operational reports	Cheap but inflexible
2. Use BI tool	More flexible, but multiple sets of metadata
3. Use EII tool	One tool, one set of metadata for all apps
Off-load OLTP systems	**Expensive; redundant data**
1. Use ODS to consolidate data	Off-loads OLTP systems, but adds a DBMS
2. Use real-time data warehouse	Consolidates reporting, but requires real-time data warehouse architecture

Table 1. Level one approaches

Direct query. Organizations that query a transaction system directly either use 1) canned reports that come with the transaction package, 2) a BI tool to query each transaction application separately, or 3) an enterprise information integration (EII) or federated query tool to query multiple applications at the same time. Canned operational reports are cheap, since they come bundled with the transaction system, but they are often static and inflexible. BI tools provide greater flexibility, but generally require a different set of metadata for each OLTP source. EII tools provide the greatest economy, since developers can create one metadata model for multiple transaction systems. However, OLTP schemas are notoriously complex and can be challenging to model.

"With operational reports, we prefer to leave the data in place if possible rather than move it."
—Jim Keene

"With operational reports, we prefer to leave the data in place if possible rather than move it," says Jim Keene of Harley-Davidson Motor Co. Keene says the company's distribution center, which ships products to dealers, runs 3,000 user-driven, operational reports a day directly against transaction systems using a BI tool.

Off-loading. However, all direct query approaches have the same downside: they can degrade the performance of transaction systems as query volumes increase. To avoid this problem, most organizations off-load transaction data into an operational data store or data warehouse equipped to handle short, tactical queries. Says Keene, "If our transaction systems are running at 70 to 80% capacity, then we'll do near-real-time replication and run the reports off of an operational data store." Organizations may also off-load transaction data when operational reports require data from several sources, and they either don't own an EII tool or don't believe an EII tool can adequately support complex, multi-way joins against multiple transaction systems.

While off-loading transaction data can ease performance issues, it can create other problems. For example, the downside of an ODS is that it adds another database and reporting environment for the IT staff to manage and support. Many companies have accumulated dozens, if not hundreds, of ODSs, which become a drain on IT time and resources.

Mixed workload. Consequently, some organizations prefer to consolidate operational reporting into an existing data warehouse so they can manage all reporting within a single infrastructure. This not only unifies operational and analytical processes from a technical perspective, but also gives business users a single point of contact (the BI team) to get reporting help or request a custom report rather than work with different IT groups based on the type of report.

Using a data warehouse for operational reporting poses challenges.

However, using a data warehouse for operational reporting poses challenges. Organizations must modify the data warehouse architecture to support both real-time updates and fast, tactical queries, two things that most data warehouses are not designed to handle. In addition, this approach requires more data integration and modeling work to get transaction data into the warehouse compared to an ODS, which typically stores mirror images of the OLTP data in the original schema. In contrast, most data warehouses (except those running on Teradata) require organizations to model the data into a dimensional schema at some point. We will discuss challenges and techniques for managing real-time data warehouses later in this report.

Level Two: Monitor Processes

Purpose: Monitor processes, alert users to exception conditions, and analyze data to determine root causes.

A performance dashboard is a graphical exception report tailored to each user.

The second level of operational BI involves monitoring business processes using performance dashboards. In reality, the difference between an operational report and a performance dashboard is not as great as some might think. A performance dashboard is, in effect, a graphical exception report tailored to each user that focuses on a few metrics that represent the performance of key processes or people. With a glance, users can ascertain whether performance is above or below expectations. Also, dashboards can alert users to exception conditions through e-mail or wireless devices, so users don't even have to log on to monitor their areas of responsibility.

Performance dashboards offer a more effective way to monitor processes and people than static, tabular operational reports. In many ways, performance dashboards are the new face of business intelligence: they conform to the way users work rather than forcing them to conform to the way BI tools or reports work. TDWI research from 2005 shows that a majority of organizations had already implemented a performance dashboard (i.e. dashboard or scorecard) and another 33% were under development.[1]

[1] *From Wayne Eckerson,* Development Techniques for Creating Analytic Applications, *TDWI Research report, 2005. Available online at www.tdwi.org/research.*

Dashboard approaches. Two years ago, most organizations with an operational dashboard—or any dashboard for that matter—built it themselves by piecing together portals, OLAP engines, and reporting systems. I call this first-generation dashboard technology. (See Table 2.)

APPROACH	COMMENTS
First-generation, custom-built dashboard	Functional, but navigation between layers is not seamless
Second-generation, purpose-built dashboard	Unified dashboard platform with seamless navigation, Flex/AJAX GUI
Event-driven analytic platforms	Tailored for real-time data acquisition and process management

Table 2. Level two approaches

Today, vendors have almost caught up with business interest in dashboards and now offer fairly robust applications. Some of these are embedded within the BI products of leading BI vendors, while others are built from scratch using the latest graphical interface techniques (such as Flash, AJAX, and Flex) and data integration technologies (such as Web services, data federation, and in-memory processing). A third generation of operational dashboards is now emerging, called event-driven analytic platforms. We'll discuss these tools in later in the report, since they are designed to do more than just populate KPIs.

Level Three: Facilitate Processes

Purpose: Embed metrics or reports within operational applications or portals.

The next level of operational BI is to use BI to facilitate processes. Companies can do this by embedding metrics or reports within operational applications or portals. In other words, instead of requiring users to use two different systems—one to run the process and another to analyze it—organizations can embed the analysis directly into the process and the application that drives it so users get the information they need when they need it without shifting software contexts.

Embedded code. There are many approaches to embedding BI, some of which are as old as the hills. (See Table 3.) For instance, companies can write custom code and embed it into a custom-developed application. Java developers, who are apt to reinvent the wheel when given the chance, have done this for years, creating custom charts and reports for their Web-based applications. Today, developers can use newer BI tools, such as InetSoft's Style Report or open source BI tools, to embed reporting functionality into their Java applications. They can also embed predictive models created with analytic modeling workbenches, such as those from SAS, into applications to support real-time decisions.

APPROACH	COMMENTS
Embed custom or open source code	Cheap to build, but can be costly to maintain
Embed proprietary API calls	How much functionality do APIs expose?
Embed Web services API calls (SOA)	How much functionality does API expose?
Extend the ODS or real-time DW	Query and update integrated data
Build composite applications	Dependent on SOA middleware
Purchase packaged application	Locked into software platform and application schema

Table 3. Level three approaches

Embedded calls. Another way to embed BI into a transaction application or portal is to embed calls to a third-party BI tool. Many independent software vendors (ISVs) have done this for years, but until recently, most BI tool APIs were proprietary in nature and exposed only a minimal set of functionality. Combined with the costs of the third-party tools and lack of API interoperability, many developers decide to write analytical components from scratch rather than integrate with a third-party tool via an API. Fortunately, many BI and predictive modeling vendors now expose functionality through Web services and XML, making it easier and more cost-effective to entertain this option.

Portals are the place most organizations would like to embed BI functionality. Most companies, however, want to display only a KPI or chart through a portal pane, not an entire report with the full set of BI functions. Most BI tools now support JSR 168, a standard adopted by most commercial portals for integrating third-party content. Even so, merging BI output into portals is not always a straightforward process, since each portal and BI vendor implements the standards in slightly different ways.

Extended ODS. Another approach is to extend an ODS to drive operational applications in real time. The ODS can be a stand-alone ODS or set of logical tables within a real-time data warehouse. The ODS delivers integrated data to a core business process and takes updates in return. A great example is a call center application where a customer service representative (CSR) queries the ODS to view an integrated profile of a customer who has called into the center. The profile contains the latest information about the customer, including recent interactions, purchases, and cross-sell recommendations. If the CSR learns something new about the customer—such as change of address or marital status—the CSR can update the profile in the real-time data warehouse. Extended ODSs can also be used to support master data management (MDM) and customer data integration (CDI) applications.

Composite applications. Recently, there has been a buzz in the industry about composite applications. Using Web services and a service-oriented architecture, developers can encapsulate functionality from any application running on any platform architected in any manner and expose them as services. Using graphical workbenches, developers can drag, drop, and connect services to create robust applications that cut across once rigid boundaries imposed by systematic differences in hardware, operating systems, programming languages, metadata, and file formats. These so-called composite applications enable a marketer, for example, to analyze customers, target market segments, download customer lists, and set up and execute campaigns, all within the same application. Composite applications promise unrivaled flexibility for developers and ease of use for workers. Although vendors are making progress toward achieving this vision, composite applications are still in the early stages of maturity.

Composite applications promise unrivaled flexibility and adaptability for developers and ease of use for workers.

Packaged applications. Perhaps a more realistic alternative today for converging operational and analytical processes is a packaged application where an applications vendor has prebuilt both the operational and analytical components and integrated them tightly for a specific business domain in a specific industry, such as sales pipeline management in the insurance industry. These packaged analytic applications provide out-of-the-box functionality, which speeds deployment time. This is great—as long as you don't try to customize the package to meet your specific requirements. Unfortunately, this is a temptation most companies are unable to avoid!

Level Four: Execute Processes

Purpose: Capture business events and apply rules to automate the execution of business processes.

In level four, BI is not just providing insights into the process, it *is* the process!

Now we come to the final—and perhaps most exciting—stage of operational BI, which is when we use analytics as an engine to execute processes and workflows. Here, BI is not just providing insights into the process—it *is* the process! The techniques described here enable companies to capture business events and apply rules to assist or automate the execution of business processes. (See Table 4.)

APPROACH	COMMENTS
Database triggers, stored procedures	Simple and inexpensive, but database-driven and requires IT to create and maintain them
BI tool alerts	Database-driven; polling overhead; coarse-grained rules
Custom event-driven analytics applications	Custom built; difficult to maintain and extend; incorporate predictive models; expensive
Event-driven analytic platforms (geared to high-volume, real-time data streams)	**BI oriented:** Celequest, SeeWhy, Syndera **BAM oriented:** Systar, Actimize, Blue Agave, TIBCO, Savvion, FileNet, Pegasystems **Database oriented:** Streambase, TimesTen

Table 4. Level four approaches

Database triggers. There are many low-cost alternatives for automating processes that developers and DBAs have used for years. For instance, database triggers and stored procedures are custom ways to monitor events and kick off actions based on predefined rules. These techniques are simple, if no longer sexy. But they don't work in every situation. For instance, database triggers don't work if the events occur outside of a database or data warehouse. In addition, they require IT to set up and maintain.

Another approach is to use BI tools, most of which have some sort of notification engine. BI tools can query a database at regular intervals and issue alerts or generate new queries when query result sets exceed a predefined threshold. Like triggers, these alerts are database-driven, and polling can add significant overhead depending on the query or polling interval. Plus, the rules can be hard to set up and tune so that the system doesn't inundate users with irrelevant alerts or trigger inappropriate actions. However, some BI vendors, such as SAS and Business Objects, have integrated their BI tools with a business process management product to provide the best of both worlds—query and report processing and rules-driven event management.

Custom-built applications. Until recently, organizations that have needed to monitor activity in real time and execute processes—such as Wall street traders, credit card companies, e-commerce firms, and high-volume manufacturers—have built custom applications using homegrown rules engines, predictive models, real-time data warehouses, and other components. For instance, some banks have created online credit systems that enable customers to apply for and receive credit or loans online in real time. (See Figure 14.) However, these homegrown applications are expensive to build, extend, and maintain.

Event-driven analytic platforms (EAP). The final approach is event-driven analytic platforms, which provide a packaged alternative for monitoring events and automating decisions in real time.

Online credit authorization

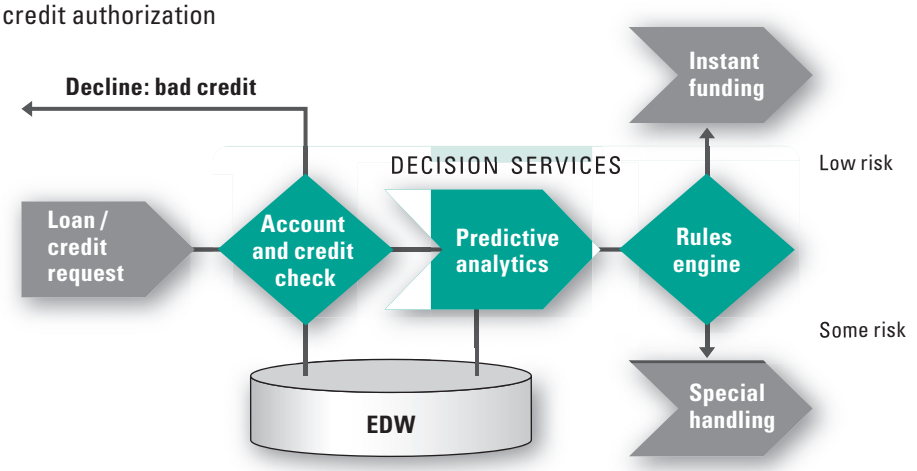


Figure 14. Many financial services firms have built custom applications that extend credit to customers online in real time.

Event-driven analytic platforms can execute processes and automate decisions.

These engines go by many names: business process management (BPM), business activity monitoring (BAM), operational dashboards, complex event processing (CPE) engines, and so on. The bottom line is that these platforms are like intelligent sensors that organizations can attach to their transaction streams. They take a continuous reading of the health of various business processes and trigger actions based on prior experience. In other words, these engines can execute processes and automate decisions.

EAP architecture. Since these products are geared exclusively to supporting operational BI applications, it is worth dissecting their architecture. (See Figure 15.)

Event-driven analytic platform architecture

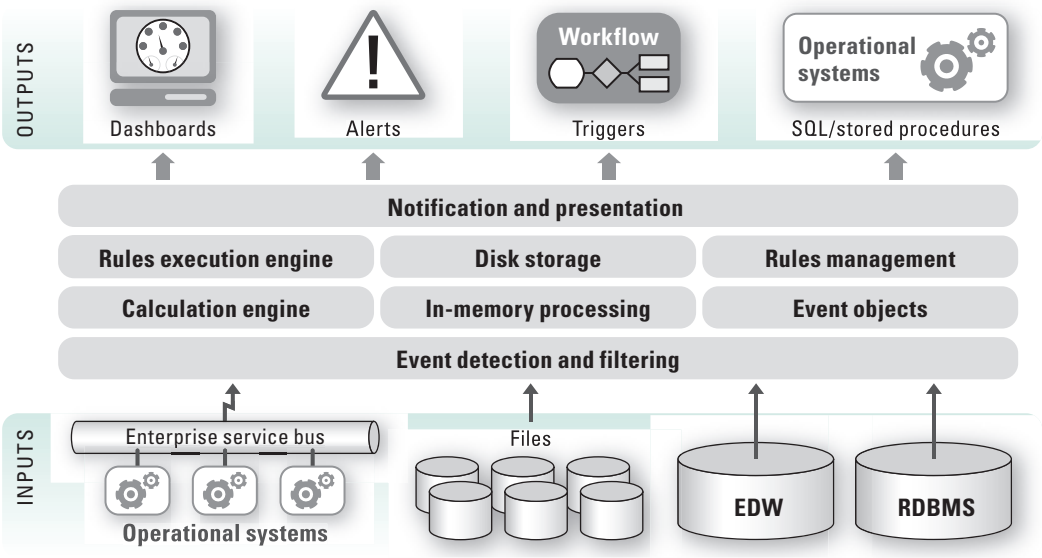


Figure 15. An event-driven analytic engine captures and correlates a continuous stream of business events to provide real-time insight into operational processes and automate responses through alerts and business workflows.

The calculation engine will hold events in memory until they can be combined or correlated with other events as they arrive.

EAPs must interface with a variety of messaging systems (WebSphere MQ, Java Message Service, etc.) and support adapters to extract data in batch from relational databases, flat files, Web services, and other systems. These adapters must be able to filter, validate, and select only the relevant events from a transaction stream to avoid overloading the system. The captured events are then processed in memory by a calculation engine that aggregates and/or correlates specific types of events depending on the business process. The calculation engine may hold certain events in memory (or persist to disk in a long-running process) until they can be combined or correlated with other events when they arrive.

Rules. The calculation engine passes the calculated objects to a rules engine, which determines whether the activity is within acceptable limits or has exceeded predefined thresholds, in which case the rules engine triggers a response. Rules can be simple or complex, but usually adhere to if/then logic: "If *x* occurs, then do *y*." Complex events can be generated from many simple events or other complex events, and the events don't necessarily need to occur in sequence or within a specific time period. While EAPs embed rules engines, organizations can also use stand-alone rules engines to drive event-driven analytic processes.

A robust rules system makes it possible to instrument an event notification system with fine-grained accuracy. The last thing you want is to inundate managers with needless alerts, which could cause them either to ignore the alerts altogether or to take unwarranted action that makes things worse. For example, the notification systems within BI tools don't have the intelligence required to model complex processes, and most organizations that have tried to implement them have turned them off.

Predictive models can improve the sophistication of rules and relevancy of alerts and triggers.

Scoring code. One way to improve the sophistication of rules is to incorporate scores generated by a predictive mining tool that applies statistical or machine-learning algorithms to historical data to predict future outcomes. The scores—usually a value between 0 and 1—define a relationship between an entity and the outcome of interest. Examples might include a customer's propensity to respond to a specific online offer or discontinue using a service (i.e. churn), or when a series of credit card transactions indicate that the card has been stolen.

For instance, fraud detection systems create predictive models based on your history of credit card purchases to predict with high accuracy which transactions are potentially fraudulent. Because they incorporate these models or rules within (mostly homegrown) EAPs, they can notify you almost immediately when someone has stolen your card and made fraudulent transactions with it. The EAP processes the transactions in real time, identifies potentially fraudulent transactions, and freezes the account. This triggers a notification that a person should examine the transaction(s) manually, corroborate the system's analysis, and call the customer to find out whether the transactions are fraudulent.

Action-oriented. Today, most EAPs power dashboards that enable users to monitor business processes and compare performance against plans or expected outcomes. But these engines also automate responses by triggering a series of activities in a predefined workflow. For example, if sales of a product that an e-commerce company is promoting have fallen below desired levels for three of the past five hours, the analytic engine will send an alert to the campaign manager and execute a procedure to disable the promotion on the Web site. EAPs can be used to dynamically optimize pricing and inventory levels, proactively monitor the obsolescence of assembly line machinery, or detect fluctuations in the prices of various security instruments to identify arbitrage opportunities. They can also be used to identify when expected events *don't* occur and launch proactive steps to address the situation.

When to automate. Obviously, before you invest in a decision automation system, you should have a good idea which applications might be candidates for the system. Consider implementing a decision automation system as the number of decisions and events per minute increases. These are generally decisions and processes managed by front-line workers, such as shipping clerks, shop floor machine mechanics, customer service representatives, and dispatchers. In contrast, general knowledge workers, analysts, managers, and executives who make fewer decisions on aggregated event data are best served by traditional BI and data warehousing systems.

Automate high-frequency decisions in well-known business processes where event volumes are large and actions must be immediate.

Bottom line, automate high-frequency decisions in well-known business processes where volumes of events are large, actions must be immediate, and the payoff is significant. Also, automate simple processes using if/then logic and complex processes using predictive models and scores. And finally, don't automate decisions in a volatile business process, because this will require you to rewrite rules frequently.

Challenges in Deploying Operational BI

There are many business and technical challenges to surmount when implementing operational BI, especially as data latency drops: there are more and diverse users to support, larger volumes of data to capture and deliver in near real time, and less time to recover from a processing error. In addition, queries must return in seconds rather than minutes or hours, and reports must update dynamically. What's more, all this extra processing power costs money and requires business users to be clear about their data freshness and uptime requirements.

When given a list of business and technical challenges to implementing operational BI, survey respondents ranked "architecting the system" at the top, followed by "managing expectations," "query performance," and "obtaining funding." Farther down the rankings were "technology limitations," "scalability," "finding experts," and "ease of use." Surprisingly, respondents ranked "system availability" in ninth place, followed only by "finding the right products" and "estimating cost." (See Figure 16.)

Ranked challenges to operational BI

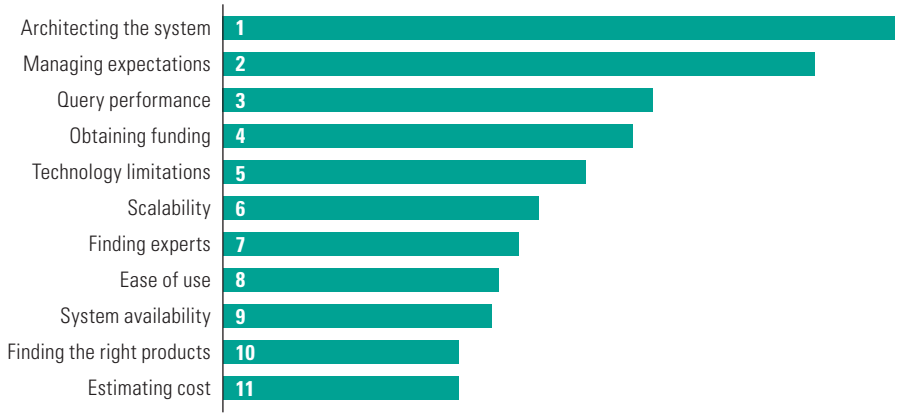


Figure 16. Respondents were asked to rank the top five challenges from the list above. The bars represent the proportional weights of the rankings.

Business Challenges

The business challenges discussed in this section revolve around gathering organizational support and funding to deliver just-in-time data through a traditional data warehousing environment.

Define requirements. A big problem with operational BI is that business users ask for just-in-time data to manage operational processes without having a clear business imperative or a true understanding of the costs and time involved. Some business managers request fresher data because they have an occasional need for it or they've heard that their competitors have implemented operational BI. Ambitious IT professionals are often too eager to help business users yield to this temptation. One BI professional interviewed for this report is designing an operational BI system even though she couldn't specify the business applications driving the request.

"We built real-time capabilities into our data warehouse because we knew users would eventually ask for it."
—Alicia Acebo

On the other hand, a BI professional with a strong knowledge of her organization and industry can anticipate requirements for just-in-time information before the business asks for it and architect solutions that are operationally ready. "We built real-time capabilities into our data warehouse from day one because I knew users would eventually ask for it," says Alicia Acebo, former data warehousing director at Continental Airlines. Acebo spent much of her career building online reservation systems and is now president of Rock Paper Data, a BI consultancy in Orlando, FL. Specifically, Acebo and her team wrapped a batch extract and load process around Continental's message queueing infrastructure that transfers data among the firm's operational systems. Then it waited.

"After the terrorist attacks on 9/11 in 2001, our management was ready to move from daily to real-time (in seconds) feeds of the data warehouse. The whole process took one week," says Acebo. Subsequently, Continental received an award from the FBI for its role in helping track down terrorists because of the timeliness of information it provided the agency.

Identify costs. By definition, operational BI delivers information about events as they happen or soon thereafter so business users can work proactively to optimize business processes. However, the costs of implementing an operational BI system increase as data latency drops. (See Figure 17.)

Data latency versus value and costs

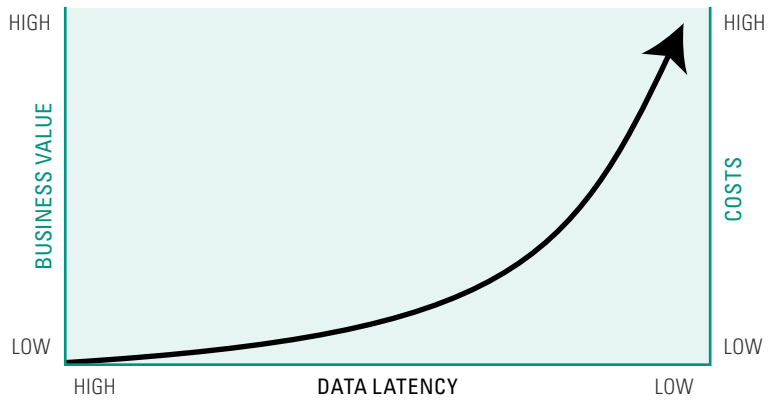


Figure 17. Business value and costs increase as the latency of the data drops in operational BI systems.

Low-latency systems require organizations to purchase new tools to capture transactions in real time, then transport and load them into the data warehouse. Many companies purchase data replication and changed data capture (CDC) products or upgrade their ETL tools to interface with message queuing backbones and enterprise service buses so they can "trickle feed" data into the data warehouse. They also parallelize their ETL and database management systems to increase throughput, or in some cases, they may write their own ETL scripts. Many also use specialized transformation software, such as Syncsort's DMExpress, to sort, aggregate, merge, calculate, and load data into a data warehouse and ensure high performance in a right-time, data delivery environment.

Operational BI systems cost an average of $1.1 million to build, including software, hardware, services, and staff.

TDWI research shows that operational BI systems cost an average of $1.1 million to build, including software, hardware, services, and staff. The bulk of this money, however, is tied up in staff costs. "The hardware and software costs are incremental, so the real expense is staff time," says Eric Lofstrom, who built an operational BI system for Quicken Loans in 2003 that cost approximately $1 million and took about 12 months to build. Lofstrom's team of six systems analysts and developers reused the company's enterprise messaging backbone, relational databases, and OLAP tools to create a hybrid data warehousing architecture that blends analytical and operational components to support dozens of operational and tactical dashboards as well as traditional OLAP desktops.

On the other hand, Blue Cross and Blue Shield of North Carolina reports that the cost of going "real time" is negligible. It only had to create two privacy tables that it updates in real time as soon as it receives HIPAA privacy forms from customers. "We already had the replication software, so our only costs were staff time to design two tables. And since our [Teradata] data warehouse is designed in third normal form, this was straightforward to do," says Naveen Narra, enterprise information architect of the company's corporate data warehouse.

Rethink business processes. The cost of operational BI goes well beyond the dollars and cents required to purchase more hardware, software, services, and support. Companies must reengineer business processes to exploit just-in-time data. For example, it's no use providing store managers with hourly sales data if they can only change prices or store displays once a week. In some cases, the systems are simply catching up with business processes that have been working in "real time" but without adequate data. For example, the "real time" BI system at Werner Enterprises will consolidate information about truck movements that managers currently patch together from multiple screens in the transaction system, says Kirchner.

In other cases, the availability of real-time data will completely change the way the organization does its work. For example, the security department at Continental Airlines researched fraudulent activity by following tips they received from the field. Today, the department functions much more proactively and effectively thanks to software "agents" that look for fraudulent activity by monitoring various data elements in the data warehouse. "The data warehouse has completely changed their processes and enabled them to recover millions of dollars in fraud," says Acebo.

Manage expectations. As in any BI project, it's important to set reasonable expectations so business users aren't surprised by the results. The temptation for IT professionals engaged on a bleeding-edge project is to overestimate what the team can deliver. Lofstrom says his team promised 15-minute updates, but given the state of the technology at the time (2003), it would have been wiser to have promised 30-minute updates. Although his team met their goals, it was not without some physical wear and tear. "Our windows were too small and, as a result, we didn't get much sleep and the team suffered burnout," he says.

"Our [update] windows were too small and, as a result, we didn't get much sleep and the team suffered from burnout."
—Eric Lofstrom

Train users. Organizations often need to retrain users to understand how to use and interpret just-in-time data. They may also need to create new incentives so workers are motivated to exploit the new information. "Many users don't want the data to change underneath them," says Acebo. "It's important to show them how to create a report as of a certain time and date." Although the data warehouse at Continental Airlines is updated in seconds, Acebo recommended that users not issue analytical queries for data less than five minutes old. "Otherwise, they may get inconsistent results if they submit a query before an update has completed."

Since many IT shops turn off referential integrity and aggregation routines and bypass data cleansing programs when loading just-in-time data, organizations need to caution users about its quality and consistency. Users need to recognize that this data hasn't gone through a formal reconciliation process and won't match numbers in the company's system of record until the next day.

Technical Challenges

To warehouse or not? The first technical question that an organization needs to address is whether to use a data warehousing architecture to deliver just-in-time data, or bypass it altogether. This is the most fundamental question, and the hardest to answer.

Many architects believe it's critical to adapt existing data warehouses to support operational BI. "The big showstopper is whether you are going to apply the same business rules to integrate, cleanse, and validate operational data streams as the rest of the data in your data warehouse," says John O'Brien, a BI consultant and former data warehousing architect. Pulling operational streams out of the data warehousing process undermines data quality and creates divergent data sets that may not reconcile, he claims.

But some disagree. They say that a data warehouse becomes a bottleneck if you try to load all data into it that users may possibly want to query. BI vendors such as Business Objects, Hyperion, SAS, and InetSoft, which support federated query, believe their tools provide an easy, low-cost way to capture real-time data and deliver it to users at the point of need with sufficient quality. Likewise, vendors of embedded BI, event-driven analytic platforms, composite applications, and in-memory analytics believe their offerings provide the most suitable way to meet high-end operational BI requirements. These vendors say that a data warehouse is critical for applying historical context to real-time data, but not necessary for managing the real-time data itself.

Organizations won't jettison their DW without trying to adapt it to deliver just-in-time data and support operational processes.

Nevertheless, data warehousing has become a well-established IT practice in corporate environments, and few organizations will jettison their DW investments without trying to adapt the architecture to support just-in-time data and operational processes. Our survey also shows that about half of organizations (51%) run both operational and analytical reporting from the same environment.

Is your operational BI environment the same as your regular BI environment?

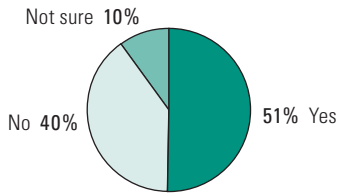


Figure 18. Based on 225 respondents that have implemented operational BI according to TDWI's definition.

Given the high number of survey respondents who equate operational reporting with operational BI, this indicates that many companies have moved operational reporting into their data warehousing environment. (See Figure 18.)

Organizations that want to deliver just-in-time data via a data warehousing environment face daunting challenges. The rest of this section describes the major challenges involved in transforming a data warehousing environment to support operational BI.

Selecting the right technology. There are many technologies that BI architects can use to build an operational BI environment. These technologies can be classified into three main categories that correspond to the way data flows through a just-in-time system:

- **Data acquisition.** Organizations must capture, transform, and move data from source systems into the analytical environment on a just-in-time basis. To do this, organizations can use ETL tools, replication tools, changed data capture technologies, messaging backbones, event-driven streaming engines, and BI and EII query tools. In practice, most companies use a combination of these.
- **Data storage.** Organizations can store acquired data in a variety of database engines, including one or more layers in a data warehousing environment (i.e. staging area, data warehouse, data mart, OLAP cube database), an ODS, a low-latency database, or an event-driven analytic engine.
- **Data delivery.** To display captured or data derived from analytic processes, organizations use custom-built applications, portals, BI tools, dashboards, composite applications, or a combination of these approaches.

Some of these technologies support multiple tasks. For instance, BI and EII tools can both acquire and deliver just-in-time data, and if you consider a report definition a storage mechanism, they support all three activities. Likewise, event-driven analytic platforms, working in concert with messaging backbones (EAI networks, enterprise service buses, publish/subscribe middleware), support all three activities. This is not surprising, since they are designed to provide end-to-end support for real-time operational BI requirements.

Our survey shows that organizations use a variety of techniques to support operational BI. The most common are ODSs, followed by more frequent batch loads and changed data capture. (See Figure 19.)

Rate the importance of the following techniques to your operational BI strategy.

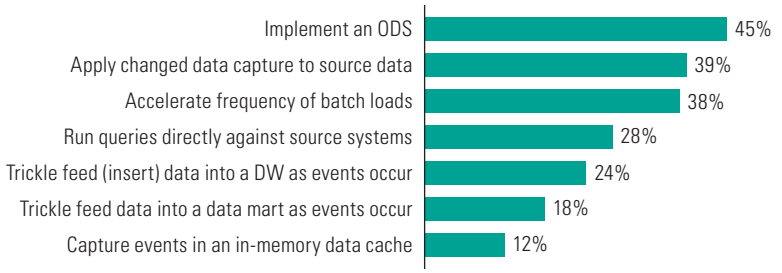


Figure 19. Based on 225 respondents who have implemented operational BI and rated the approach of "high" importance.

Increase scalability and throughput. Another key issue is building systems that scale to support greater numbers of users, more data sources, and higher volumes of data with increased rates of throughput—while ensuring high data quality and security. Organizations need to implement high-performance data warehousing platforms from leading companies such as Sybase, Teradata, and HP, and robust data integration platforms like those from Syncsort and others. To ensure scalability and throughput, companies may need to upgrade their networks and hardware, and parallelize key extraction, transformation, and load processes so they can adequately meet these increased throughput demands and eliminate processing bottlenecks.

For example, the Congressionally mandated TRICARE Encounter Data (TED) System of the Department of Defense Military Health System (MHS) collects, verifies, and tracks billions of dollars annually in claims made by more than 9 million health plan subscribers. The system processes millions of claims a day worth more than $13 billion. Claims data is pumped nightly into an ODS that is used to monitor, process, and audit the claims management process using thousands of business rules that are changed frequently. The ODS, in turn, populates a data warehouse that feeds downstream data marts used to analyze claims trends to detect fraud and abuse.

"One of our biggest challenges was to reduce the initial TED system's processing time through parallelization," says Joe Russell, director of software development at ABSi, a defense contractor that built the system and now runs it for the federal government. By parallelizing its database and ETL processes and upgrading its cluster of Unix servers, ABSi reduced nightly processing time from an average of 18 hours to 6. This not only enables same-day monitoring and analysis by the government, but also speeds up the reimbursement process.

"Avoid updates at all costs."
—John O'Brien

Use inserts. To increase ETL throughput, some (but not all) architects recommend inserting new records (i.e. loading data) rather than adding, changing, or deleting existing records. "Avoid updates at all costs," says O'Brien. "With large data sets, it takes too much time to find the record and update it." Although this creates lots of duplicate data, administrators can use SQL distinct and group-by statements to identify and delete duplicates, O'Brien says.

O'Brien also recommends turning off referential integrity and database logging. "Why log inserts if you have the original load file?" he says. In addition, there is no reason to create indexes and aggregations during the ETL process, since they consume a lot of CPU and slow down the inserts. Like many architects, O'Brien now advocates using the database to perform these tasks once the data is loaded into the data warehouse to exploit the parallelism of the underlying database. This process is known as extract, load, and transform (ELT) versus the more traditional extract, transform, and load (ETL).

Increase availability and recoverability. In a just-in-time data warehousing environment there is little time to recover from errors, because batch windows are small or nonexistent. "In a real-time system, you have no time to fix problems, and messages can start piling up," says Lofstrom. The best way to avert these problems is to build high-availability systems to drive the acquisition and deployment process. These systems parallelize single points of failure, run in a clustered environment, support failover and backup processes, and have an offsite backup.

For example, the Hong Kong Hospital Authority, which manages 159 public clinics and hospitals in Hong Kong, built a system in 2001 that provides an integrated view of patient clinical records on top of a right-time data warehouse updated every 15 minutes. The team realized that any system downtime would affect the accuracy of patient records. As a result, the team built a hot standby

system that contains a mirror copy of all data, duplicate connections to all sources, and staging servers. The failover system can resume full ETL processing with 30 minutes of an outage.

Likewise, ABSi implemented a failover capability to handle system processing in the event of a hardware failure within TED's primary server complex. The company is also in the process of implementing an offsite disaster recovery or continuity of operations site that can go live within 48 hours after a major outage at TED's primary data center.

To avoid losing data, John O'Brien recommends using "microbatches."

Microbatches. To avoid losing data, John O'Brien recommends using "microbatches" instead of real-time streaming, since microbatches provide a buffer with which to recover from a network or server outage. His rule of thumb is to set the ETL batch cycle three times larger than the time required to process the data. "If it takes 90 seconds to process a file, then I'll set the batch cycle to five minutes and have 3.5 minutes of sleep time."

However, other IT practitioners believe it is perfectly safe to stream events into a data warehouse. Lofstrom, for example, recommends logging each message as it comes off a queue, which may slow down the speed of the streaming data, but prevents the loss of data caused by server outages. If an outage occurs, Lofstrom recommends loading accumulated messages into the warehouse in a single batch job before turning event streaming back on. In other cases, companies use two-phase commit to update streamed messages from a queue into a staging area, so updates are never lost if there is a planned or unplanned outage in the database server.

Deliver adequate performance. A challenging problem for data warehousing designers is ensuring rapid query response times while simultaneously loading or updating a data warehouse and performing other tasks, such as monitoring events, triggering alerts, running backups, and scoring models. A mixed workload can cause RDBMS performance to degrade to the point where it blocks incoming queries or causes load processes to fail. Achieving adequate performance in a mixed workload environment can be a vexing challenge.

Users expect near-instantaneous response times in an operational BI environment.

Complicating matters is that most users expect near-instantaneous response times in an operational BI environment. For example, customer service representatives can't wait more than a few seconds for an integrated view of customer activity and cross-sell offers before their callers get impatient and hang up. Most data warehouses aren't designed to support a mixed workload of tactical queries that return a few rows of data in less than a second along with complex requests that scan millions of records and take minutes or hours to process.

So how do you balance these mixed workloads and maintain SLAs without causing the system to crash? There are several options, all of which have trade-offs:

- **Throttle incoming requests in response to heavy system loads.** You can schedule queries or jobs to only run at night, or queue or delete them using query governors or job managers during periods of heavy load to ensure the integrity of system updates. This approach obviously won't sit well with users, and it's the reason companies began building data warehouses in the first place: to provide the business with a dedicated analytical environment that would not interfere with operational processes.
- **Separate analytical and tactical queries and jobs.** Another approach is to off-load tactical queries and predictive scoring jobs to an ODS or specialized data mart and leave the data warehouse to support long-running analytical queries. Here, the ODS contains current transaction data, usually no more than 30 days' worth, stored in third normal form models

that reflect operational schemas. The downside of this approach is that it separates historical and current data into different systems, making it difficult for users to view just-in-time data in historical context.

- **Leverage RDBMS mixed-workload capabilities.** Many database management systems (DBMS) have significantly improved their ability to optimize the performance of long-running strategic queries, short-running tactical queries, load and update processes, and other types of workloads. Teradata, for example, has made big improvements in its priority scheduler for optimizing tactical queries, says Dan Graham, marketing manager at Teradata. The priority scheduler is a real-time traffic cop that always allocates CPU to tactical queries first to guarantee sub-second response times. It continuously monitors in-flight tasks and dynamically adjusts CPU allocation to optimize performance. In addition, administrators can configure and prioritize workloads by user groups, types of activity, and other variables. HP's new Neoview data warehousing platform also boasts workload management capabilities.

An RDBMS that supports mixed-workload processing lets organizations have their cake and eat it, too. But evaluate carefully.

An RDBMS that supports mixed-workload processing lets organizations have their cake and eat it, too: they can load current and historical data into the same database and optimize performance across all types of queries. Without an RDBMS that supports mixed workloads, many organizations decide to avoid using a data warehouse for operational BI. However, not all mixed workload capabilities are created equal, so evaluate your database vendor's capabilities carefully. Also, running multiple workloads on a single platform may require hardware upgrades to maintain adequate performance, so calculate these additional costs when deciding whether to use a data warehouse to support operational requirements.

Avoid contention problems. When user queries and ETL processes hit the same table, there is the potential for one process to block the other and cause query or load performance to degrade. Along with mixed workload requirements, the issue of contention causes many people to maintain historical and just-in-time data in distinct data stores. There are three ways to minimize the contention problem, but they are not for the faintheated:

- **Allow simultaneous reads and writes.** If you are inserting rather than updating data, the database won't lock the table. The only problem here is that the data can be out of sync if the user reads the table before all the inserts have completed. To avoid confusion, it's imperative to provide users with guidelines for submitting ad hoc queries against real-time data, and it helps if you dynamically time-stamp all reports. It's also important to turn off caching in BI servers, since there is no point in loading data every hour if the BI tool requests data from a cache that gets flushed once a day.
- **Insert and flip partitions.** Another approach is to create a duplicate partition of a fact table and load it with current data. Then, on a periodic basis, swap the table with the live fact table to update the data warehouse. This approach may require you to pause the application server while the flip takes place so no new requests are initiated while the swap occurs, says Justin Langseth, CTO of Claraview, a BI consultancy in Reston, VA. Many organizations use this "insert-and-flip" approach.
- **External real-time cache.** Another approach Langseth recommends is to load data into an in-memory cache outside of the data warehouse. Requests for real-time data are fulfilled from the cache, while queries that require a combination of real-time and historical data

merge the requested data in a set of temporary tables in either the data warehouse or real-time cache, depending on which database holds the largest volume of data requested. This type of just-in-time merging of information requires complex SQL and may be challenging for most BI tools to support.

For more on how to avoid contention issues in a just-in-time data warehouse, read Langseth's excellent article, "Real-Time Data Warehousing: Challenges and Solutions." (http://dssresources.com/papers/features/langseth/langseth02082004.html)

Synchronize aggregates. A problem related to query contention is when users query aggregation tables that are not dynamically recalculated with real-time data. Most organizations wait until the end of the day to recalculate aggregates with data in real-time partitions. Again, this requires administrators to educate users about the time sensitivity of the data they are querying. However, if the data volumes are small, it is possible to recalculate aggregates dynamically in a separate partition, and once the calculation has completed, expose updated aggregates using the insert-and-flip technique just described.

Summary. There are many challenges involved in implementing operational BI. The challenges discussed here deal primarily with transforming a data warehousing architecture to support just-in-time data delivery. Although there are other ways to deliver just-in-time data to support operational BI, many architects recommend using a data warehousing environment to ensure a consistent set of data for the entire organization.

Recommendations

Operational BI encompasses a broad set of applications, approaches, and technologies. While this makes it hard to provide advice that covers every potential use, these nine recommendations can help steer you in the right direction.

1. **Don't maintain a rigid boundary between analytical and operational processes and technologies.** As time windows shrink and the velocity of business increases, traditional boundaries between analytical and operational processes and technologies start to fade. Operational BI is more of a mindset than a specific set of technologies or architectural approaches, says William McKnight, senior vice president of data warehousing at Conversion Services International, a BI and data management consultancy. "To support operational BI, you need to think of operations first. You need to consider a smorgasbord of technologies—basically, anything that can add intelligence to a business process." For instance, McKnight says he now often uses database triggers, message queuing systems, EII tools, and Web services rather than data warehousing or BI tools to monitor and manage operational processes.
2. **Ascend the levels of operational BI.** We've outlined four levels of operational BI in this report. To start, organizations need to decide whether to run operational reports within the data warehousing environment or in a separate environment managed by another team. Next, organizations should experiment with higher levels of operational BI such as operational dashboards, embedded BI, or event-driven analytics, in response to or anticipation of business requirements.

3. **Put just-in-time data in historical context.** Most users need to view real-time data within historical context. If possible, adapt an existing data warehouse to deliver real-time data to maintain a single, consistent source of information for both real-time and historical data. Otherwise, make sure that your real-time environment can capture, assimilate, and deliver suitable historical data.

4. **Strive for architectural simplicity.** There is a temptation with operational BI to build complex processing environments with a lot of new technologies and moving parts. However, sticking with tried and true technologies and adapting them to deliver just-in-time data is often the best recipe for success. The most resilient architectures are simple and straightforward.

5. **Understand the costs of operational BI.** Recognize that costs increase as the latency of data decreases. Operational BI systems are mission-critical, and an outage can cost an organization severely. Therefore, you need to beef up the resiliency and recoverability of the data delivery infrastructure. To accommodate high-speed data flows and fast query performance, you'll need to boost processing speeds through faster hardware and parallelized software.

6. **Simplify delivery.** In an operational BI environment, it's important to deliver data to users in the most easily digestible form possible. Avoid complex reports and overly cluttered dashboards. You may need to create a custom GUI to deliver data in a straightforward way that makes sense for the process or processes that users are monitoring and managing.

7. **Close the loop.** Organizations that use dashboards to measure and monitor business processes gain a better understanding of how their businesses work and what actions to take in response to new or repeat events. In short order, these organizations can use this knowledge to create intelligent alerts that notify the right people about an impending problem, supply relevant reports to view, and suggest next steps to take.

8. **Incorporate analytics to streamline processes.** Applying predictive models to operational processes can improve efficiency. For example, a retailer that creates sales forecasts by applying predictive algorithms to current and historical sales data can better optimize inventory and minimize stock-outs than a retailer that relies on gut intuition, or on current activity or historical data alone. Analytics can also improve the sophistication of rules that automate actions in response to a stream of incoming business events.

9. **Automate with caution.** Automated systems can easily get out of control, flooding users with unwanted alerts or triggering inappropriate actions. To avoid these problems, automate only well-known business processes with a high volume of decisions that require immediate action. It's likely that you will need to apply predictive models to improve the accuracy and relevance of the alerts. This increases the sophistication and cost of the decision automation system.

Business Objects™

Business Objects has been a pioneer in business intelligence (BI) since the dawn of the category. As the world's leading BI software company, Business Objects transforms the way the world works through intelligent information. The company helps illuminate understanding and decision-making at more than 43,000 organizations around the globe. Through a combination of innovative technology, global consulting and education services, and the industry's strongest and most diverse partner network, Business Objects enables companies of all sizes to make transformative business decisions based on intelligent, accurate, and timely information.

www.businessobjects.com

hp invent

HP focuses on simplifying technology experiences for its customers—from individual consumers to the largest businesses. With a portfolio that spans printing, personal computing, software, services, and IT infrastructure, HP is among the world's largest IT companies, with revenue totaling $97.1 billion for the four fiscal quarters ended April 30, 2007.

The HP Neoview data warehousing platform is an integrated system that cost-effectively provides the dependability, performance, and functionality to meet high-end enterprise data warehousing needs. A key component of a comprehensive BI portfolio of products and services, it leverages technology and innovation across HP, and offers the full range of HP global services.

www.hp.com

ORACLE® | Hyperion®

Hyperion Solutions Corporation, a wholly owned subsidiary of Oracle, is the global leader in enterprise performance management (EPM) software. More than 12,000 customers in 90 countries rely on Hyperion both for insight into current business performance and to drive performance improvement. With Hyperion software, businesses collect, analyze, and share data across the organization, linking strategies to plans and monitoring execution against goals. Hyperion integrates financial management applications with a business intelligence platform into a single management system for the global enterprise.

www.hyperion.com

InetSoft
open standards innovation

InetSoft operational BI software offers a unique approach in delivering the right information to the right people—when and how they need it. inetSuite™ is a highly scalable, full-function operational BI suite that utilizes an innovative patent-pending technology to create secure, performance-tuned "Atomic Data Blocks" for assembly, mash-up, reporting, visualization, and collaboration by domain experts and business users across the enterprise. InetSoft zero-client, open standards, Web-based solutions allow companies to improve operational productivity and business performance through easy integration with business applications and processes. InetSoft delivers innovative, award-winning operational BI and enterprise reporting solutions to over 3,000 customers worldwide.

www.inetsoft.com

sas

SAS® business intelligence gives you THE POWER TO KNOW® by integrating data from across your enterprise and quickly transforming that data into shared insights. With SAS, IT maintains control of the data to ensure quality and consistency throughout the organization while providing business users the right information, at the right time, in the right format. SAS transforms decision making. We embed analytics in operational processes, providing fast, simple, and complete context to BI. SAS' unmatched BI platform provides data integration, analytics, BI, targeted business solutions, and specific industry solutions that take your organization Beyond BI™.

www.sas.com

SYBASE®

Sybase is the largest global enterprise software company exclusively focused on managing and mobilizing information from the data center to the point of action. Our highly optimized analytics server, Sybase IQ, is designed specifically to deliver dramatically faster results for mission-critical business intelligence, analytics, and reporting solutions on standard hardware and software platforms. Sybase IQ delivers unsurpassed query performance and storage efficiency for structured and unstructured data. Sybase IQ combines extraordinary speed and agility with low total cost of ownership, enabling enterprises to perform analysis and reporting that was previously impossible, impractical, or cost prohibitive.

www.sybase.com

syncsort

Syncsort Incorporated is a leader in performance innovation. For nearly 40 years, Syncsort has built a reputation for superior product performance and reliable technical support.

DMExpress™ 3.0, Syncsort's latest innovation, is the high-speed data integration tool specifically designed to meet the challenges of ETL data warehousing, BI, and other mission-critical applications. When dealing with large data volumes, DMExpress replaces long-running processes so that applications run faster on less expensive machines. With hours, even days of time saved, and a lower total cost of ownership, DMExpress can add tangible business value to your organization.

www.syncsort.com

Teradata
a division of NCR

Teradata, a division of NCR, is the global leader in enterprise data warehousing and enterprise analytic technologies and services. Organizations use Teradata® solutions to gain a single, integrated view of their businesses so they can make better, faster decisions that drive growth and profitability. Companies leverage this integrated source of consistent, accurate data to identify opportunities to increase revenues, decrease costs, and improve relationships. Teradata customers are industry leaders and include 60 percent of the top global companies. With our proven expertise and best-in-class technology, Teradata can help you see your business like never before.

www.teradata.com

TDWI RESEARCH

TDWI Research provides research and advice for BI professionals worldwide. TDWI Research focuses exclusively on BI/DW issues and teams up with industry practitioners to deliver both broad and deep understanding of the business and technical issues surrounding the deployment of business intelligence and data warehousing solutions. TDWI Research offers reports, commentary, and inquiry services via a worldwide Membership program and provides custom research, benchmarking, and strategic planning services to user and vendor organizations.

1201 Monster Road SW
Suite 250
Renton, WA 98057

T 425.277.9126
F 425.687.2842
E info@tdwi.org

www.tdwi.org