FOURTH QUARTER 2009

TDWI BEST PRACTICES REPORT

# NEXT GENERATION DATA WAREHOUSE PLATFORMS

By Philip Russom

www.tdwi.org

tdwi
THE DATA WAREHOUSING INSTITUTE

## Research Sponsors

Aster Data Systems

HP

IBM

Infobright

Kognitio

Microsoft

Oracle/Intel

Sybase

Teradata

# NEXT GENERATION DATA WAREHOUSE PLATFORMS

By Philip Russom

## Table of Contents

## About the Author

**PHILIP RUSSOM** is the senior manager of TDWI Research at The Data Warehousing Institute (TDWI), where he oversees many of TDWI's research-oriented publications, services, and events. Before joining TDWI in 2005, Russom was an industry analyst covering BI at Forrester Research, Giga Information Group, and Hurwitz Group. He also ran his own business as an independent industry analyst and BI consultant and was contributing editor with *Intelligent Enterprise* and *DM Review* magazines. Before that, Russom worked in technical and marketing positions for various database vendors. You can reach him at prussom@tdwi.org.

## About TDWI

The Data Warehousing Institute, a division of 1105 Media, Inc., is the premier provider of in-depth, high-quality education and training 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 data warehouses. It also fosters the advancement of data warehousing research and contributes to knowledge transfer and the professional development of its Members. TDWI sponsors and promotes a worldwide Membership program, quarterly educational conferences, regional educational seminars, onsite courses, solution provider partnerships, awards programs for best practices and leadership, resourceful publications, an in-depth research program, and a comprehensive Web site (www.tdwi.org).

## About the TDWI Best Practices Reports Series

This series is designed to educate technical and business professionals about new business intelligence technologies, concepts, or approaches that address a significant problem or issue. Research for the reports is conducted via interviews with industry experts and leading-edge user companies and is supplemented by surveys of business intelligence professionals.

To support the program, TDWI seeks vendors that collectively wish to evangelize a new approach to solving business intelligence problems or an emerging technology discipline. By banding together, sponsors can validate a new market niche and educate organizations about alternative solutions to critical business intelligence issues. Please contact TDWI Research Director Wayne Eckerson (weckerson@tdwi.org) to suggest a topic that meets these requirements.

## Acknowledgments

TDWI would like to thank many people who contributed to this report. First, we appreciate the many users who responded to our survey, especially 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 production team: Jennifer Agee, Bill Grimmer, Denelle Hanlon, and Deirdre Hoffman.

## Sponsors

Aster Data Systems, HP, IBM, Infobright, Kognitio, Microsoft, Oracle/Intel, Sybase, and Teradata sponsored the research for this report.

## Research Methodology and Demographics

**Report Scope.** This report is designed for business and technical executives who are responsible for planning and implementing programs for data warehousing. This report is essentially a catalog of the many new technologies and techniques that have arisen for data warehouse platforms in recent years. The report's goal is to help data warehouse professionals and their business sponsors understand what options are now available and for which business and technology use cases they are best suited.

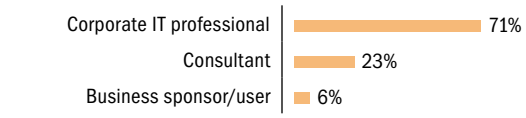
**Research Methodology.** Most of the market statistics presented in this report are based on a research survey. In May 2009, TDWI sent an invitation via e-mail to the data management professionals in its database, asking them to complete an Internet-based survey. The invitation was also distributed via Web sites, newsletters, and conferences from TDWI and other firms. The survey drew complete responses from 452 survey respondents. From these, we excluded respondents who identified themselves as academics or vendor employees, leaving the responses of 417 respondents as the core data sample for this report. The report also references data from several TDWI Technology Surveys.

**Survey Demographics.** The wide majority of survey respondents are corporate IT professionals (71%), whereas the remainder consists of consultants (23%) or business sponsors/users (6%). We asked consultants to fill out the survey with a recent client in mind.

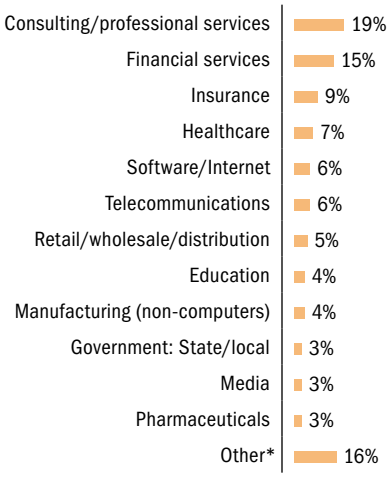
The consulting (19%) and financial services (15%) industries dominate the respondent population, followed by insurance (9%), healthcare (7%), software (6%), telecommunications (6%), and other industries. Most survey respondents reside in the U.S. (53%) or Europe (19%). Respondents are fairly evenly distributed across all sizes of companies and other organizations.

**Other Research Methods.** In addition to the survey, TDWI Research conducted numerous telephone interviews with technical users, business sponsors, and recognized experts in the field of data warehousing. TDWI also received product briefings from vendors that offer products and services related to the best practices under discussion.

### Position

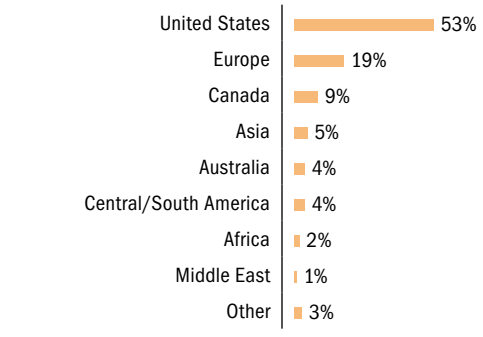



### Industry

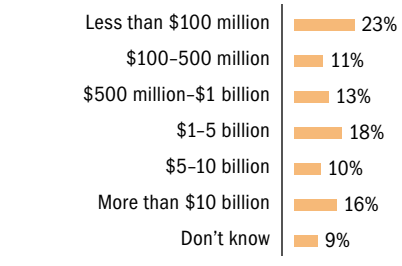



**"Other" consists of multiple industries, each represented by 2% or less of respondents.*

### Geography




### Company Size by Revenue




*Based on 417 survey respondents.*

# Introduction to Next Generation Data Warehouse Platforms

If you're a data warehouse professional—or you work closely with one—you've probably noticed the many new options for data warehouse platforms that have appeared this decade.

**Data warehouse platforms support many more options today than just a few years ago.**

We've seen the emergence of new categories of data warehouse (DW) platforms, such as data warehouse appliances and software appliances. A new interest in columnar databases has led to several new vendor products and renewed interest in older ones. Open source Linux is now common in data warehousing, and open source databases, data integration tools, and reporting platforms have come out of nowhere to establish a firm foothold. In the hardware realm, 64-bit computing has enabled larger in-memory data caches, and more vendors now offer MPP architectures. Leading database vendors have added more features and products conducive to data warehousing.

Those are mostly features within the data warehouse platform, especially its database. There are also growing practices that are demanding support from the platform, including real-time integration between the data warehouse platform and operational applications, various types of advanced analytics, and reusable interfaces exposed through Web services or service-oriented architecture (SOA). Furthermore, a number of data warehouse platforms and other business intelligence platforms are now readily available through software-as-a-service (SaaS) and cloud computing.

**Understanding which options to use, when, is a challenge.**

The good news is that the options for data warehouse platforms have recently become far more numerous. The bad news is that it's difficult for data warehouse professionals and their business sponsors to keep track of these advancements and select the ones that are appropriate for their needs.

To help organizations understand the many new options available to them, this report catalogs the new data warehouse platform products, features, and techniques that have appeared this decade, plus notable advances in more established data warehouse platforms. As examples, the report mentions many vendors and their products. From the survey data cited here, you'll see that many organizations are planning the next generation of their data warehouse, and this report provides information that can be instrumental for such planning. The focus is on technology, but this report also explains how technology's adoption in next generation data warehouse platforms is driven by real-world business and organizational needs and requirements.

## Definitions of Terms and Concepts

### Data Warehouse Platform

**A DW platform has five basic components: server, operating system, database, storage, and networking.**

For the purposes of this report, a data warehouse platform consists of one or more hardware servers, an operating system, a database management system (DBMS), and data storage. These communicate via a LAN or WAN, although a multi-node data warehouse platform may have its own specialized network. Note that a data warehouse platform manages a data warehouse, defined as a collection of metadata, data model, and data content, designed for the purposes of reporting, analyzing information, and making decisions. But the data warehouse is not part of the platform per se. (See Figure 1.) All these components and more have seen generational advances in recent years.

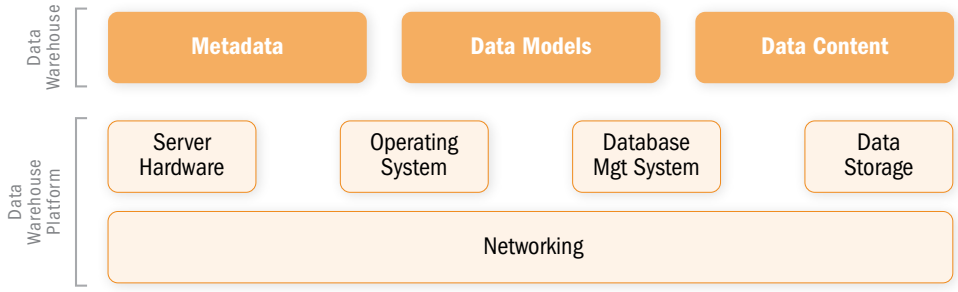



***Figure 1**. A data warehouse platform manages a data warehouse, but the two are separate.*

### Generations of Data Warehouses

TDWI's position is that certain relatively new technologies, techniques, and business practices are driving the majority of data warehouses and their platforms toward a redesign, major retrofit, or even replacement that we can recognize as a generation. TDWI takes the term literally, meaning that the current generation of a data warehouse will beget the next generation. In many cases, generational change is an evolutionary process that adapts the resulting data warehouse to changing business and technology requirements. In fact, generational change is often driven by these requirements, as is explained in detail in the next section of this report. In other cases, generational change is more of a maturation process that steps a data warehouse through multiple stages of a lifecycle.

### Next Generation Data Warehouse Platforms

**Your next generation isn't necessarily leading edge.**

What's next for a given organization's data warehouse platform can vary tremendously. For example, a next generation data warehouse platform may tap into leading-edge features, such as appliances, open source, and cloud computing. It may simply get you caught up with somewhat more established practices for real-time operation, advanced analytics, and services. Sometimes, the next generation addresses administrative issues, such as hardware upgrades (from 32-bit to 64-bit), data migrations (from one DBMS to another) or architectural changes (from SMP to MPP). So, let's keep in mind that a next generation data warehouse platform is a relative concept, because it depends on where you're starting, what new requirements you must address, and how many resources you have.

## Why Care about Data Warehouse Platforms Now?

- **Businesses face change more often than ever before.** Recent history has seen businesses repeatedly adjusting to boom-and-bust economies, a recession, financial crises, and shifts in global dynamics or competitive pressures. Increasingly, businesses rely on the data warehouse and related business intelligence infrastructure to understand change and react appropriately.
- **DW platforms need updating to support changing business requirements.** In fact, many of the technologies associated with the next generation DW relate to change in some way, such as advanced analytics, scalable architectures, virtualization methods, reusable services, real-time integration with operational applications, and so on.
- **Successful DWs mature through multiple lifecycle stages.** This usually provokes changes in the underlying DW platform and elsewhere in the business intelligence (BI) infrastructure.

- **There's probably a new generation in your near future.** TDWI survey data shows that almost half of respondents are planning a data warehouse platform replacement in 2009–2012. Many others anticipate keeping their current platforms, but updating them significantly.

**USER STORY MANAGEMENT REQUIREMENTS OFTEN DICTATE THE DESIGN OF A NEXT GENERATION DW AND ITS PLATFORM.**

"We pulled together our current data warehouse a couple of years ago," said Karl Mikula, the data and BI manager at Hagerty Insurance Agency, America's leading provider of products and services for collectors of classic cars and boats. "Now that the company sees the value, we're building our next generation data warehouse and BI solution atop a platform that'll do what the company needs. In a nutshell, upper management wants to adapt a performance management methodology with scorecards. And they want self-service BI, where they can search a repository and pull data into reports or spreadsheets of their own design, presented through a corporate portal. To support this, we're designing a data warehouse that stores metrics and KPIs in a searchable repository. For the next generation platform, we have a database management system, a data integration tool, a reporting tool, a search engine, and an enterprise portal. All these come from Microsoft, and they're all tightly integrated out of the box."

# The Evolving State of Data Warehouse Platforms

## Technology Drivers for New Generations of Data Warehouses

There's more than one path to the next generation of your organization's data warehouse platform:

**For the next generation, you can update the old platform or replace it with a new one.**

**Retain the current platform, but do more with it.** Many users still haven't tapped all the capabilities of their current platforms. Many interviewed for this report spoke of early-to-mid-life project phases where it was time (defined by business readiness) to embrace the platform's more sophisticated capabilities, especially real-time functions, data federation, in-memory processing, and analytics (whether based on OLAP or data mining). Since there's a difference between a data warehouse and the platform that manages it, you can remodel the DW significantly to add value without replacing the platform. Although this report focuses on data warehouse platforms, some new generations involve tools that are tangential to the platform, such as solutions for data integration, quality, master data, reporting, and so on. Incremental additions to hardware are common (to add more CPUs, memory, or storage), and these satisfy next generation requirements (fast queries, in-memory databases, and scalability) by doing more with the current platform.

**Replace the current platform, then build out the new one.** Compared to other approaches, ripping out and replacing a data warehouse platform is rather expensive for IT budgets and intrusive for business users. Therefore, this path to the next generation should be avoided, in general. And yet, this is exactly what almost half of users surveyed by TDWI Research report they are contemplating. (See Figure 2.) Forty-six percent of respondents say they anticipate replacing their current primary data warehouse platform during the years 2009 through 2012. Another 5% will make replacements during 2013 through 2015. Note that roughly half of respondents (49%) have no plans to replace their current data warehouse platform.

**When do you anticipate replacing your current primary data warehouse platform?**

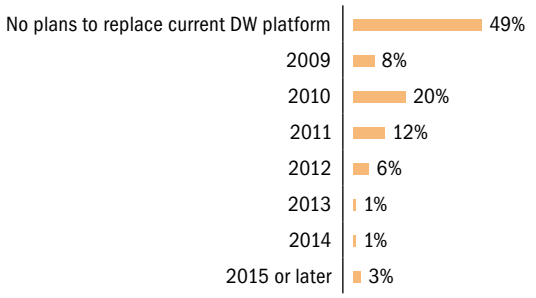



***Figure 2.*** *Based on 417 responses.*

**The next generation is enabled by hefty technology.**

To quantify why so many organizations are planning to replace their data warehouse platform, TDWI asked: What problems will eventually drive you to replace your current primary data warehouse platform? (See Figure 3.) If we group survey responses, five areas emerge where users are having problems with their current data warehouse platforms. Each of these problems is a technology driver that leads organizations toward a next generation data warehouse:

**Analytics of various types.** Many data warehouses have evolved (based on users' designs and usage) to become platforms for reporting or basic online analytic processing (OLAP). As these users try to move beyond reporting and OLAP, they find that their platform "can't support advanced analytics" (40%). There are multiple forms of advanced analytics, including those based on data mining or statistics and those based on complex ad hoc SQL statements. The former may or may not run in a DBMS (depending on the vendor's analysis tool capabilities), which is a problem when it forces users to move data out of the data warehouse for the sake of analysis, then back in. The latter methods (based on SQL) are hamstrung if the platform suffers "poor query response" (45%). Of course, many other DW and BI functions suffer when queries are slow.

**Real-time and related technologies.** An ongoing trend involves integrating an organization's data warehouse with its transactional and operational applications. Such efforts are stymied when the DW platform is "poorly suited to real-time or on-demand workloads" (29%). A number of issues relate to real-time operations. For example, a DW isn't real time or on demand if it's hampered by "inadequate high availability" (19%), "inadequate data load speed" (39%), or "inadequate support for Web services and SOA" (16%).

**Scalability, in many senses.** A DW platform can't cope with growth over time if it "can't scale to large data volumes" (37%) or it "can't support a large concurrent user count" (20%). Achieving these goals is difficult when the "cost of scaling up is too expensive" (33%).

**Addressable memory space.** One of the leading reasons for replacing old server hardware is because the "current platform is 32-bit, and we need 64-bit" (15%). The primary advantage of 64-bit hardware and software is its large addressable memory space. Hence, by comparison, 32-bit systems may suffer from "inadequate support for in-memory processing" (16%).

**Warehouse architecture and related practices.** Architectural issues voiced in users' survey responses include "we need a platform that supports mixed workloads" (21%), our "current platform is SMP, and we need MPP" (14%), and "we need a platform better suited to cloud or virtualization" (13%).

**What problems will eventually drive you to replace your current primary data warehouse platform? (Select nine or fewer.)**

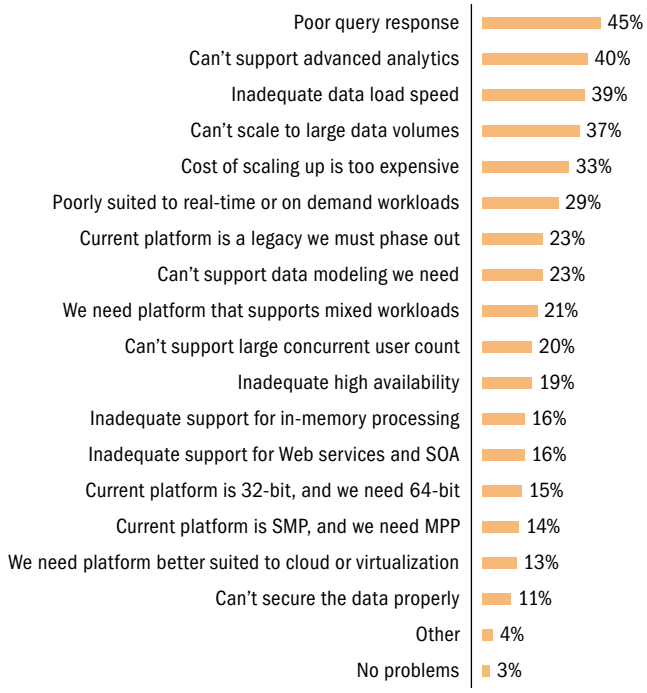



***Figure 3.*** *Based on 1,752 responses from 417 respondents; 4.2 responses per respondent on average.*

**USER STORY TECHNICAL DEFICIENCIES CAN MAKE AN OLD DATA WAREHOUSE PLATFORM RIPE FOR REPLACEMENT.**

"We fought our data warehouse platform for years, and we lost most of the battles," said a data warehouse architect. "Because of our unpredictable increases in data volume, we were constantly reconfiguring the server hardware and repartitioning the database. To maintain decent query response times, we were constantly re-orging and re-indexing databases and tables. The architecture of the data warehouse environment was a complicated mess, because we had lots of separate databases to enable multiple workloads, analytic methods, detail data storage, and data integration speeds. Any changes we'd make—and we made a ton of them—would ripple through all these complicated pieces, generating many hours of testing and bug fixing.

"Over time, the intense administrative work required of the platform took over our schedules, such that we couldn't do as much new development as the business needed from us. Once business management felt the pain, they took more seriously our pleas for a more modern data warehouse platform. As we begin our evaluations of products for our next generation data warehouse platform, we'll primarily be looking for databases and other components that can handle multiple workloads and real-time requirements in a single, seamless architecture. I'm not sure yet, but along the way, we'll probably move from SMP to MPP, from a scale-up to a scale-out hardware strategy, and from a relational database designed for transaction processing to one purpose-built for data warehousing."

## Business Drivers for New Generations of Data Warehouses

All five groups of technology drivers just discussed have real-world business situations behind them. In other words, most of the technology drivers that lead to a next generation data warehouse are actually IT's reaction to urgent business drivers:

**Technology drivers are really business drivers.**

**Analytics of various types help the business cope with change and discover opportunities.** According to this report's survey, 38% of organizations surveyed are practicing advanced analytics today, and 85% say they'll be practicing it within three years. The increasing use of advanced analytics is driven by organizations' need to understand constantly changing business environments, as well as to discover opportunities for cost reductions and new sales targets.

**Real-time technologies should support some form of operational excellence.**

**Real time and related technologies are enablers of operational excellence.** Because of economic, competitive, and quality issues, many organizations are under pressure to achieve unprecedented levels of organizational excellence. Many are pursuing this goal by embedding data and functionality available from their BI and data warehouse infrastructure into their operational and transactional applications. This enables time-sensitive business practices such as operational BI, on-demand dashboards and performance management, just-in-time inventory and manufacturing, and so on. Technology people talk about integrating BI and operational systems through real-time data warehousing, which they may call by other names such as active, dynamic, or on-demand data warehousing. But these are just enabling technologies that help satisfy more fundamental business requirements for operational excellence.

**Scalability, in many senses, enables business growth.** As businesses grow, as they automate more business processes with software, and as they depend more on BI and data warehousing, they generate more data that needs processing for BI and to run the business. Likewise, there's growth in user communities, reports, analyses, and so on. With the enterprise data warehouse at the heart of BI—and more and more at the heart of operational excellence—warehouse scalability has become a critical success factor across the board.

**Addressable memory space automates new time-sensitive business practices.** In particular, various types of in-memory databases can now be far larger than before, and data operations in memory are far faster than those that involve input/output with disks. Imagine putting an entire data mart or warehouse in memory; reporting and analysis functions are now so fast that they are more easily embedded in operational applications. This speedy intelligence takes business practices to a new level, such as up-sell/cross-sell guidance in a telemarketing scenario, automated recommendations in an e-commerce situation, improved service in call center and similar online applications, and fraud detection in a variety of contexts.

**Warehouse architecture and related practices affect a DW's ability to support a business.** Most of the technology and business drivers mentioned here involve creative adjustments to the logical architecture of the data warehouse's data model and the systems architecture of its hardware configurations. When moving to a next generation data warehouse platform, expect to make architectural changes to accommodate new technical functions and their related business requirements.

**USER STORY MOST BUSINESS TRANSFORMATIONS REQUIRE A TRANSFORMATION OF DATA WAREHOUSING AND BI.**

"Our BI and DW solutions were very successful for years, largely because they were conservative and old-fashioned," said an experienced BI director. "Everything changed when we got a new CEO who expects modern BI and DW solutions. Luckily, he doesn't mind sponsoring and funding them, and he's not shy about getting everyone on board.

"The first order of business was to give managers dashboard-style reports, to complement the static reports we'd provided for years. The second order was to tie the dashboards into operational systems for real-time and on-demand monitoring of the business. To support this mix of performance management and operational BI, we extended our history-oriented data warehouse to include real-time data for operational metrics and KPIs. To get fresh data into the warehouse, we replaced our hand-coded ETL with tool-based ETL, and we tied it into our message-oriented middleware. Long story short, this was a lot of work, but well worth it, because we've transformed the business into an agile, fact-based operation. To make the transformation sustainable, we extended our current data warehouse platform and related tools for BI and ETL, to accommodate new data warehousing practices, while also continuing to support older ones."

## Your Data Warehouse Today and Tomorrow

### Origins of Your Next Generation Data Warehouse Platform

Remember that almost half of survey respondents say they'll replace their current data warehouse platform by 2012. If they do, what strategies will they follow in assembling their next generation data warehouse platform? (See Figure 4.)

**Most users assemble their own DW platform, but this may change.**

**Today, most DW platforms are assembled by in-house personnel.** According to our survey, most data warehouse platforms are custom solutions created internally by members of IT or the data warehouse team (55%). Even so, the system integration required of custom solutions is a time-consuming distraction, which is why some user organizations offload it to consultants or system integrators (28%). Still other users turn to pre-integrated hardware/software bundles from vendors (8%), and a few early adopters are using data warehouse appliances (6%).

**Users are open to DW appliances and similar hardware/software bundles.** According to the survey results in Figure 4, if given the chance to replace a data warehouse platform, more respondents would go with an appliance (20%) than with consultants or a system integrator (16%). Users are likewise open to a vendor's pre-integrated hardware/software bundle (15%) for their next generation DW platform. However, a custom in-house design (44%) would still be the preference for most users choosing a new platform.

**For your organization's current, primary data warehouse platform, which of the following best describes its origins? For your organization's next generation data warehouse platform, what would you prefer that its origins be?**

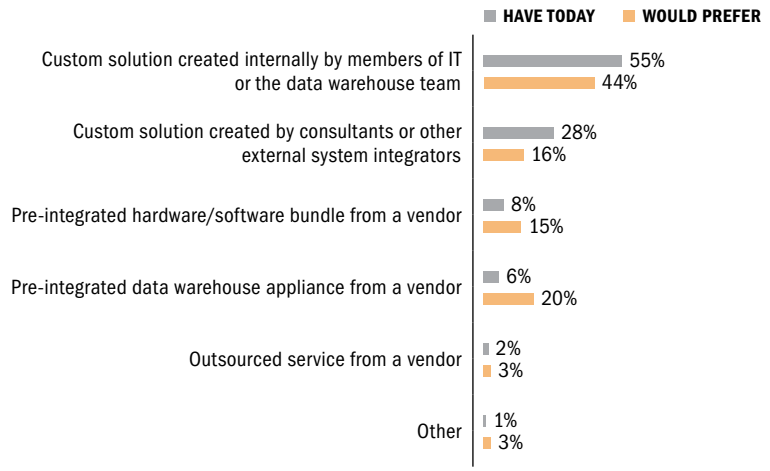



***Figure 4.*** *Based on 417 responses.*

## Data Volume Growth as a Generational Driver

Let's recall that one of the reasons so many users are planning to replace their current data warehouse platform is that it "can't scale to large data volumes" (37%, as seen in Figure 3). To quantify data growth, this report's survey asked: What's the approximate total data volume that your organization's data warehouse environment manages, both today and in three years or so? (See Figure 5.)

**The big get more numerous, and the small get less so.** As you might expect, the percentage of organizations surveyed with large DWs (3 terabytes [TB] or more) will increase substantially over the next three years. Likewise, the number of small DWs (less than 3 TB) in this survey population will decrease, because they'll move up to manage more terabytes.

**Scaling to large data volumes can force a new generation.**

**The "10-terabyte club" is set to double.** In recent years, many users have told TDWI Research that exceeding 10 TB is difficult given today's technology and user best practices. Many have admitted that a change of data warehouse platform (especially the DBMS) was required to complete the journey. The survey shows that the percentage of organizations in the 10 TB range will double in the next three years, from 17% to 34%. Hence, many of the 17% of surveyed organizations that will move into the 10 TB club within three years will probably need to replace their data warehouse platform to make the move possible and sustainable.

**What's the approximate total data volume that your organization's data warehouse environment manages, both today and in three years or so?**

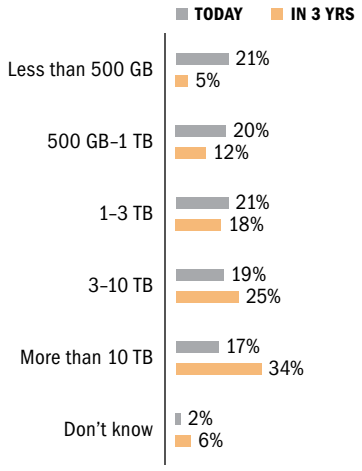



***Figure 5.*** *Based on 417 responses.*

## The Economy as a Generational Driver

Cost is an issue for almost everything in today's economic recession, including the funding of data warehouse platforms.

**Interest is rising for open source, SaaS, and clouds.**

**The recession has definitely affected BI and data warehousing.** When asked how the current economic recession is affecting data warehousing teams, respondents reported reduced budgets (57%), frozen hiring (41%), projects on hold (30%), frozen tool and platform acquisitions (25%), and even layoffs (19%). (See Figure 6.) Only 27% said there's no impact so far.

**Users and organizations are very open to low-cost DW platforms.** A whopping 57% of survey respondents said that yes, the recession has made them consider low-cost DW options more seriously, whereas only 34% said no. (See Figure 7.) This has spurred much more interest in open source software (especially databases), plus creative approaches to licensing (SaaS) and low-investment deployments (clouds). And, of course, many forms of hand coding are back with a vengeance.

**How is the current economic recession affecting data warehousing teams and projects in your organization? (Check all that apply.)**

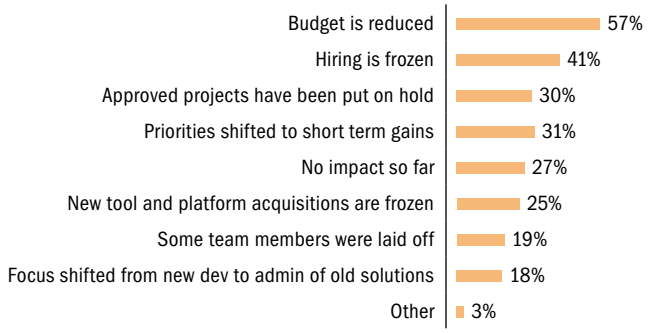



***Figure 6.*** *Based on 1,049 responses from 417 respondents; 2.5 responses per respondent on average.*

**Has the current economic recession made you or your organization more seriously consider low-cost options for data warehouse platforms?**

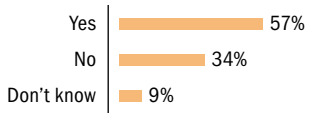



***Figure 7.*** *Based on 417 responses.*

**USER STORY** **NEW GENERATIONS MUST BALANCE PERFORMANCE AND COST.**

KnowledgeBase Marketing (KBM) is a marketing services provider that helps direct marketers find, know, keep, and grow profitable customers. "KBM hosts our clients' data about their customers and prospects for sales and marketing," said Dave Rajala, KBM's vice president of development. "Besides hosting, we also cleanse, augment, and analyze our clients' data, which helps them launch better targeted and more cost-effective direct marketing campaigns."

Like most service providers, KBM has deployed a lot of servers in its data center, typically several servers per client. "This worked fine for years," according to Randy Herzog, KBM's vice president of configuration services. "However, as our clients' needs continued to change, we realized we needed to find a different platform strategy for our hosted data business. We needed higher performance with a lower total cost of ownership for our clients. After looking at several platforms, we are now utilizing the HP Oracle Database Machine with the Exadata Storage Server for some of our most robust environments. It allows us to preserve client data integrity and safeguards, but they still get higher performance than with our older, standalone configurations. And, in our estimation, a large database appliance will cost less than multiple standalone machines. Plus, Oracle software is still a heavy standard in our shop, so our skills in Oracle Database and PL/SQL transfer well to the new platform."

# Quantifying Data Warehouse Generations

At this point in the report, we've defined the terms and concepts of next generation data warehouse platforms, and we've listed the drivers that push organizations into a new generation. Now it's time to start drawing the big picture by answering questions such as:

> What are the many features and techniques that users need to incorporate into the next generation of their data warehouse platform? Which ones are users adopting and growing the most? Which are in decline? At what rate is change occurring?

**New options for data warehousing are quite diverse in terms of maturity.**

To help quantify these and other questions, TDWI presented survey respondents with a long list of options for next generation data warehouse platforms. (See the left side of Figure 8.) These options include a mix of vendor-oriented product features and product types, as well as user-oriented techniques and best practices. The list includes options that have arrived fairly recently (clouds, SaaS, open source, appliances), have been around for a few years but are just now experiencing broad adoption (real-time data warehousing, advanced analytics, MDM, MPP), or have been around for years and are firmly established (DBMSs designed for transaction processing, SMP, EDWs). After all, generational change (when managed well) addresses features and techniques based on business requirements, not the vintage or novelty of available options.

**Usage of most new DW options is set to grow.**

Concerning the list of DW platform options presented in the survey, TDWI asked: "Which are you using today on or around your primary data warehouse platform?" To get a sense of how this will change over time, TDWI also asked: "Which do you anticipate using in three years or so?" Survey responses for these two questions are charted as pairs of bars on the left side of Figure 8. The "Delta" chart in the middle of Figure 8 simply shows the difference between responses for "Using Now" and "Use in 3 Years" to provide an indication of change over time—whether growth or decline—per DW platform option.

The survey question told the respondents: "Checking nothing on a row means you have no plans for using that DW option." This enables us to quantify the approximate percentage of user organizations surveyed that are committed to using a particular DW platform option, whether now, in the future, or both. These percentages are charted on the right side of Figure 8.

## Growth or Decline of Usage versus Breadth or Narrowness of Commitment

Figure 8 is fairly complex, so let's explain how to read it. First, Figure 8 is sorted by the Delta column, in descending order. In this sort, "real-time data warehousing" appears at the top of the chart, because—with a delta of 75%—this DW option has the greatest growth potential. However, not all organizations are planning to use this option. In the "Plan to Use" column, we see that 58% of survey respondents have committed to implement some form of real-time data warehousing. Among the 58% of organizations who've made a commitment to real-time data warehousing (whether now, in the future, or both), 17% (of the 58%) are already using real-time data warehousing today, and 92% anticipate using it three years from now.

**Some of the greatest growth rates are within small niches of the data warehouse community.**

From this, we see that there are two forces at work in Figure 8, as well as in the planning processes of user organizations:

- **Change.** The Delta chart indicates the amount of change in usage per data warehouse platform option we can expect over the next three years. The change may be growth or decline.
- **Commitment.** The Plan to Use chart quantifies the percentage of organizations that have made a commitment to use a particular DW platform option. The commitment may be realized today, sometime in the near future, or both.
- **Balance of Change and Commitment.** To get a complete picture, it's important to look at the metrics for both change and commitment. Some features or techniques may have significant growth rates, but within a rather small segment of the user community (e.g., private clouds and open source DBMSs). Or they could have low growth rates, but be strongly established (e.g., EDW, analytics within the EDW).

To help you visualize the balance of change and commitment, Figure 9 (page 16) includes the Delta and Plan to Use numbers from Figure 8 as opposing axes of a single chart. DW platform options are plotted on Figure 8 as a balance of increasing or decreasing usage (*x*-axis) and narrowness or breadth of commitment (*y*-axis).

**Note:** The rest of this report will refer frequently to Figure 8.

**For the techniques and technologies in the following list, which are you using today on or around your primary data warehouse platform? Which do you anticipate using in three years or so? Checking nothing on a row means you have no plans for using that DW option.**

| GENERATIONAL DW FEATURE OR TECHNIQUE | USING NOW | USE IN 3 YRS | DELTA | PLAN TO USE |
|---|---|---|---|---|
| Real-time data warehousing | 17% | 92% | 75% | 58% |
| Master data management (MDM) solution | 21% | 89% | 68% | 62% |
| Private cloud | 24% | 85% | 61% | 22% |
| In-memory database | 25% | 83% | 58% | 39% |
| 64-bit hardware and software | 25% | 78% | 53% | 58% |
| Open source DBMS | 32% | 80% | 48% | 17% |
| Trickle or streaming data loads | 34% | 82% | 48% | 41% |
| Service-oriented architecture (SOA) | 35% | 82% | 47% | 47% |
| Advanced analytics (e.g., mining, predictive) | 38% | 85% | 47% | 85% |
| Power-efficient hardware | 31% | 77% | 46% | 38% |
| Massively parallel processing (MPP) | 36% | 82% | 46% | 58% |
| Software as a service (SaaS) | 34% | 76% | 42% | 29% |
| Open source reporting or analysis tool | 36% | 77% | 41% | 26% |
| Open source data integration tool | 33% | 74% | 41% | 27% |
| Web services | 35% | 75% | 40% | 59% |
| Data quality tool | 42% | 82% | 40% | 69% |
| Public cloud | 41% | 78% | 37% | 14% |
| Software appliance | 40% | 72% | 32% | 19% |
| High availability for the DW | 46% | 78% | 32% | 66% |
| Mixed workloads | 46% | 72% | 26% | 43% |
| Data warehouse appliance | 53% | 78% | 25% | 47% |
| Column-oriented storage engine | 51% | 75% | 24% | 39% |
| Data federation | 50% | 72% | 22% | 43% |
| Data encryption and other security features | 51% | 73% | 22% | 55% |
| Open source operating system | 53% | 68% | 15% | 24% |
| DBMS purpose-built for data warehousing | 58% | 71% | 13% | 69% |
| Server virtualization | 60% | 70% | 10% | 41% |
| Central enterprise data warehouse (EDW) | 62% | 67% | 5% | 82% |
| Bundled DW platform from multiple vendors | 54% | 58% | 4% | 34% |
| Analytics processed within the EDW | 63% | 67% | 4% | 74% |
| Analytic databases outside the EDW | 66% | 60% | -6% | 61% |
| Blade servers in racks | 68% | 57% | -11% | 53% |
| Symmetrical multi-processing (SMP) | 85% | 41% | -44% | 36% |
| DBMS designed for transaction processing | 89% | 37% | -52% | 46% |

*Figure 8.*

**DW Platform Features and Techniques Plotted for Delta (Growth) and Plan To Use (Commitment)**

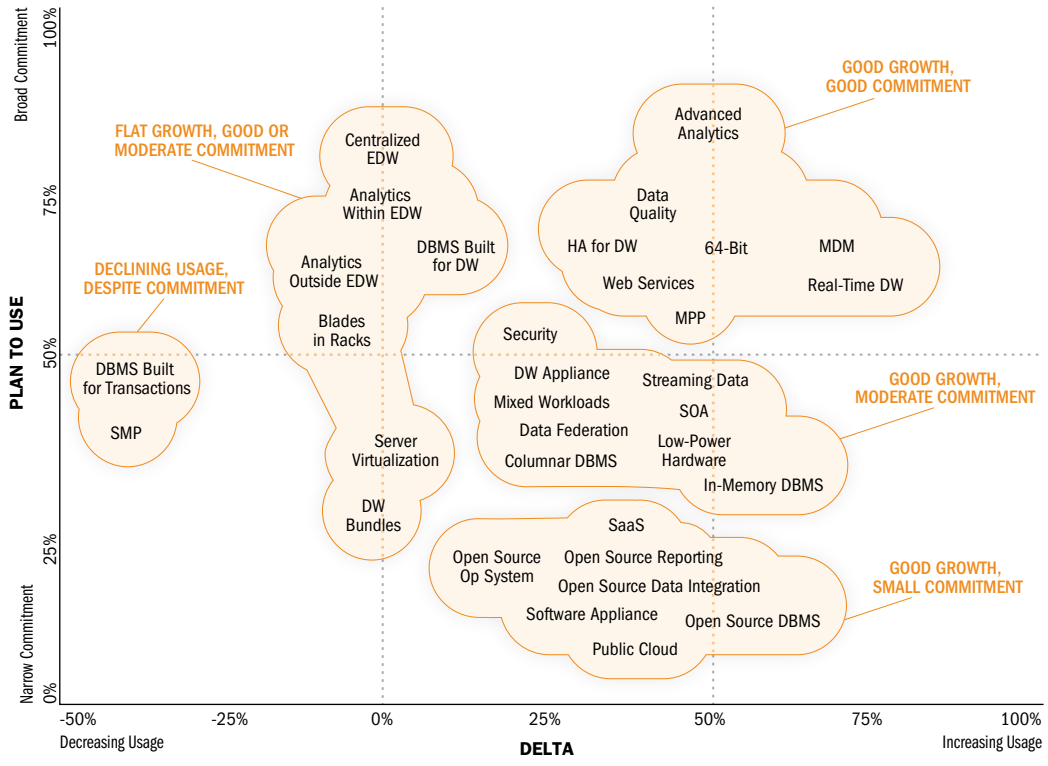



***Figure 9.*** *Plots are approximate, based on values from Figure 8.*

## Trends for Next Generation Data Warehouse Platform Options

**Rates of growth and commitment identify five groups of next generation options.**

Figures 8 and 9 reveal a number of trends concerning how users plan to apply various options to their next generation data warehouse platforms. In particular, five groups of options stand out based on combinations of growth and commitment. (See the groupings in Figure 9.)

- **Good growth, good commitment.** As you can see in Figures 8 and 9, most DW platform options will experience growth in the near future. However, the options most likely to live up to our great expectations and sustain growth over the long haul are those that have good survey results for both delta (growth) and plan to use (commitment). These include some of the most hotly pursued features and techniques of recent years, namely real-time data warehousing, MDM, 64-bit processing, advanced analytics, MPP, data quality, Web services, and high availability for the DW. In a lot of ways, this category is the epitome of the next generation data warehouse platform, because of its mix of leading-edge options supported by real-world organizational commitment.
- **Good growth, moderate commitment.** This is a mixed bag of features and techniques that share one thing in common: all are relatively new and have recently "arrived" by achieving a moderate commitment from the data warehouse community. Considered too new to touch a few years ago, this category includes technologies such as data federation, data warehouse appliances, mixed workloads, and columnar databases. Some of these options are poised for significant growth, namely streaming data, SOA, and in-memory databases.
- **Good growth, small commitment.** It's interesting that this category includes some of the newest options for data warehousing, including SaaS, public and private clouds, and four types of

open source software. These options are so new to data warehousing that they have only a minimal commitment, as yet, from the broader data warehouse community. Yet, among the few organizations making the commitment, these new option types will see very good growth.

- **Flat growth, good or moderate commitment.** The options in this category are all common data warehousing features and techniques, including essential options such as centralized enterprise data warehouses (EDWs), analytics within the EDW, and DBMSs built for data warehousing. Despite the good to moderate commitment users have made in this category, growth is flat (or close to flat). For half of these options (the top half, as seen in Figure 9), flat growth is probably due to saturation brought on by the popularity of these options. In the bottom half, growth may be flat due to the newness of these options to data warehousing (e.g., server blades in racks and server virtualization).
- **Declining usage, despite commitment.** This category includes two of the great pillars of data warehousing: relational DBMSs designed for transaction processing (though used for data warehousing) and server processing architectures based on symmetrical multi-processing (SMP). In fact, these are by far the most common components found in data warehouse platforms deployed today. If these are so popular, then why does the survey show them in decline? As we'll see later in this report, many users have come to the conclusion that their organizations would be better served by using DBMSs designed for data warehousing and server architectures based on massively parallel processing (MPP). For this reason, survey respondents anticipate a decline in SMP and DBMSs designed for transaction processing, as they migrate to MPP and DBMSs purpose-built for data warehousing.

# Next Generation Data Warehouse Platform Options

This final section of this report segments the many options available for next generation data warehouse platforms into 10 functional groups that are based on related features and techniques. Within each group, one or more primary options were identified by TDWI survey responses as high-priority items. The rest of the group consists of features and techniques that are similar to the primary one or that achieve a similar goal. The goal is to boil down the dozens of available options to a short list of 10 groups.

**Functional groups of DW options define the next generation.**

Note that the 10 groups define the essential (or most pressing) requirements of a next generation data warehouse platform. Since the order of groups listed here is based on projected growth (as defined by survey responses), their sequence in this discussion suggests a priority order. It is TDWI's sincere hope that these requirements and priorities will assist users as they plan their next generation data warehouse platforms.[1]

## Real-Time Data Warehousing

Looking back at Figure 8, we see that real-time data warehousing (RTDW) topped the chart. This is because it has the greatest projected growth rate of the DW options surveyed. It also has a strong commitment from user organizations (58% plan to use it). A mere 17% of respondents committed to RTDW reported using some form of real-time functionality with their data warehouse platform today. Yet, a whopping 92% claim they'll be using RTDW within three years. Extrapolating from the survey data, in three years, roughly half of DWs will be retrofitted or replaced to enable RTDW. Why have organizations delayed the adoption of RTDW techniques?

[1] Most of the DW options discussed here are supported by products from software vendors, and so representative vendors and products are listed. The vendors and products mentioned here are representative, and the list is not intended to be comprehensive.

**Making a DW operate in real time is not trivial.**

For one thing, retrofitting RTDW to a traditional, history-oriented DW is complex and expensive, as is upgrading or replacing a database management system to get a platform more conducive to RTDW. For another thing, business managers have been slow to understand the true value of RTDW, namely the intelligence, analytics, and broader view of business entities that operational processes and operational applications gain from tight integration with a data warehouse.

**Real time is more of a business practice than a technology.**

Real-time and similar functions are key enablers for RTDW, but it's about much more than technology. RTDW unifies operational and analytic processes, plus their supporting technologies. It transforms how the business runs (and how its IT systems interoperate), so that as many operations as possible are enlightened by the full informational view, historical context, and analytic power of the data warehouse and related business intelligence infrastructure.

Real-time data warehousing is a primary option that relates to other, similar options:

- **Trickle or streaming data loads** (48% delta). This is a base requirement for RTDW, and it usually involves capturing fresh data from source systems as it's created or altered. The data is then processed for reports and analysis that are refreshed frequently from the data warehouse. This option makes tough demands on data integration technology, as well as the DBMS that manages the data warehouse.
- **High availability for the DW** (32% delta). RTDW supports time-sensitive business practices such as operational BI, on demand performance management, and just-in-time inventory management. These practices are not "real time" and they aren't in continuous communication with operational systems (which are mostly non-stop) unless the data warehouse is up and running 24x7. Hence, a next generation data warehouse platform that gives priority to RTDW must also have a strategy for DW high availability.
- **Data federation** (22% delta). In contrast to capturing and loading data continuously (as with trickle or streaming approaches), data federation gets fresh data from source systems only when an application, report, or user asks for it. This reduces the overhead of continuous data capture, but usually lacks the full array of data processing options available through other approaches.

Vendors with database management systems (DBMSs) have added new features to their products and helped develop best practices for RTDW and similar DW options. For example, Teradata's "active data warehousing" and IBM's "dynamic data warehousing" both encompass practices and technologies for loading a data warehouse continuously from various source systems, as well as closing the loop by enabling the DBMS to push time-sensitive data out to a variety of applications, reports, users, and so on. HP Neoview is built on a transactional foundation—but fully optimized for data warehousing—to provide low latency data loading and high throughput of transactional queries concurrent with enterprise data warehouse workloads. RTDW has hefty data integration requirements, and Oracle enables RTDW for Oracle Database through integration products like Oracle Streams (for application integration), Oracle Data Integrator (for real-time ETL), and Oracle Data Services Integrator (for data federation).

**USER STORY A NEED FOR REAL-TIME DATA WAREHOUSING CAN FORCE A NEW DW GENERATION.**

"When we replaced our data warehouse platform around 18 months ago, we focused on finding one that could operate at near real time, right out of the box," said Deborah Schanda, a senior manager at outdoor clothier Timberland. "That's because the business needs the enterprise data warehouse to interoperate tightly with operational applications, to support near real time reporting and analytics. For example, one of the first things we did was give business people constantly updated data for all wholesale orders per region, so they can adjust inventories and shipments on the fly. Next on the horizon—other 'active data warehousing' techniques, like pushing warehouse data back to operational systems."

## Data Management Practices

Curiously, some of the most pressing needs for a next generation data warehouse platform involve technologies and practices that we generally don't think of as part of the platform. In particular, many users need to update the data management tools that process data for use through the DW.

**Technologies outside the DW platform can be generational, as well.**

For example, the survey for this report predicts strong growth for master data management (MDM) (68% delta), and it reveals a good commitment from users (62% plan to use it with the DW platform). This makes MDM one of the highest priorities for data warehousing. A similar (but slightly less urgent) data management practice is data quality (DQ), which likewise should see strong growth (40% delta), bolstered by committed users (69% plan to use it).

There are good reasons why MDM and DQ are high priorities for next generation DWs:

- **Most DWs still lack MDM and DQ functions.** MDM and DQ certainly aren't new, but they're still underutilized by data warehousing professionals. For many, extending data integration and other data management practices around the DW to include MDM and DQ is an act of catching up, not stepping out on the leading edge. The goal is to achieve quality decisions based on quality data.
- **Tighter integration with operational systems demands MDM.** In these situations, MDM ensures better integration by identifying data properly, so that the best sources are found for a specific use and "apples to apples" data exchanges are made. Likewise, some organizations appoint the EDW as the "system of record" for master and reference data, which obviously requires a hefty MDM solution for the EDW.
- **Regulatory and financial reports must be squeaky clean and accurate.** More and more, organizations produce these reports through their BI and DW infrastructure. These reports must be as accurate and credible as possible. Data management technologies like DQ and MDM help achieve that end, though most DWs lack these today, forcing a generational change. Given the current economic and political environment, regulatory requirements are set to increase rapidly in the next few years.[2]

**USER STORY SOMETIMES IMPROVING YOUR DATA WAREHOUSE PLATFORM ISN'T ABOUT THE PLATFORM PER SE.**

"One of the barriers to our next generation data warehouse was our old hand-coded jobs for extract, transform, and load (ETL)," said Tony Simpkins, the data warehouse supervisor at Grange Insurance. "The business needed us to integrate more data into the warehouse so they'd have a longer history to analyze for actuarial purposes. But we couldn't do that with our legacy ETL solution, which took up our entire five-hour batch window. Now that we've migrated the solution to a vendor's ETL tool, it runs in only 22 minutes. That frees up plenty of time to integrate more data, so we can proceed with other plans for our next generation."

[2] For a survey of data quality vendors and tools, see the TDWI Technology Market Report *Enterprise Data Quality Tools* (Q2 2006). For a survey of MDM vendors and tools, see the TDWI Technology Market Report *Segmenting Master Data Management Solutions* (Q4 2006). Both reports are available online to TDWI Members at www.tdwi.org/research.

## Cloud Computing and Software-as-a-Service (SaaS)

Cloud computing is one of the newest and most innovative platform choices to come along in years. This is where numerous hardware servers and other resources are pooled and virtualized, so that they can be freely allocated to applications and software platforms as resources are needed. This enables software applications to dynamically scale as workloads increase. As the workload of an application decreases, resources are freed up for use by other systems. Furthermore, dynamic allocation and re-allocation give the provider of the cloud fuller utilization of server resources, with less administrative work, as compared to traditional data center approaches.

**Early adopters set up analytics in a cloud, sometimes an EDW.**

Although cloud computing and similar virtualization techniques are firmly established with operational applications today, they're just now starting to be used as DW platforms of choice. The dynamic allocation of a cloud is useful when the data volume of the warehouse varies unpredictably, making capacity planning difficult. According to users TDWI interviewed for this report, a data warehouse (or analytic database) in a cloud is often set up for analytics, to accommodate the large and unpredictable volumes of data that business analysts and other power users collect, analyze, and then archive. But TDWI has also encountered users who have moved their entire enterprise data warehouse to a cloud.

There are different kinds of clouds and related services:

- **Public cloud.** The cloud is available off-site through a third-party provider. As mentioned earlier in this report, public clouds are set for substantial growth (37% delta), but within a rather short list of committed organizations (14% plan to use a public cloud).
- **Private cloud.** The cloud is set up in-house (usually as part of IT infrastructure) and made available to departments within an enterprise. Private clouds are set for even more growth than public ones (61% delta), but still within a minority of organizations (22% plan to use a private cloud). The preference for a private cloud over a public one for DW applications (as seen in survey responses) probably reflects user concerns over security, the volatility of new public cloud providers, and the difficulty of moving data into and out of a cloud.

**Cloud and SaaS aren't the same thing, but they usually go together.**

- **Software-as-a-service (SaaS).** This may refer to Internet-based software applications that happen to run in a public cloud, or it may refer to a cloud as a platform service. Nowadays, SaaS and cloud are regularly mentioned in the same sentence. That's because, more and more, SaaS offerings are set up to run in a cloud. But keep in mind that SaaS is also available via non-cloud infrastructure. For BI users, finding SaaS-based analytic applications that meet specific BI requirements can be challenging. But SaaS also has appealing benefits such as automatic software updates for SaaS applications and easy outsourcing of IT and BI infrastructure.

A number of vendors are addressing diverse needs for clouds and data warehouses. For example, as part of their data center product and service families, both IBM and HP offer public and private clouds that are suited to data warehouse platforms, as well as other applications. Some DBMSs have a reputation for running well in a cloud, like the Aster nCluster DBMS from Aster Data Systems. And Kognitio offers a dynamic platform for data warehousing and advanced analytics via a cloud that's priced in a SaaS business model. Oracle Database has recently introduced integration with public clouds for backup and recovery; this provides access to relatively cheap storage and easy growth for data warehousing backup processes.

**USER STORY** **CLOUDS PROVIDE A HEFTY INFRASTRUCTURE QUICKLY, CHEAPLY, AND FLEXIBLY.**

Lenin Galli is the director of business intelligence at ShareThis: "ShareThis is a sharing network that makes it simple to share any online content quickly. ShareThis allows users to save and share publisher content, including text and pictures to many services including Facebook, Twitter, instant messages, e-mail, SMS, Digg, and others. The compilation of all sharing services into a single, collapsible widget reduces clutter on the page, saves time, and increases page traffic.

"At ShareThis, we've made a deep commitment to cloud computing. That's because we're a start-up, and we need a low-cost investment. Also, we don't want to be an infrastructure company, although we have to handle very big data. That would be a distraction that would keep us from our engineering goals. Plus, resource planning is tough, given our unpredictable growth. Our commitment to cloud affects the tools we select, and we chose the Aster nCluster database from Aster Data [Systems], because it has a proven track record running in the public cloud we use. So, our multi-terabyte warehouse of aggregated data and our analytic applications run in the cloud. And we even cache our widget software in an edge cloud."

## In-Memory Processing and 64-Bit Computing

**Large memory space is the most compelling reason for 64-bit DWs.**

Sixty-four-bit systems typically offer faster CPUs and more power-efficiency hardware than older systems. But, for DW professionals, the most compelling benefit of 64-bit systems is the large space of addressable memory. For example, users whom TDWI Research has talked with recently report that a leading reason for upgrading to 64-bit systems is to deploy an in-memory database for reporting or analytic applications that need very fast query response. In-memory databases provide such speed because they don't have disk input/output (IO) to slow them down. The in-memory database is usually a function of a DBMS, but some BI platforms for reporting and analysis also support in-memory data stores and related processing.

Tools for extract, transform, and load (ETL) commonly support in-memory processing in a 64-bit environment, so that complex joins and transformations are executed in a large memory space without the need to land data to disk in temporary tables. This makes an ETL data flow a true "pipe," which means the ETL tool can scale up to large data volumes that are processed in relatively short time periods.

**Future generations: in-memory, multi-terabyte DWs.**

Disks aren't going away; they're needed for storage, and in-memory data structures are typically loaded from disk. One of the barriers is that main memory is still rather expensive. As the price comes down, however, data warehousing should experience an upswing for in-memory processing, and DW platforms of the future will commonly support multi-terabyte memory spaces. Furthermore, solid-state disks (which have many of the performance characteristics of memory) now have a foothold in the IT market, and they are also coming down in price. In many ways, MPP architectures are an alternative to 64-bit-based, in-memory processing, because MPP pools memory resources from many servers (whether 32- or 64-bit) to create a large virtual space.

In-memory databases and the 64-bit hardware and software that enable them are high priorities for next generation data warehouse platforms. According to this report's survey, in-memory databases should experience strong growth (58% delta), bolstered by a moderate commitment (39% plan to use it). Likewise, 64-bit hardware and software is set for growth (53% delta), driven by a strong commitment (58% plan to use it). Whenever BI/DW teams plan a next generation data warehouse platform, 64-bit systems and in-memory processing should be a required part of the design.

As an example of a relevant vendor product, the Sun Oracle Database Machine is a 64-bit platform that leverages the extended in-memory capabilities of Oracle Database 11g R2.

## Open Source Software

**SURVEY SAYS: Over one-third of users surveyed use open source in BI, DI, or DW.**

According to survey data from TDWI Research, open source tools are being used at an unprecedented level in business intelligence (BI), data integration (DI), and data warehousing (DW). In a TDWI Technology Survey of May 2009, over one-third of organizations surveyed reported using open source software in BI, DI, or DW applications. There are good reasons for the upswing of open source software used in data warehousing:

- The recession has driven up interest in low-cost open source software.
- Open source tools are coming into a new level of maturity.
- Open source software augments traditional enterprise software without replacing it.

**SURVEY SAYS: Open source has benefits and barriers.**

According to the TDWI Technology Survey of May 2009, the three leading barriers to using open source software are: Getting adequate support and maintenance (82%), selling it to peers and management (51%), and learning a new tool (24%). The three leading benefits of open source are: Its low price (56%), the ability to download and try the software before buying (50%), and its community of collaborative developers (48%).

The survey for this report shows that three types of open source software are poised for good growth in terms of usage with data warehouse platforms. (See the Delta column in Figure 10.) Yet, the growth will occur within a minority of committed organizations. (See the Commitment column in Figure 10.) This is typical of a new-but-promising technology; it's hard to get users to try it because it's new, but they tend to grow its usage once they try it because it's cost-effective.

| OPEN SOURCE SOFTWARE TYPE | DELTA (GROWTH) | PLAN TO USE (COMMITMENT) |
|---|---|---|
| DBMS | 48% | 17% |
| Data integration tool | 41% | 27% |
| Reporting or analysis tool | 41% | 26% |

***Figure 10.*** *Growth and commitment for types of open source software.*

Many vendors are providing open source software suitable for use in BI, DI, and DW. Open source DBMSs are available from Infobright, MySQL, and PostGres. Open source data integration tools are available from Apatar, JitterBit, Pentaho, and Talend. Open source reporting or analysis tools are available from Actuate and Jaspersoft.

**USER STORY COLUMN ORIENTATION SPEEDS UP SCANS OF WIDE RECORDS.**

Kevin Galligan is an independent consultant who specializes in Web applications. "I recently got a gig where I needed to make a database query-able in real time over the Internet. That's simple enough, except that each row hit by the queries had over 900 fields. The wide record choked every brand of database management system that I tried. Finally, someone suggested that I try a column-oriented database. Unfortunately, I got this suggestion late in the day, and I needed to have something up and running the next business day.

"After a few Internet searches, I found Infobright's open source database with a column-oriented data store. I downloaded it at 10:00 p.m. and had it running with my application by 3:00 a.m. Later that morning, I successfully demo'd my new query capability. The experience taught me a couple of valuable lessons. The wider the record, the more a column-oriented data store makes sense. And the 'try before you buy' business model of open source software lets you prove that it works in your application before you make an investment. Plus, you can download open source software in the middle of the night!"

## Advanced Analytics

In the next few years, advanced analytics will experience good growth (47% delta), bolstered by the strongest commitment seen in Figure 8 (a whopping 85% plan to use advanced analytics). The increase in usage of advanced analytics is driven by organizations' need to understand constantly changing business environments, as well as to discover opportunities for cost reductions and new sales targets.

There are different analytic methods users can choose as they move beyond basic OLAP-based methods and into advanced analytics. Some users choose advanced analytic methods based on data mining, predictive analytics, statistics, artificial intelligence, and so on. The majority of users, however, seem to be choosing SQL-based methods. This is probably because they know and trust SQL, and they can leverage the SQL-based tools and skills they already have.

**"Load and go" is an advanced analytic method based on SQL.**

This trend is pushing SQL-based analytics to a new extreme. With "load and go" methods, users quickly load a few terabytes of raw operational data and go at it with ad hoc queries until the data reveals the answers they need. The ad hoc queries get more complex with each iteration by a business analyst or similar power user. This method doesn't allow time and resources for data transformation, cleansing, or remodeling, so users compensate with lots of WHERE clauses, table joins, and temporary tables (when necessary). SQL-based analysis at this advanced level is powerful, but it succeeds only when supported by a DBMS that can quickly execute extremely complex SQL statements run against multi-terabyte volumes of raw data, in a schema-neutral fashion that supports "load and go" practices.

**Most EDWs are optimized for reporting, not advanced analytics.**

Many organizations depend on an enterprise data warehouse (EDW) to fulfill most of their analytic requirements. The problem is that most EDWs are optimized for standard reports and recurring analytic questions based on online analytic processing (OLAP). It may be that EDWs have lost their analytic prowess as users have evolved them into reporting and OLAP databases. Hence, there's a need for DBMSs that can support free-form, ad hoc analysis against multi-terabyte data stores of mostly source data in simple data models—but still handle simpler workloads, like standard reports and OLAP.

**Where to store analytic data is a key design decision.**

Whether to store analytic data in the EDW or in a separate analytic database is one of the most critical design and architecture decisions adopters of next generation DW platforms must make.

- **Analytics processed within the EDW.** This is a very well established practice (in Figure 8, 74% plan to use it), but survey results indicate flat growth (4% delta). Many of the analytic tools based on data mining technologies require users to dump analytic data into flat files with a specific record structure, because that's the data structure the tool is optimized for. In recent years, data mining and predictive analytic tools have gotten better at processing data while it's stored in a DBMS. This is called "in-database analytics." With any luck, the trend toward in-database analytics will continue, because most users would rather manage data with an EDW or similar database and leave the data in place when analyzing it. Naturally, SQL-based analytics demand that data be managed by a SQL-compliant DBMS.
- **Analytic databases outside the EDW.** This, too, is a well established practice (61% plan to use it), but with indications of decline (-6% delta). This takes many forms. At one extreme, data marts proliferate outside the EDW until IT and DW teams are forced to rein them in through time-consuming data mart consolidation projects. This may well be what survey respondents were thinking of when they said there will be a decline in analytic databases outside the EDW. At the other extreme, a new best practice is to isolate disruptive analytic workloads on data warehouse appliances and other analytic databases outside the EDW.

Support for advanced analytics has been an area of great activity for software vendors in recent years. For example, many new software firms offering DBMSs built specifically for data warehousing and analytics have sprung up this decade, including 1010data, Aster Data Systems, Greenplum, Illuminate, Infobright, Kognitio, ParAccel, and Vertica. Neoview is a new EDW and analytics platform from established vendor HP. Many established DBMSs have updated their support for query optimization and SQL standards (partially for better SQL-based analytics), including from IBM, Oracle, Microsoft, Sybase, and Teradata. Likewise, some of the established DBMS vendors have improved their in-database analytic capabilities, as seen in the partnership between SAS and Teradata.

Other vendors have taken analytic processing to where data lives in storage. For example, the Oracle Database 11g R2—when used in combination with Sun Oracle Exadata—pushes the scoring of data mining models down into the storage tier (Exadata). This produces data mining results faster because far less data is moved over the network to central CPUs. The Dataupia Satori Server and the Netezza Performance Server provide similar storage-level processing for queries and advanced analytics. Note that this approach requires specialized hardware for the storage tier, which Dataupia, Netezza, and Oracle Exadata all provide.

> **USER STORY** **ANALYTICS AND BIG DATA ARE CRITICAL TO SUCCESSFUL BUSINESS ADJUSTMENTS.**
>
> "About 18 months ago, we saw a sudden change in customer behavior," said an intelligence director at a global consumer electronics firm. "To understand the behavior, we needed to start collecting terabytes of online data. We knew our current data warehouse platform couldn't scale up, so we went in search of a next generation platform.
>
> "We looked at data warehouse platforms that can cope with multiple terabytes, are proven in the marketplace, and come from an established vendor. These requirements led us to HP Neoview. For the configuration we needed, HP's pricing was similar to that of appliances, so we conducted a proof of concept at our site with a 12 TB box. It took two days to set up Neoview, fully configured and ready for loading data. It only took us six weeks to load data, model it, implement two BI platforms, and develop reports.
>
> "So, within a few short weeks, we showed our management KPIs and metrics they hadn't seen before, from queries running in seconds across millions of rows. In the proof of concept, we included streaming data via Web services from Web servers to Neoview, so new events register in reports refreshed by managers on demand. These features really sold management on our new generation of the data warehouse. And the new platform has been instrumental in my company's adjustment to new customer behaviors."

## Services

**SOA and Web services are hot DW options.**

Service-oriented architecture (SOA) and Web services are two of the most anticipated data warehouse options, according to this report's survey. Fifty-nine percent of respondents said they plan to use Web services in their next data warehouse, and Web services growth is projecting a delta of 40%. SOA isn't too far behind, with both a good commitment (47% plan to use it) and good growth (47% delta).

Users interviewed for this report explained why they're adapting services to data warehousing. Services are becoming the preferred type of interface because they're easily callable and consumable by many platform types, as compared to the proprietary interfaces supported by data integration and application integration platforms. The reuse of services has turned out to be as good as promised. In addition, best practices are now known for handling data through services.

Corporate cultures craving hand coding can hand-code Web services, if they insist. Even so, many different types of enterprise software tools can automate and manage services, including DBMSs, data integration tools, and BI platforms.

**Services can be hand-coded or tool based.**

"Elsewhere in IT—outside the data warehouse team, that is—my company has made a serious investment in SOA tools and infrastructure," said a user interviewed for this report. "As a second-tier priority for our next generation, we'll tie the new data warehouse into this infrastructure for broader data access, more real-time interfaces, and tighter integration with operational systems."

## Processing Architectures

**SMP is declining as MPP is rising.**

Symmetrical multi-processing (SMP) will suffer declining usage (-44% delta), despite a moderate commitment (36% plan to use it) and three decades of successful use in data warehousing. How could a technology so entrenched in data warehousing decline so quickly? Well, in a complementary trend, massively parallel processing (MPP) is poised for good growth (46% delta) fueled by good commitment (58% plan to use it). In a sense, these are competing and complementary technologies. Many users have plans to migrate from SMP to MPP when they next change their data warehouse platform.

To further quantify users' desires for MPP, the TDWI survey for this report asked what their processing architecture is today versus what they'd prefer. (See Figure 11.) Today, roughly two-thirds have SMP, and one-third has MPP. In terms of what respondents would prefer, the numbers shift to one-third SMP and two-thirds MPP.

**Today, which processing architecture is your data warehouse on? Ideally, which processing architecture would you prefer for your data warehouse platform?**

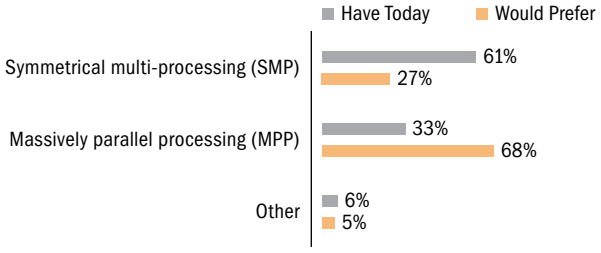



***Figure 11.*** *Based on 417 responses.*

**MPP yields fast queries and data scalability, and that's what DWs need.**

Why are so many DW professionals experiencing MPP envy? MPP involves multiple "shared nothing" nodes that work in parallel on the same computational problem, but without sharing memory or other resources. This is good for problems that can be parallelized effectively, but bad in its complexity of configuration and networking. In a nutshell, MPP's shared-nothing parallel processing is good for data warehousing because it yields fast query responses and scalability for data volumes. It also supports the trend toward scaling out, instead of scaling up. And some configurations of MPP work well with inexpensive commodity hardware.

Vendor DBMS products have supported MPP for many years, as seen in Kognitio's WX2 and most of Teradata's products. Many of the newer DBMSs on the market were built for MPP from the beginning, including Aster Data Systems, HP Neoview, ParAccel, and Vertica. Oracle Exadata is a hybrid, delivering an MPP storage grid underneath a shared-everything architecture. Microsoft's acquisition of DATAllegro in 2008 underscores the importance of MPP; Microsoft is moving the

MPP capabilities of DATAllegro into SQL Server (through a project code-named "Madison") to deliver massive scalability, support for multi-node hardware configurations, and an appliance-like user experience.

## Data Warehouse Appliances and Similar Platforms

The most widely discussed and argued new DW option of the decade has to be the data warehouse appliance (DWA). It has undergone considerable evolution since first appearing early this decade. DWAs now include diverse product types that any definition of data warehouse appliance should encompass.[3]

### Data warehouse appliance

**Use of proprietary hardware was part of the original DWA definition.**

Netezza was the first vendor to offer a data warehouse appliance (introduced around 2002), so early DWA definitions were based upon Netezza products, which provide a whole-technology stack for data warehousing. That is, the Netezza Performance Server combines database and operating system software with server and storage hardware in a complete data warehouse platform. Before Netezza, Teradata, Sequent, and WhiteCross (now Kognitio) had for years offered similar single-vendor combinations of hardware and software purpose-built for data warehousing, though not necessarily in an appliance package or described as an appliance.

DATAllegro launched in mid-2005 with a whole-technology stack solution involving proprietary hardware. DATAllegro soon left its proprietary hardware in favor of commodity hardware from other vendors, before being acquired by Microsoft in 2008. Teradata announced in 2008 a new family of data warehouse packages that includes DWAs, some running on commodity hardware. In 2003, Kognitio moved from proprietary to commodity hardware, after offering a data warehouse platform on proprietary hardware since 1989. The movement from proprietary to commodity hardware has good reasons behind it. Commodity hardware is relatively inexpensive and thus helps keep down the price of DWAs. This is important, since DWAs compete largely on their low cost.

### Software appliance

**DWAs experienced a trend toward commodity hardware.**

Starting in 2006, a new wave of vendors emerged with database management systems (DBMSs) purpose-built for data warehousing. These include DBMSs based on the relational model (Aster Data Systems, Greenplum, and Kognitio) and the columnar model (Infobright, ParAccel, and Vertica). Most of these DBMSs are sold and licensed stand-alone (like any DBMS software) or embedded in an appliance (usually with a certified or recommended hardware configuration). In the context of the embedded license model, most of the new DBMS vendors call their product a *software appliance*. This somewhat oxymoronic term refers to a software component (namely a DBMS) that may be embedded in a full data warehouse appliance. Hence, each of these vendors offers a partial-stack appliance, called a software appliance.

The software appliance has proved to be a good starting point for the new DBMS vendors. It allows them to focus on database software (not designing and building hardware), which is their point of greatest innovation and therefore their value proposition. The software appliance product enables the new, small DBMS vendors to partner with commodity hardware vendors and to benefit from these larger firms' resources.

### Bundled DW platform from multiple vendors

**DW bundles provide DWA benefits, so the two are similar.**

As DWAs entered the marketplace, relational database vendors (IBM, Microsoft, Oracle, Sybase, and HP) stepped up their offerings of hardware and software bundles that assemble a whole-technology stack for data warehousing. Most of these bundles are not DWAs per se, yet they offer many of the benefits of a DWA. In particular, a preconfigured technology stack reduces system integration work,



[3] See the article "Defining the Data Warehouse Appliance" at www.tdwi.org/publications/display.aspx?id=7784. An update of this article appears in Volume 27 of TDWI's *What Works* publication (www.tdwi.org/publications/whatworks). For further analysis, replay the archived 2009 TDWI Webinar "Data Warehouse Appliances: An Update on the State of the Art."

reduces time to use, and comes from a single vendor that supports the whole stack. Furthermore, vendor size matters in that some user organizations avoid start-up vendors. For these users, the hardware/software bundles are significant, because they come from large, stable vendors and include familiar, mature DBMSs.

Examples of bundles from leading database vendors include HP Neoview, IBM Smart Analytic System, and IBM InfoSphere Balanced Warehouse. Most of the hardware and software components in the IBM bundles come from IBM. In the case of HP Neoview, all the hardware and software components come from HP. Launched in late 2008, the HP Oracle Database Machine and the HP Oracle Exadata Storage Server are both based on Intel processors, hardware from HP, and software from Oracle.[4] Sybase's Analytic Appliance (announced in May 2008) combines pSeries hardware from IBM with Sybase IQ (a columnar database, purpose-built for data warehousing). In February 2009, Microsoft launched SQL Server Fast Track Data Warehouse, which accelerates DW deployments through reference configurations. Fast Track Data Warehouse is available on inexpensive hardware from Bull, Dell, EMC, and HP.

### The Data Warehouse Appliance Redefined

**DWAs of various types are now established DW platforms.**

Due to the trend among vendor offerings moving from whole-technology stacks to partial ones, a newly revised definition must encompass DWAs that complies with the original definition (from Netezza and DATAllegro), as well as the newer software appliances (from Aster Data Systems, Infobright, Greenplum, Kognitio, ParAccel, Vertica, and so on). Furthermore, hardware/software bundles assembled for data warehousing (from HP, IBM, Microsoft, Oracle, Sybase, Teradata, and so on) share many characteristics and benefits with DWAs, so these should be mentioned whenever DWAs come up.

Prospects of growth and adoption vary across the three definitions of appliances:

- **Data warehouse appliance.** Prospects for data warehouse appliances are very positive, based on the survey's indications of good growth (25% delta, as seen in Figure 8), driven by a moderate commitment (47% plan to use a DWA). This balance of growth and commitment shows that the DWA has definitely "arrived" as a common DW platform.
- **Software appliance.** This product type is even newer than DWAs, so it's not surprising that interest in it comes from a minority of committed organizations (only 19% plan to use it). Within this niche community, however, software appliances will experience good growth (32% delta).
- **Bundled DW platform.** The good news is that the survey predicts a moderate commitment for DW bundles (34% plan to use). The bad news is that growth will be near flat (4% delta). For a more bullish view of bundled DW platforms, see the discussion of Figure 4.

**Users manage large data marts and analytic databases with DWAs, sometimes EDWs.**

The user community continues to redefine how it uses data warehouse appliances. As discovery-driven advanced analytics based on SQL become more common and more mission critical, users are in dire need of a data warehouse platform that can respond quickly (with little or no tuning) to ad hoc and/or complex queries against multi-terabyte data sets of less-than-ideal structure and quality. DWAs fulfill this need, as do the analytic DBMSs discussed elsewhere in this report. For this reason, the vast majority of DWAs and software appliances manage multi-terabyte data marts and other analytic databases. EDWs are rare on DWAs and software appliances today (though common on DW bundles), but this should change over time as user confidence grows and appliance capabilities increase.

[4] Late in 2009, these Oracle products were rereleased on a new hardware platform under the names Sun Oracle Database Machine and Sun Oracle Exadata Storage Server, respectively.

**USER STORY DATA AGGREGATION AND SAAS PLACE TOUGH DEMANDS ON DW PLATFORMS.**

"Our data warehouse is critical infrastructure for the aggregated data and SaaS-based analytic applications we sell to retailers and their suppliers," said Groupe Aeroplan's IT Director Fiachra Woodman. "So, the data warehouse has to satisfy stringent requirements. That's why we acquired WX2, a data warehouse appliance from Kognitio. For example, we process 30 million records daily, cleansed and made ready for buyers overnight. We do ELT—not ETL—because the appliance is far better equipped than the average data integration server. Our customers expect the data warehouse to be available 24x7, so—in addition to our in-house data warehouse appliance—we maintain a shadow database at Kognitio's remote SaaS site. Using in-memory processing, the data warehouse platform we chose provides fast responses, although our customers are hitting us with very complex queries and joins against very large data sets. All this gives our customers many terabytes of fresh, cleansed, and enhanced data that's available continuously through high-performing analytic applications."

## New Database Management Systems as Alternative Options

To review for a moment, a data warehouse platform consists of five basic components (as illustrated in Figure 1). But, among these, which components are the top priorities for data warehouse professionals? To quantify platform priorities, TDWI asked, "Of the components of your data warehouse platform, which do you feel is the most critical to success?" (See Figure 12.)

**Of the components of your data warehouse platform, which do you feel is the most critical to success?**

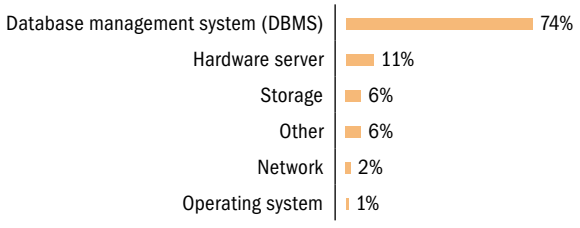



***Figure 12.*** *Based on 417 responses.*

The DBMS is the heart of the next generation DW platform.

**The DBMS is the most critical component of a data warehouse platform.** At least, that's how the wide majority of survey respondents feel (74% in Figure 12). That's a natural response, given that the DBMS is the component of the platform where the most work must be done to implement a data model and optimize it for query performance. It's where the battles of scalability are fought. And the focus on the DBMS is a sign of both deep DBMS expertise and deep brand dependence. After all, data warehouse professionals tend to prefer DBMS brands they already know, which affects data warehouse platform decisions.

**The DBMS is where many next generation innovations have occurred this decade.** DBMS brands from all vendors have made great strides in recent years, but these accomplishments are too numerous to list here. Instead, let's focus on new types of DBMSs and new vendors that have arrived recently. These new arrivals are significant and worth discussing, because they are revitalizing the DBMS market with creative applications of open source, analytic capabilities, appliance packaging, columnar storage engines, MPP, competitive pricing, cloud and SaaS deployment options, and so on.

Users need more DBMS choices to evaluate.

**The DBMS is an area where many data warehouse professionals want more alternatives.** The vendor marketplace early this decade had consolidated down to three leading vendors (IBM, Microsoft, and Oracle), each providing a relational DBMS originally designed for online transaction

processing (OLTP), but successfully retrofitted with data warehousing capabilities. Undeniably, these are powerful DBMSs for data warehousing, as proved by the fact that the vast majority of data warehouses are managed by these DBMSs today. But few viable alternatives existed that were purpose-built for data warehousing (just Kognitio WX2, Sybase IQ, and Teradata). As we'll see from survey data, DW professionals are actively looking for such alternatives. In many ways, the next generation data warehouse platform is about exploring new options, and new DBMSs give users more options where they truly need and want them.

### Open Source DBMSs

**An open source DBMS is a low-cost alternative.**

The economy has driven up usage, not just interest, in low-cost open source DBMSs. TDWI regularly finds them supporting departmental data marts and operational data stores (ODSs). Due to the low price, open source DBMSs are becoming common in poorly funded projects such as skunkworks BI solutions. Open source DBMSs have a strong reputation for perky query response, so they are getting used as the DBMS for analytic applications or Web-based applications that depend on query performance (like product catalogs in e-commerce). Most open source DBMSs support Web technologies (like Java, PERL, R, and so on), which explains why they're gaining traction on the Web. And—because they can be stripped down to a minimal footprint—open source DBMSs are often used as embedded databases within various applications.

Representative products include Infobright Community Edition (ICE), PostGres, and MySQL (recently acquired by Sun, which is being acquired by Oracle). Between MySQL and PostGres, PostGres is used most often in data warehousing because of its good query performance. Infobright offers a column-oriented data store and other enhancements for MySQL, which make it better suited to data warehousing.

### Analytic DBMSs

**An analytic DBMS is an alternative platform that complements the EDW.**

In the discussion of advanced analytics, we saw that the adoption of analytics forces users to make a critical design and architecture decision: Do we store analytic data in the EDW or in a separate analytic database? This is an important question for next generation data warehouses, since the raison d'être of many of these is to accommodate advanced analytics. TDWI consistently sees its Members and others—when faced with this decision—choosing to employ a secondary, alternative platform outside the EDW. This is to isolate demanding and unpredictable analytic workloads from the EDW, where they might degrade EDW performance. Likewise, users may isolate real-time workloads on a complementary platform next to the EDW.

Given this new demand for alternative DBMSs optimized for analytics, a number of vendors have stepped in to supply the demand. Under the rubric "analytic database," new vendors include 1010data, Aster Data Systems, Illuminate, and Infobright. Sometimes, the analytic database is available in an appliance package, as with Aster Data, Greenplum, Kognitio, Netezza, ParAccel, and Vertica. Established vendors are also supplying the demand for analytic DBMSs, as seen with HP Neoview, the Sun Oracle Database Machine, IBM Smart Analytics Optimizer, and Teradata's new appliance packages.

### Data Warehouse Appliances

**By nature, a DWA is an alternative platform.**

A pre-integrated data warehouse appliance is an alternative platform choice, compared to the in-house assembled data warehouse platform that's most common today (as described in the discussion around Figure 4). Besides being an alternative itself, the DWA has wielded considerable influence by making data warehouse professionals aware of other alternatives:

- **DWAs have revitalized MPP.** Data warehousing too often repurposes transactional technologies and practices, and SMP is a prominent one. Most appliances support MPP and therefore remind us that MPP has advantages over SMP for data warehousing.
- **DWAs remind DW professionals that alternative databases exist.** In particular, the software appliance has raised awareness of DBMSs from Aster Data Systems, Greenplum, Kognitio, ParAccel, and Vertica.
- **DWAs have introduced DW professionals to open source databases.** Note that appliances from Netezza and ParAccel began with code borrowed from the open source community.

### Columnar DBMSs

**A columnar DBMS provides an alternative data storage method.**

Many people are confused about what a columnar DBMS is and does, so let's clear up the confusion with some quick definitions of relevant terms:

- **Relational database (RDBMS).** A database management system (DBMS) where data is modeled/structured mostly as tables and keys.
- **Row-oriented data store (or storage engine).** In most relational DBMSs, data is physically stored as table rows. This makes sense in OLTP environments, where each row is a transaction that's committed to or retrieved from the database. But it makes less sense with analytic queries that parse columns.
- **Columnar (or column-oriented) data store (or storage engine).** Data is physically stored as table columns, even if the data is structured in a relational model. This makes it easier for the DBMS to create statistics and lists about the content of columns, which in turn greatly speeds up column-based queries. (As proof, see the User Story by consultant Kevin Galligan on page 22.)
- **Columnar (or column-oriented) database.** This is a relational DBMS with a columnar data store, optimized for column-oriented queries.
- **A columnar DBMS is also an analytic DBMS.** After all, one of the leading technology requirements of analytics is high performance for complex queries against large data sets, and that's what the column store enables. A columnar DBMS is also "purpose-built for data warehousing" (as the saying goes).

**Most users prefer row data stores, but they're aware of column stores.**

In the survey for this report, TDWI asked, "What kind of storage engine would you prefer for data warehousing?" (See Figure 13.) A slight majority (51%) prefers a row-oriented storage engine. The fact that 42% prefer a column-oriented storage engine shows that the data warehouse community is aware of this somewhat new option. Oddly enough, no one would prefer to use both types of storage engines.

**What kind of storage engine would you prefer for data warehousing?**

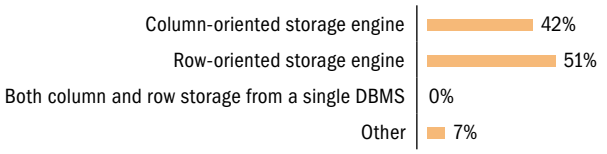



***Figure 13.*** *Based on 417 responses.*

A few new vendors have emerged in recent years with DBMSs that support columnar data stores. This includes Infobright, ParAccel, and Vertica. And let's not forget that Sybase IQ is a columnar DBMS that's been available for about 10 years now. Sybase IQ proved the concept of the columnar data store early on, which makes it the "mother" of all columnar DBMSs.

# Recommendations

- **Plan for the next generation data warehouse that *is* in your near future.** TDWI survey data shows that almost half of respondents are planning a data warehouse platform replacement in 2009–2012. Many others anticipate keeping their current platforms, but with significant updates.
- **Recognize that next generation technology drivers are really business drivers.** For example, real-time data warehousing (RTDW) is a high priority for the next generation. But you should never implement this hefty technology in a vacuum. RTDW should support time-sensitive business methods like operational BI and just-in-time manufacturing or inventory management.
- **Avoid assembling your own data warehouse platform.** Today, most DW platforms (55%) are assembled by in-house personnel. This is time-consuming, a distraction, not your major deliverable, and so on. Be open to pre-assembled and integrated DW appliances and similar DW bundles, or outsource the work to a system integrator or consultants.
- **Plan for big data.** The number of organizations in the "10-terabyte club" will double within three years. Many of the organizations moving into the club will probably need to replace their data warehouse platform to make the move possible and sustainable.
- **Be open to low-cost DW platform options.** When asked if the economy has made them more open to cheap platforms, a whopping 57% answered yes. If budget constraints are blocking your BI projects, look into open source software (especially DBMSs) and data warehouse appliances, plus creative approaches to licensing (SaaS) and low-investment deployments (clouds).
- **Don't forget options outside the DW platform.** Technologies outside the DW platform can be generational, as well, especially data management practices like master data management, data quality, and data integration.
- **Expect analytics to be a priority for your next generation DW platform.** Your peers in other firms report they'll step up analytics, so perhaps you should, too. Many are moving data for advanced analytics off the EDW onto next generation platforms like clouds, appliances, analytic databases, and columnar databases, so this might make sense for you, too.

- **Note that some next generation options are a critical path to others.** For example, you may need to upgrade to 64-bit hardware and software before you can achieve goals with in-memory databases, faster query performance, and data scalability. Likewise, you may need to implement Web services and SOA before you get the broad access and real-time functionality needed to tightly integrate the EDW with operational systems.
- **Realize that your next generation DW platform may require multiple platforms.** Trends toward advanced analytics and real-time business methods lead many organizations to deploy separate (sometimes alternative) platforms for these to off-load and complement an EDW.
- **Be open to alternative DBMSs.** Thanks to the next generation, you now have more alternatives to consider, including open source, analytic, appliance, and columnar DBMSs. Decide carefully which of these are appropriate to secondary platforms versus the primary EDW.


**Aster Data Systems**
**www.asterdata.com**

Aster Data Systems is a proven leader dedicated to providing the best data analytics and management platform for frontline data warehousing—the first DBMS to tightly integrate SQL with MapReduce—providing rich insights on data managed on clusters of inexpensive commodity hardware. The Aster *n*Cluster database cost-effectively powers rich analytic applications for companies such as Coremetrics, MySpace, aCerno (an Akamai company), and ShareThis. Running on low-cost off-the-shelf hardware, and providing "hands-free" administration, Aster enables enterprises to meet their data warehousing and analytics needs within their budget.


**HP**
**www.hp.com**

HP, the largest technology company in the world, brings together a portfolio of services, software, hardware, and partner technologies to create business intelligence solutions that allow organizations to make better, faster decisions that optimize business performance, capitalize on customer relationships, and maximize productivity. HP BI solutions connect the information scattered across an organization, reconciling data silos and allowing the knowledge gained to flow across the enterprise. As a trusted advisor, HP combines unbiased guidance with a deep understanding of client environments to help businesses realize the full value of their BI investments.


**IBM**
**www.ibm.com/software/data/infosphere**

IBM is a leader in data warehousing solutions, offering a full range of information management solutions that enable organizations to innovate with information and perform real-time analytics on data to optimize business processes, provide trusted information to end users, and reduce costs across the enterprise. Key offerings include the IBM InfoSphere family of products, such as InfoSphere Data Warehouse, and the new IBM Smart Analytics System.


**Infobright**
**www.infobright.com**
**www.infobright.org**

Infobright, the open source data warehousing company, delivers a self-managing, low-cost analytic database that enables companies with limited staff and budget to implement a highly scalable data warehouse. Infobright combines a column-oriented database with unique Knowledge Grid technology that reduces ongoing administration by up to 90% and cost by more than 50%. Infobright is being used by companies in online marketing, financial services, telecommunications, and other industries to provide rapid access to critical business data with unmatched operational simplicity.


**Kognitio**
**www.kognitio.com**

Kognitio is the provider of Kognitio WX2, a high-performance analytical in-memory database solution that allows organizations to understand more about their business and their customers in shorter timescales. Companies run Kognitio WX2 to analyze large volumes of data quickly, allowing them to make more informed, better business decisions that help them to drive growth and reduce costs. Kognitio WX2 is available as a software-only solution, as a fully configured data warehouse appliance running on industry-standard hardware, or on-demand via DaaS, Kognitio's data warehousing as a service offering.


**Microsoft**
**www.microsoft.com**

Founded in 1975, Microsoft (Nasdaq "MSFT") is the worldwide leader in software, services, and solutions that help people and businesses realize their full potential.


**Oracle**
**www.oracle.com**

Oracle's business is information—how to manage it, use it, share it, protect it. For nearly three decades, Oracle, the world's largest enterprise software company, has provided the software and services that let organizations get the most up-to-date and accurate information from their business systems. Today, Oracle is the world's largest business software company, with more than 320,000 customers.


**Sybase**
**www.sybase.com**

Sybase is an industry leader in delivering enterprise and mobile software to manage, analyze, and mobilize information. We are recognized globally as a performance leader, proven in the most data-intensive industries and across all major systems, networks, and devices. Our information management, analytics, and enterprise mobility solutions have powered the world's most mission-critical systems in financial services, telecommunications, manufacturing, and government. Sybase IQ is the world's leading column-based analytics server, handling the most challenging business reporting and analytics requirements with ease, and delivering dramatically faster, consistently more accurate analysis to address the growing need to anticipate risks and opportunity in a volatile global climate.


**Teradata Corporation**
**www.teradata.com**

Teradata is the acknowledged global leader in data warehouse innovation and analytical solution development. Every day we raise our customers' intelligence to higher levels, making them more focused and competitive by gathering enterprise information and extracting actionable insight. Teradata elevates enterprise intelligence by giving every decision maker the insight required for smarter, faster decisions. We add value and reveal opportunity across more dimensions than any competing solution. In every industry and geography, our technologies and expertise make the difference. Simply put, Teradata solutions make companies smarter and give them the competitive advantage to win.

## 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