TDWI report series

A 101communications Publication

August 2004

In Search of a Single Version of Truth: Strategies for Consolidating Analytic Silos

By Wayne Eckerson

tdwi

Research Sponsors

Ascential Software Corporation

Hummingbird Ltd.

Informatica Corporation

Teradata, a division of NCR

Unisys Corporation

TDWI report series

August 2004

In Search of a Single Version of Truth: Strategies for Consolidating Analytic Silos

Table of Contents

About the Author

Wayne Eckerson is the director of research for The Data Warehousing Institute (TDWI), the leading provider of high-quality, in-depth education and research for data warehousing and business intelligence professionals worldwide. Eckerson oversees TDWI's research services and its BI Strategies education program at TDWI's World Conferences.

As an industry analyst, Eckerson has written and spoken on data warehousing and business intelligence subjects since 1995. He has published in-depth reports and articles on a variety of data warehousing and business intelligence topics. In addition, Eckerson has delivered presentations at industry conferences, user group meetings, and vendor seminars. He also consults with vendor and user firms.

Prior to joining TDWI in 2000, Eckerson was a senior consultant at the Patricia Seybold Group in Boston, where he started the firm's data warehousing and business intelligence practice. He can be reached at weckerson@dw-institute.com.

About the TDWI Report Series

The TDWI Report Series is designed to educate technical and business professionals about critical issues in business intelligence and data warehousing (BI/DW). TDWI's in-depth reports offer objective, vendor-neutral research consisting of interviews with practitioners and industry experts and a survey of BI/DW professionals worldwide. TDWI in-depth reports are sponsored by vendors who collectively wish to evangelize a BI/DW discipline or emerging technology.

Acknowledgements

TDWI would like to thank many people who contributed to this report. First, we appreciate the many users who responded to our survey, as well as those who responded to our requests for phone interviews. Second, we'd like to thank our report sponsors who reviewed outlines, survey questions, and report drafts. We would also like to recognize TDWI's account and production team: Denelle Hanlon, Theresa Johnston, Marie McFarland, Jeff Reeds, and Heather Truchot.

About TDWI

The Data Warehousing Institute (TDWI), a division of 101communications LLC, is the premier provider of in-depth, high-quality education and training in the BI and DW industry. TDWI offers quarterly educational conferences, regional seminars, onsite training, professional membership, leadership awards, print and online publications, and a public and private (Members-only) Web site.

Definitions

The TDWI survey defined terms used in the survey on an opening page and in hyperlinks within relevant questions. The definitions are below:

Data Warehouse: A shared, analytic data structure that supports **multiple** subjects, applications, or departments. There are three types of data warehouses—centralized, hub-and-spoke, and operational data stores.

Centralized Data Warehouse: A data warehouse residing within a single database, which users query directly.

Hub-and-Spoke Data Warehouse: A data warehouse that stages and prepares data for delivery to downstream (i.e., dependent) data marts. Most users query the dependent data marts, not the data warehouse.

Operational Data Store (ODS): A "data warehouse" with limited historical data that supports one or more operational applications with sub-second response time requirements. An ODS is also updated directly by operational applications.

Data Mart: A shared, analytic data structure that generally supports a **single** subject area, application, or department. A data mart is commonly a cluster of star schemas supporting a single subject area.

Dependent Data Mart: A data mart that extracts data from a hub-and-spoke data warehouse and uses data definitions and a data model that are consistent with that data warehouse.

Independent Data Mart: A stand-alone data mart with a model and rules that don't conform with other marts or warehouses—it often extracts data directly from transaction systems, not a data warehouse. Also called "analytic silos," independent data marts by definition are non-integrated.

Spreadmart: A spreadsheet or desktop database that functions as a personal or departmental data mart whose definitions and rules are not consistent with other analytic structures.

Federated Marts or Environments: An architecture that leaves existing analytic structures in place, but links them to some degree using shared keys, shared columns, global metadata, distributed queries, or some other method.

Research Methodology

Report Scope. This report is designed for business and technical executives who would like to learn more about the various ways to consolidate data marts and other analytic structures (i.e., data warehouses, operational data stores, and spreadsheets.) The report provides an overview of the benefits and challenges of consolidating analytic structures, and describes the tradeoffs for potential strategies.

Methodology. The research for this report is based on a survey that TDWI conducted in the winter of 2004, as well as interviews with data warehousing practitioners, consultants, industry analysts, and report sponsors.

Survey Methodology. TDWI contacted business intelligence and data warehousing (BI/DW) professionals in its database and 101communications' database. (TDWI is a business unit of 101communications.) In total, 672 individuals responded to the survey, but only 615 of them were qualified to complete the survey in its entirety. Results from respondents who work at a vendor firm or are a professor or student (57 total) were not calculated.

Survey Demographics. A majority of the qualified survey respondents (59 percent) who took the survey are corporate IT professionals. The remainder include independent consultants (30 percent), and business sponsors/users (11 percent). The largest percentage of respondents were project managers (19 percent) or chief architects (17 percent), but respondents are employed in a range of BI and DW roles. Totals that do not equal exactly 100 percent are due to multi-choice questions and rounding.

Most Respondents Are U.S.-based Corporate IT Professionals

Demographics

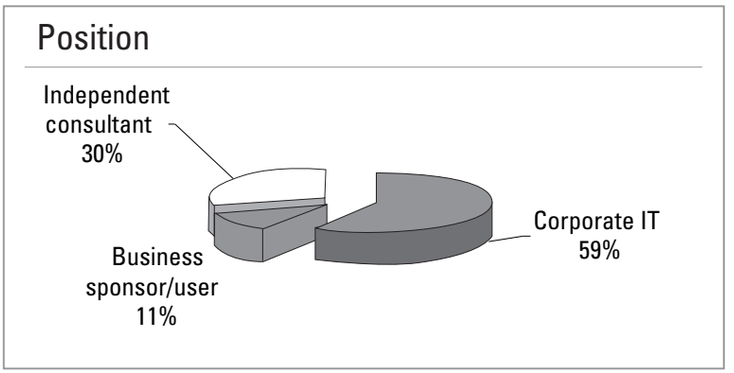


Illustration 1. Based on 615 respondents.

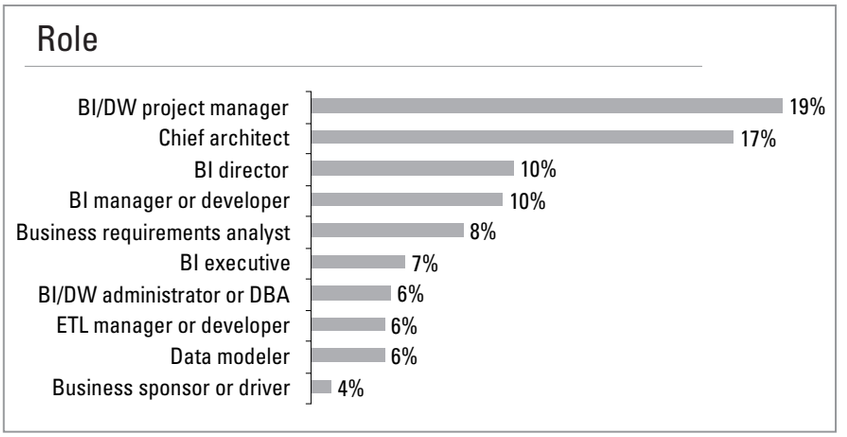


Illustration 2. Based on 615 respondents. Seven percent of respondents selected "other."

More than one-third of the qualified respondents (38 percent) work at companies that have revenues in excess of $1 billion, while 51 percent have less than $1 billion. (The remainder aren't sure about their firm's revenues.) Most respondents are based in the U.S. and work in a range of industries, the largest of which are the consulting and financial services industries. Consultants were asked to fill out the survey with their most recent client in mind.

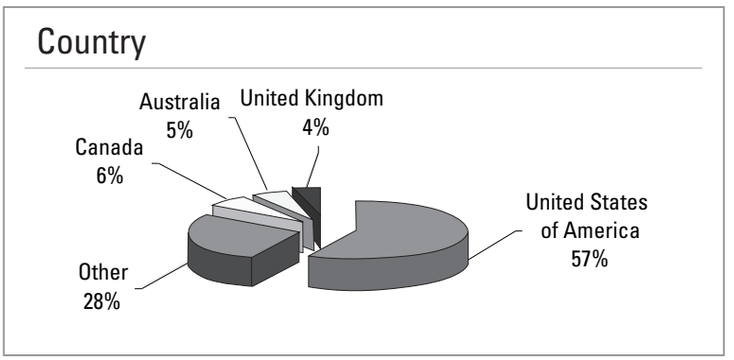


Illustration 3. Based on 615 respondents.

Drivers of Analytic Consolidation

Causes and Consequences of Proliferation

The one word that best describes the state of analytic data in large organizations is "fragmented." Despite their best intentions, CIOs are struggling to deliver consistent data that provides a single view across the enterprise. CIOs who seek this so-called "single version of the truth" must feel like they are playing an endless game of Whack-A-Mole—every time they stamp out a renegade analytic silo, another pops up elsewhere.

Empowered Groups Create Analytic Silos

It's easy to see why analytic silos proliferate. Most organizations empower business units, functional groups, and individuals to adapt products and services to local markets. Empowered with budgetary authority, these groups build both operational and analytical systems to support their initiatives. Sometimes, they build new analytical systems even when another system contains the data they need.

"I'm amazed at the power of the 'not invented here' syndrome," says a data warehouse manager at a large forest products company who wished not to be named. One business group at his firm spent a year and a half delivering information already contained in the central data warehouse. "They did this partly to protect their turf, partly so they could continue using tools they're familiar with, and partly because they could."

Business Events. But organizational dynamics aren't the only reason for the proliferation of analytic data. External events, such as mergers or acquisitions, create a multiplicity of redundant analytic structures. This is also true for internal reorganizations, especially when a decentralized firm decides to centralize its operations. In both cases, formerly autonomous groups now own multiple overlapping analytical systems

Firms Are Cleaning up the Excesses of the Late 1990s

Mergers, acquisitions, and reorganizations occurred at a lightening-fast pace in the late 1990s as the U.S. economy sizzled. Now that the economy has slowed, organizations are trying to clean up from the excesses of those years and bring order to chaos by consolidating redundant, analytic silos, among other things.

Types of Analytic Structures. Whatever the impetus, organizations now find themselves loaded down with many different types of analytic structures, including data warehouses, data marts, operational data stores (ODSs), and spreadmarts. (See definitions on page 3.)

On average, organizations have two data warehouses, six independent data marts, 4.5 ODSs, and 28.5 spreadmarts. (See the section: "Trends in Analytic Consolidation Projects" to follow.) But, we interviewed data warehousing managers at several organizations who estimated that they had dozens of data marts and hundreds of spreadmarts. Some couldn't even begin to estimate how many spreadmarts exist in their organizations.

Dueling Spreadmarts Frustrate Executives

Consequences of Proliferation. The problems that result from this rampant proliferation are twofold. First, executives become frustrated because they can't get the data they need to assess the performance of their company. They get apoplectic when managers spend more time in meetings arguing about whose data is right rather than developing strategies and plans to achieve corporate goals. Without a consistent view of performance across the enterprise, executives know they can't do a good job running the company.

Analytic Silos Increase Costs by 30 to 50 Percent

Second, these redundant, non-integrated analytic structures are wasteful. When deployed as physically distinct systems with separate hardware, storage, software licenses, data feeds, and staff, these systems can drive up data warehousing costs by 30 to 50 percent, according to some data warehousing managers.

Consolidating these structures can save organizations millions of dollars a year and deliver a quick return on investment (ROI). (See "Trends in Analytic Consolidation Projects" for specific costs and ROI figures.)

Remedies for Proliferation

The Disease Is the Cure. Ironically, the cure for the proliferation of analytic silos is to create another analytic structure. To stamp out analytic silos, organizations need to implement an enterprise data warehouse (EDW) that provides sustenance for all past, present, and future analytical structures. The key is to plant the EDW deeply enough and broadly enough within the information architecture and corporate culture that it becomes the defacto analytic structure within the organization.

The EDW only works, however, if the organization comes together to hammer out definitions and rules for commonly used terms and calculations, such as "net margin" or "sale" or "profit." Standardizing the meaning of shared data elements—often referred to as metadata—is often more challenging than consolidating the actual physical structures. Once this exercise is completed, the EDW becomes the repository for shared data, rules, definitions, and other metadata used by multiple analytic applications.

"Successful consolidation projects require organizations to standardize metadata but also provide a way to give users and administrators ready access to those definitions and rules so they can better understand the nature of the data they need to analyze or manage," says David Lyle, vice president of product strategy at Informatica Corp.

Success Requires: (1) CEO Mandate, (2) Corporate Commitment, and (3) Time and Money

Consolidating myriad analytic structures into a single EDW is challenging, but often the hardest task is to overcome legacy attitudes and culture surrounding the use of information. Bringing order out of analytic chaos requires three things: (1) a mandate from the CEO's office, (2) an enterprise commitment to delivering consistent information, and (3) a substantial investment in time and money

Consolidation in a Decentralized Firm. For example, the forest products company mentioned previously had four data warehouses and about 30 data marts, most of which were acquired through a series of acquisitions prior to 2000. Coupled with a decentralized organization structure, the firm spent several years trying to reconcile these multiple environments without much success. Then, a new CIO decided to consolidate the IT department into a central group and standardize on a single vendor for both operational and analytical systems.

"In a large, multinational organization with fairly autonomous divisions and operating groups, it was impossible to contain the proliferation of independent data marts and other decision support systems," says the firm's data warehousing manager. "Not until our executives decided to centralize IT could we consolidate and harmonize our data."

Benefits of Consolidation

According to TDWI research, there are many reasons organizations want to consolidate analytic silos. However, by far and away, the most prevalent driver is the need to have consistent data across the enterprise. Almost all survey respondents (90 percent) rated this factor as "very high" or "high." This was followed by "reducing costs and overhead" (71 percent) and "standardizing the IT portfolio" (50 percent). (See illustration 4.)

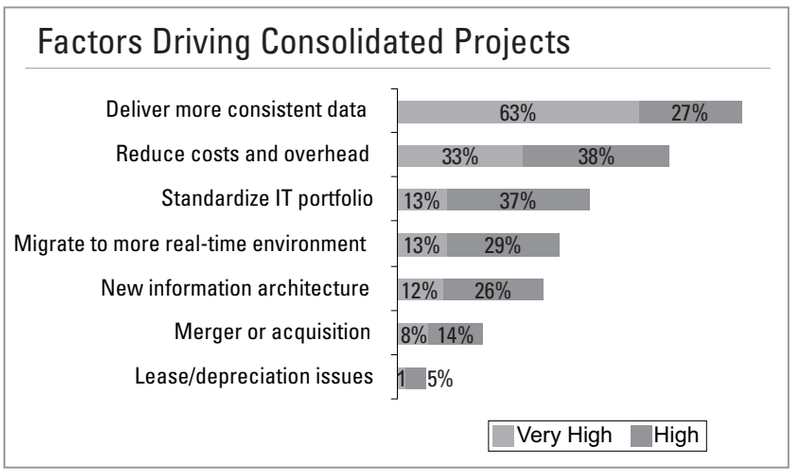


Illustration 4. Organizations primarily consolidate analytic structures to improve data consistency and reduce costs. Based on 615 responses and filtered by "very high" and "high" ratings.

"Our executives got fed up with seeing different numbers," says Dirk de Wilde, data architect at Canadian National Railway Company (CN). "The lack of consistent data made it impossible for them to deploy a performance management system to monitor key metrics. They got frustrated."

Canadian National Railway Company: From Silos to Standard Data

Canadian National Railway Company. In 2000, CN had 17 independent data marts and reporting systems running on various platforms and database management systems. The data marts were created by fairly autonomous business groups with their own IT staffs, and three distinct development groups within IT. Although a central IT group was building an EDW on a mainframe, it met considerable resistance from the business groups.

Finally, senior executives decided that the only way to get the situation under control was to outsource IT. However, when this option proved too expensive and inflexible, they decided to consolidate all BI operations under a single manager. With sufficient executive backing, the new BI manager initiated an overall redesign of the architecture and created an EDW that has gradually subsumed most of the firm's independent data marts and reporting systems.

Today, CN's EDW integrates data from more than 15 source systems, supports 15 dependent star schema data marts, and delivers reports and queries to more than 2,000 users. Strategically, the UNIX-based EDW supports an executive performance management system that generates hundreds of standard KPIs for executives and managers throughout the firm.

"Everyone Now Looks at the Same Metrics and Data."

—Dirk de Wilde

"Every morning, executives and managers look at the same reports and numbers in our corporate measures system and take immediate action to address problems. This has enabled the firm to operate more efficiently and offer better customer service without increasing costs," says de Wilde.

In addition, the new EDW has enabled CN to shut down eight data marts and legacy reporting systems, saving money in extra license fees, disk storage, and staff time required to maintain the independent marts. The firm plans to consolidate several remaining data marts and reporting applications in the next couple of years.

Trends in Analytic Consolidation Projects

Project Status and Drivers

While most organizations excel at proliferating analytic structures, few are experienced at consolidating them.

A Majority of Organizations Are Engaged in a Consolidation Project

According to TDWI research, only 11 percent of organizations have completed a project to consolidate analytic silos. However, a majority (56 percent) are in the process of designing or implementing a consolidation project and another 26 percent are exploring the idea. Only 7 percent have "no plans." (See illustration 5.)

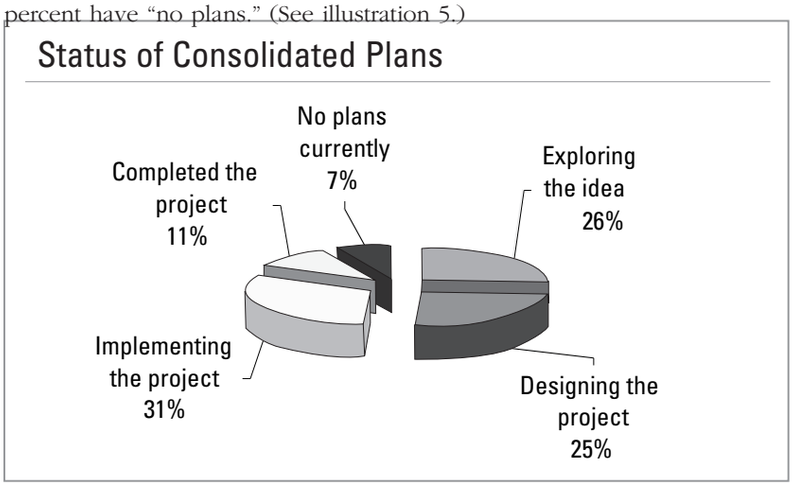


Illustration 5. Most organizations are starting a consolidation project. Based on 615 responses.

Surge in Interest. One reason for the current surge of interest in consolidation projects is the recent economic downturn, which has caused a slowdown in the number of IT projects. During this lull, IT groups designed new information architectures to streamline processing, reduce the number of suppliers, and save the company money. These initiatives have been reinforced by executive mandates to reduce costs and reduce the number of suppliers.

Strategically, many executives know that the strongest organizations leverage information as a competitive weapon. These executives understand that consolidating analytic silos and delivering a single version of internal truth is a prerequisite for most strategic initiatives today.

Business Events Often Kickstart a Consolidation Project

Precipitating Factors. Often, a business event serves as a catalyst to kick-start a consolidation project. These events thrust the need for more consistent information out of the shadows and into the limelight. Precipitating events help galvanize executive attention and support for a consolidation project.

Typical precipitating events are:

- **A New Top Executive**. A CEO, CFO, COO, or CIO with a vision for integrated data or who previously worked at a company with integrated data often initiates consolidation projects. These executives often call upon the head of IT to drive the projects. (See illustration 6.)
- **Strategic Business Initiative**. Executives who want to enhance business performance, improve customer relationships, streamline the supply chain or initiate some other technology-driven project always run smack into the wall of inconsistent data. As part of their project, they fund a consolidation project.
- **A Merger or Acquisition.** The combined company inherits multiple, incompatible analytical silos which must be integrated quickly. In fact, some companies use an enterprise data warehouse as a vehicle to facilitate acquisitions.
- **Organizational Restructuring.** Companies that recentralize IT or other functional groups undertake consolidation projects to integrate the organizations and their data.
- **Standardizing Application Suppliers.** To better integrate operational applications, organizations often standardize on a single applications supplier. This decision often spills over to the analytic side of the house, when application vendors also provide data warehousing solutions. Here, all data warehouses and data marts are converted to the ERP vendor's analytic platform.
- **Obsolete Technology.** When a vendor sunsets its technology, organizations use the conversion process to also re-architect their analytic environments. For example, many Red Brick and Informix shops are converting to new database management systems and consolidating analytic silos in the process.
- **New Regulations.** New financial regulations, especially the Basel Accord and international privacy regulations, are forcing organizations to consolidate analytic silos and re-architect information systems to improve control and data accuracy.

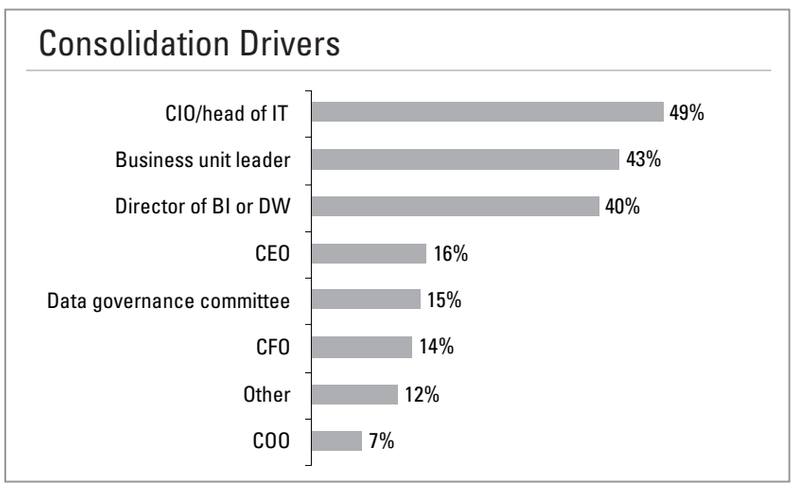


Illustration 6. While CXOs typically initiate, approve, and fund consolidation projects, CIOs and IT directors drive them. Based on 615 respondents.

Target Analytic Structures

The More the Better. When it comes to deciding which analytic structures to consolidate, organizations are pretty clear: consolidate them all! The most popular candidates are central data warehouses and independent data marts, followed by operational data stores and operational reporting systems. Only hub-and-spoke data warehouses aren't much of a target since they integrate data marts by default. Hub-and-spoke data warehouses would be a target for consolidation if a merger, acquisition, or reorganization created duplicate data warehouse environments. (See illustration 7.)

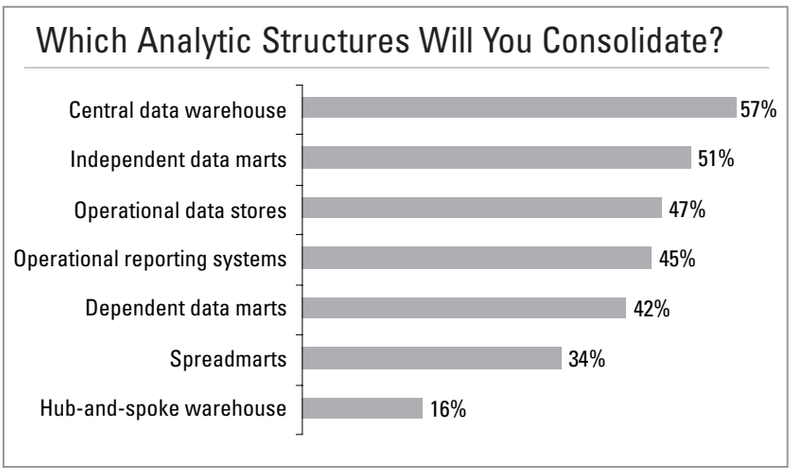


Illustration 7. Organizations plan to consolidate all types of analytic structures, especially central warehouses and independent marts. Based on 615 respondents.

Duplicate Data. One reason organizations are eager to merge these disparate structures is that they contain a lot of redundant data. In fact, two-thirds of our survey respondents said that a third or more of the data in their analytic structures is redundant. And 42 percent said at least half or more of the data in these diverse structures is redundant. (See illustration 8.)

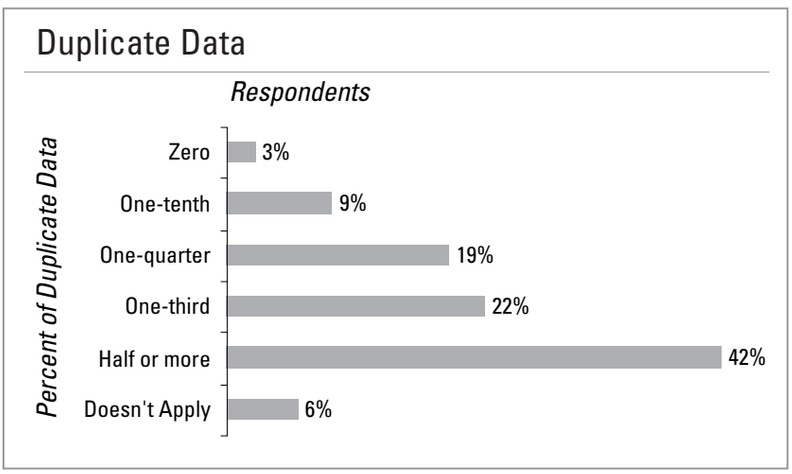


Illustration 8. Most organizations have a lot of duplicate data among their analytic structures. This is not a problem if the data is integrated, as in a hub-and-spoke architecture. Based on 615 respondents.

Of course, redundant data doesn't always mean non-integrated data. A hub-and-spoke architecture duplicates data between the data warehouse and its marts to ensure integration. Also, an ODS often acts as a source of data for the warehouse since it is more operational in nature.

A Big Job. Consolidating duplicate data and redundant analytic structures is not for the faint of heart. That's because most organizations have dozens of analytic structures that they would like to consolidate. As mentioned above, organizations have an average of two data warehouses, six independent data marts, 4.5 ODSs, and 28.5 spreadmarts left to consolidate. To date, they've only consolidated about one-third of all the structures in their environment. But acquisitions, mergers, and reorganizations make this number a perpetually moving target. (See illustration 9.)

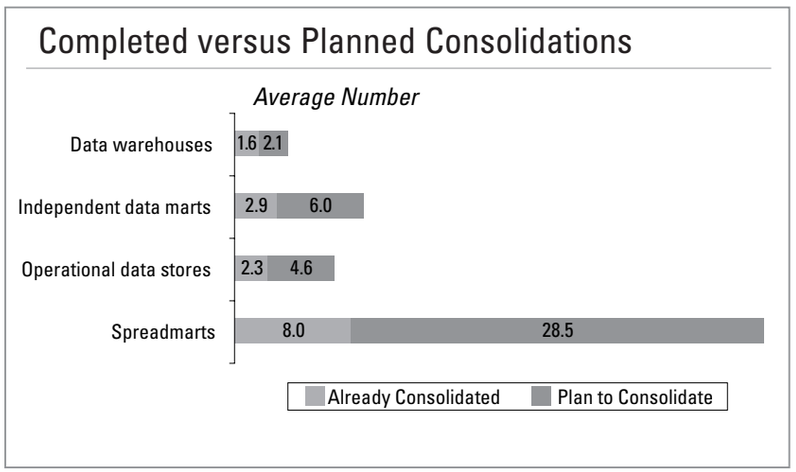


Illustration 9. Organizations plan to consolidate all types of analytic structures, especially central warehouses and independent marts. Based on 615 respondents.

On average, organizations have consolidated a slightly higher percentage of data warehouses (42 percent) and a slightly lower percentage of spreadmarts (22 percent). This makes sense given the relative numbers of these structures. In fact, it's safe to say that the smaller the structure, the more likely it is to proliferate and the harder it is to consolidate.

The Smaller the Analytic Structure, the More It Will Proliferate

Insidious Spreadmarts. Thus, it's safe to say that spreadmarts are the most insidious analytic structures. One power user with Microsoft Excel or Access on his or her desktop can easily create and maintain one or more spreadmarts for various executives or groups. Most data warehousing managers don't know how many spreadmarts actually exist in their organizations.

"Every week during our assessment phase, I discovered a new spreadsheet that was a source for customers or contacts. In the end, I found more than 23 files in spreadsheets, Foxpro, Act, Goldmine, and Paradox databases that contained customer information," says Wanda Black, director of information resource management at a privately held manufacturing firm.

ODSs Have Unique Issues. While spreadmarts may be difficult to identify and tear away from their protective owners, ODSs pose a different problem. Since ODSs usually support multiple operational applications, these applications will need to be rewritten or at least, redirected. In addition, ODSs are usually highly optimized to deliver sub-second response times.

"Consolidating ODSs Risks Ruining the Optimization." —Philip Russom

"When you consolidate ODSs, you run the risk of ruining the optimization. As a result, a lot of companies talk about [consolidating ODSs] but don't consolidate them," says Philip Russom, principal analyst at Forrester Research.

Project Duration

Time to Consolidate. Not surprisingly, the most complex analytic structures take the most time to consolidate and the simplest structures take the least time. Our survey shows that data warehouses take an average of 9.75 months to consolidate (the average of planned and completed consolidations), followed by operational data stores at 6.82 months, data marts at 6.03 months, and spreadmarts at 5.69 months. (See illustration 10.)

It Will Take 23 Years to Consolidate All Analytic Silos, if Done Sequentially

Twenty-three Years of Hard Labor. If we multiply the number of analytic structures that organizations plan to consolidate (see the section "Completed versus Planned Consolidations") by the average time to consolidate each type of structure, then it will take organizations 277 months—or 23 years—to complete their projects if done sequentially! Obviously, organizations need to consolidate multiple structures in parallel—or implement a true EDW that can replace all warehouses, marts, and spreadmarts at once—or more likely—over a period of several years. (See illustration 10.)

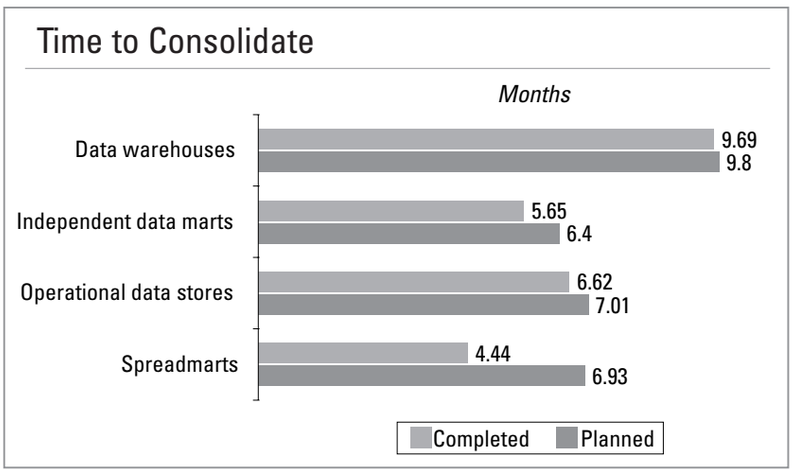


Illustration 10. Data warehouses take the longest time to consolidate, approximately 9.75 months. Based on 209 respondents who have completed a consolidation project.

Long-Term Perspective. Given the numbers of analytic silos within organizations and the time required to consolidate them, it's clear that most organizations have a long way to go before completing their consolidation projects.

"It may take us five to ten years, but we'll get there. We may not win every battle, but if we stick to the fundamentals and keep our eyes on the goal, we'll get consistent data," says Jim Gallo, enterprise architect at Worthington Industries, a $2 billion metals processing manufacturer in Columbus, OH.

Worthington Is Ready to Embrace Centralized Data and Analysis

Organizations with autonomous divisions or business units face an uphill battle trying to deliver a consistent view of data across the enterprise. Their decentralized organizational structures make them fertile breeding grounds for analytic silos. To bring order to such chaos, organizations usually require a strong-willed executive to fundamentally change the structure and culture of the firm. However, sometimes a bottom-up, grassroots strategy succeeds when autonomous units or divisions see the benefits of coming together to share critical resources.

For example, Gallo said that many divisional executives at Worthington Industries are ready to embrace an enterprise approach. "Our executives see the value in receiving consistent, accurate information and are willing to provide the necessary resources to create a system that works for the whole company."

To deliver an enterprise resource while preserving divisional autonomy, Gallo is planning to incrementally develop data standards and a normalized data model to populate divisional data warehouses. "Each division will get its own data warehouse, but since we'll conform dimensions and share metadata, we will preserve an enterprise view as well."

Cost Justification

Although reducing costs is an important reason to consolidate analytic silos, it usually takes a backseat to more strategic concerns.

"We did some ROI analysis, but that's not the driver. Although we expect to reduce costs by 30 percent or more, what is really driving this project is the need to comply with the Basel Accord and improve the bank's risk analysis capabilities," says the lead architect at an Australian bank.

According to our survey, less than a third of organizations analyze what it costs to support independent data marts or data warehouses or migrate them to a new environment. Similarly, less than one-third calculate the ROI of migrating analytic structures to a new environment. (See illustration 11.)

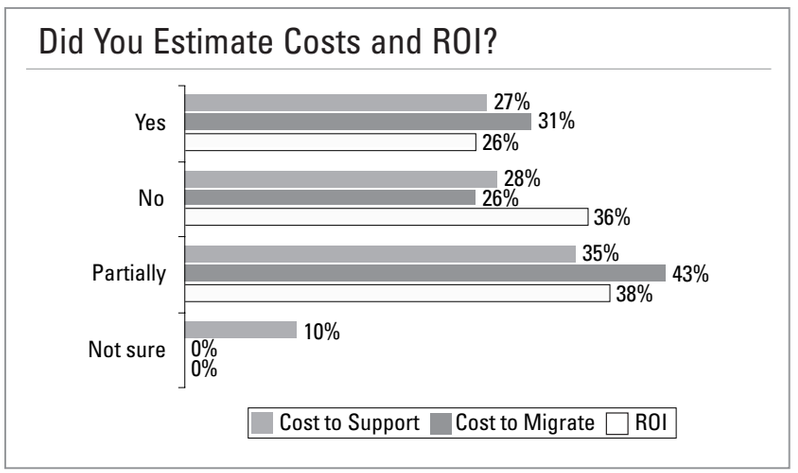


Illustration 11. Less than a third of organizations calculated what it costs to support or migrate analytic structures. Based on 573 respondents.

Russ Vaughn, senior vice president of data warehousing at Bank of America, agrees that while ROI and saving money is important, it's not the primary reason business people fund or endorse consolidation projects. "You have to present the project as a business benefit to them. They get more interested if you say you can help them generate $2 million in additional revenue rather than reduce costs by $3 million by consolidating systems."

Average Costs and Projected ROI. TDWI research shows that consolidation projects deliver a hefty payback on the investment.

TDWI research indicates that it costs $614,000 a year to maintain an independent data mart and $1.597 million a year to maintain a data warehouse. (See illustration 12.) If a company has six independent data marts that it wants to consolidate (see the "Completed versus Planned Consolidations" chart on page 11), then it spends about $3.7 million a year to support them.

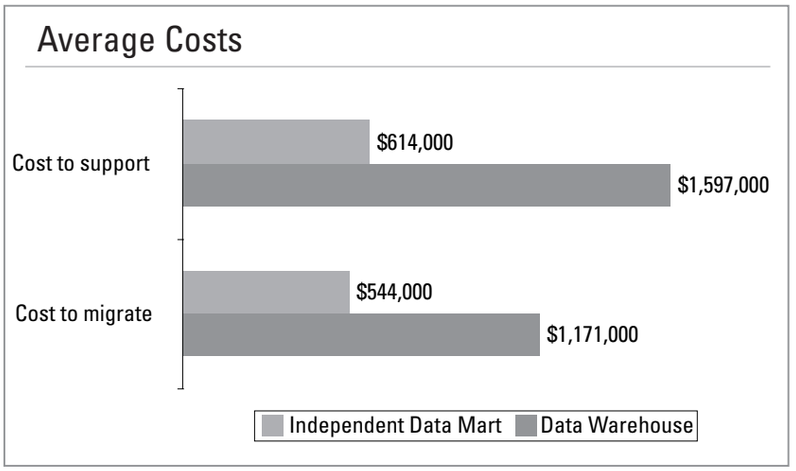


Illustration 12. Data warehouses cost more to support and migrate to a consolidated environment than independent data marts. Based on 150 respondents who have already consolidated analytic structures.

Migration Costs. Then, if the company decides to undertake a consolidation project, it will need to spend $3.3 million over the course of three years (six marts times six months each) to migrate the independent data marts. At the same time, it will spend $1.6 million per year to build and support the new data warehouse, or $4.8 million over three years.

Total three-year migration costs are $8.1 million. On the flip side, the organization will save $6.4 million over three years by shutting down one independent data mart every six months. At the completion of the project, it will have spent $1.7 million more than it saved.

Consolidation Projects Yield $2.1 million in Cost Savings Two Years After Project Is Complete

Break Even After 1-2 Years. However, every year thereafter, the project nets the organization $2.1 million in cost-savings ($3.7 million from shuttering six data marts minus $1.6 million to support the new data warehouse). Thus, the company breaks even the first year after the completion of the project with a $400,000 cost savings. The second year yields a total of $2.5 million in overall cost savings. If we figure in the net present value of money, then the break-even point for consolidation projects occurs even earlier.

Alstom Power Generates High ROI. Our estimates match those generated by Alstom Power, which is now building an EDW to consolidate numerous analytic structures, including a data warehouse, a reporting repository, two departmental data marts, and numerous Access databases. Alstom expects a $3.5 million payback over 2.8 years from its project, according to Michael Sykes, U.S. manager of data warehousing at the company.

Our survey data also confirms the same level of ROI. Among organizations that calculated ROI for their projects, the average ROI was $3.34 million over 2.1 years. (See illustrations 13 and 14.)

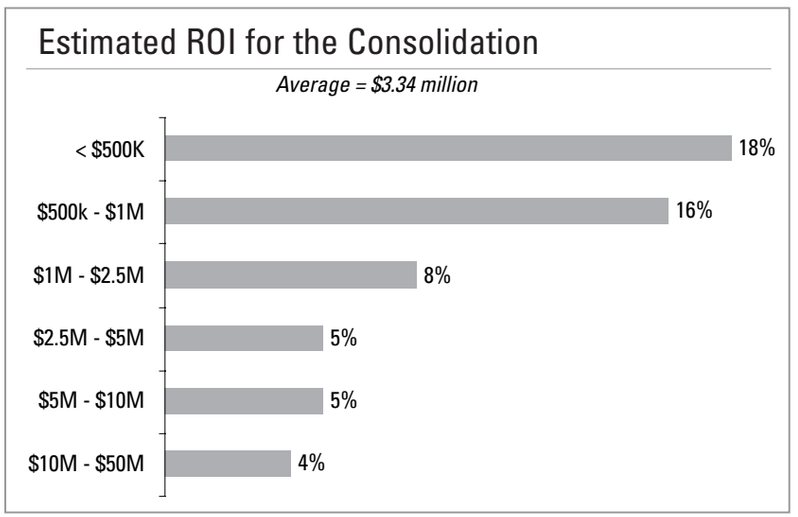


Illustration 13. The average ROI for consolidation projects is $3.34 million. Based on 150 respondents. Forty-four percent of respondents selected "don't know."

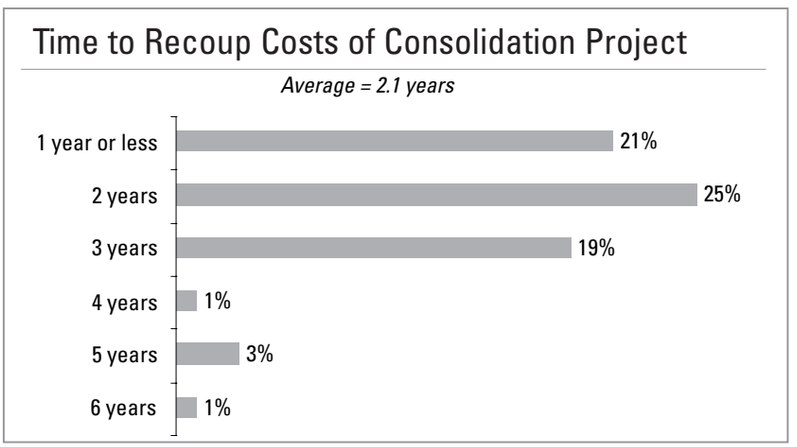


Illustration 14. The average payback period for consolidation projects is 2.1 years. Based on 150 respondents. Thirty percent of respondents selected "don't know."

Cheik Daddah, director of the value consulting team at Teradata, a division of NCR, has done exhaustive studies with dozens of clients on the cost of supporting data marts. He believes that the fully loaded costs of supporting data marts range between $1 million and $2 million. Daddah's cost estimates include all hardware, software, networking, support, labor, and development work. Most of the marts Daddah examined were physically distinct systems.

"Not looking at this entire cost picture is one of the reasons why companies end up with data marts. Managers think data marts are cheap but the real costs are hidden/buried, invisible to the naked eye," says Daddah.

If Daddah's estimates are correct, then the ROI for consolidation projects is better than TDWI's estimates. Daddah's research is validated by AMR Research, which believes that data marts cost firms $1.4 million a year on average. Given these estimates, AMR Research says the cost savings of a data mart consolidation project outweigh the migration costs after the third or fourth data mart is consolidated. However, the break-even point slips to six or seven marts in a moderate cost-savings scenario or 16 to 17 data marts in a conservative cost-savings scenario, according to Bob Parker, vice president of industry strategies at AMR Research in Boston.

Not All Data Marts Are Good Candidates for Consolidation

Target Candidates. Of course, not all data marts are good candidates for consolidation. "We don't believe you need to consolidate all your data marts," says Teradata's Daddah. "Often, there are data marts that can't be consolidated because of political or legal reasons or because the cost savings don't justify it."

In contrast, the most promising candidates for consolidation are data marts that run on different platforms and are managed by separate IT staffs. Consolidating these independent data marts can save companies more money than those that run on the same platform and are managed by the same IT staff. Of course, as we have seen, saving money is not the prime incentive for consolidating marts; delivering a consistent view of the enterprise is.

End-State Architectures

After the Pain of Analytic Silos, Most Firms Want to Centralize As Much As Possible

Once an organization has made a commitment to consolidating analytic structures, it needs to plan how to migrate from its current chaotic environment to a standardized architecture. Generally, once an organization has lived through the pain of analytic silos, it is adamant about centralizing as much data, resources, and infrastructure as it possibly can.

"A centralized environment ensures that people are looking at the same source of information and getting consistent reporting and interpretations," says Vaughn of Bank of America. "Once something leaves our hands, we can't guarantee the validity of the information."

When asked to describe the architecture of a new or planned consolidated environment, almost two-thirds of respondents (61 percent) selected "central data warehouse" while only 16 percent selected "hub-and-spoke data warehouse." Only a fraction of respondents chose more distributed architectures, such as "conformed data marts" or "federated." (See illustration 15.)

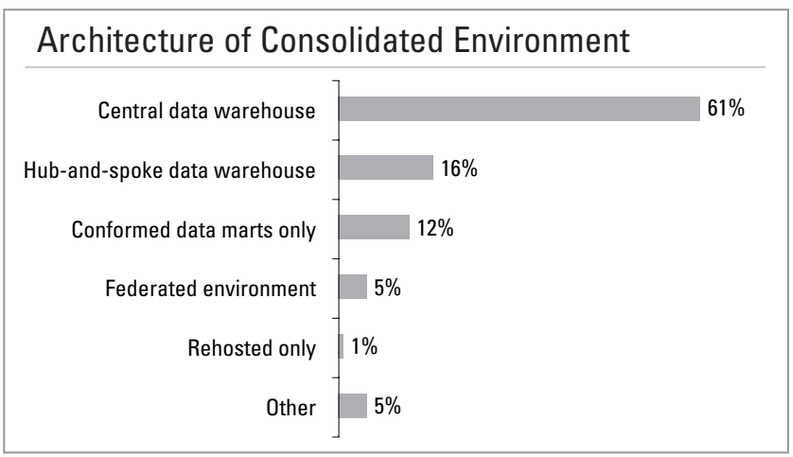


Illustration 15. A central data warehouse architecture is the preferred post-consolidated environment, although this apparently doesn't preclude the use of data marts, contrary to TDWI's definition. Based on 615 respondents.

Architectural Approaches to Consolidation

TDWI's Survey Data Is Misleading. Hub-and-spoke DWs Are the Most Prevalent

User Perceptions. Although TDWI's definition of a central data warehouse precludes the use of data marts, it's clear that most survey respondents don't share this viewpoint. Even though TDWI posted its definition at the beginning of the survey and in links within the relevant questions, most respondents who selected "central data warehouse" indicated in a follow-up question that their data warehouse also supports data marts, which is the definition of a "hub-and-spoke" data warehouse.

Thus, if most data warehouses support dependent data marts, then the most popular data warehouse architecture—by TDWI's definition—is a hub-and-spoke data warehouse. (Author's note: This shows that organizations aren't the only victims of inconsistent semantics and metadata; surveys experience the same problems!)

Dependent Data Marts Abound. However, the trend towards greater centralization holds true. The most popular types of data marts are centralized—that is, they run within the same database or server as the data warehouse. Thus, a "central data warehouse" really means "logical data marts" for most respondents.

Survey respondents said their data marts consisted of "database views" (36 percent), or logical tables in the warehouse database (36 percent), or separate database instances on the same server (33 percent.) A much smaller percentage of respondents said they had "physically distinct" data marts. About one-fifth of the respondents (22 percent) said their data marts ran on different "local" servers, while 14 percent support data marts running on remote servers. (See illustration 16.)

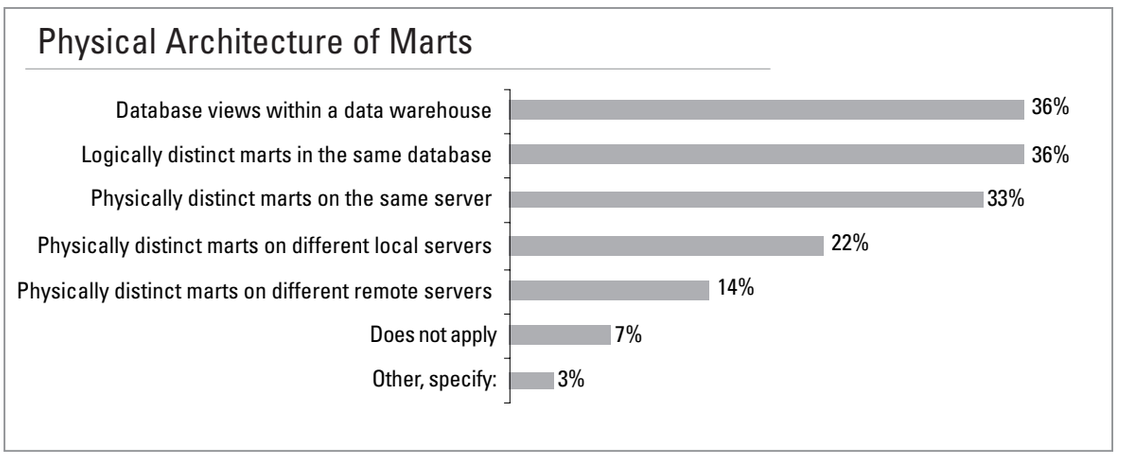


Illustration 16. Although most organizations have "central data warehouses," they also support data marts, mostly running within the same database or server as the data warehouse. Based on 615 respondents..

Hub-and-Spoke Redefined. When we examine the data more closely, it's clear that most respondents believe that a central data warehouse supports logical data marts, while a hub-and-spoke data warehouse supports physically distinct data marts.

For respondents with central data warehouses, 55 percent have data marts that are either database views (30 percent) or logically distinct marts in the same database (25 percent). In contrast, 60 percent of respondents with "hub-and-spoke" data warehouses have physically distinct marts of some type. (See illustration 17.)

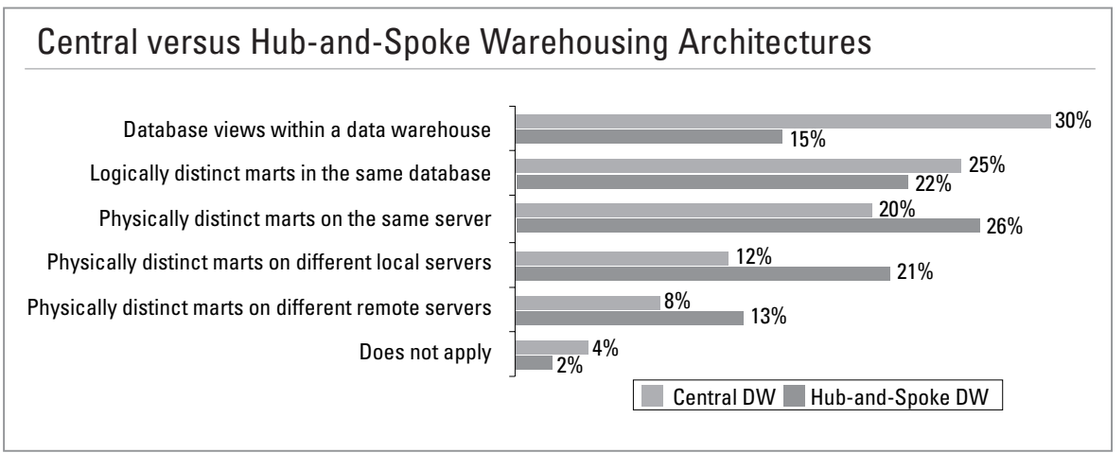


Illustration 17. Central data warehouses support mostly logical data marts (database views and logically distinct marts), while hub-and-spoke data warehouses support mostly physically distinct data marts.

Although our results show a huge move to "centralized" data warehousing, this does not mean that organizations are abandoning data marts and letting users query the data warehouse directly. In fact, it appears that the opposite is true: a majority of organizations now implement logical data marts co-located with a "central" data warehouse.

Migration Options

Whatever the target architecture, organizations need to devise a technical strategy to consolidate silos and deliver a single version of the truth. Unfortunately, there is no single path to analytic nirvana, although there are a number of well-worn routes, which we describe below.

To figure out the best consolidation strategy for your organization, you need to assess a number of factors. These include:

- Strategic focus of executives (cost-cutting versus strategic initiatives).
- Organizational structure (centralized versus decentralized) and degree of autonomy within divisions or lines of business.
- Type of merger or acquisition. Merger of equals or acquisition of a smaller firm by a larger firm.
- Time available to complete the project.
- Money available for the project, either within an existing budget or a special project-specific allocation.
- Nature of analytic silos—data warehouses, data marts, operational data stores, reporting systems, or spreadmarts.
- Stability of transactional environment, i.e., does the firm plan to replace legacy systems in the near future?
- User satisfaction with existing reports and query tools.
- Geographic scope of organization and analytic silos.
- Available tools/technologies.

Eight Migration Options. Through interviews with dozens of organizations that have consolidated analytic silos, we have defined eight migration strategies. Some organizations only use one strategy; others adopt different strategies at various phases in the migration to a consolidated environment; and others are forced to switch strategies as business events change.

The eight consolidation strategies fall into two major categories: rehosting and integration. Rehosting involves moving existing analytic structures to a new platform without combining or changing them. Some call rehosting a "forklift" operation because it simply moves structures to a new shared platform without integrating them. Integration on the other hand, involves merging metadata and data models into a new analytic structure.

In general, a majority of firms (55 percent) rehost and integrate at the same time. Only a third rehost first and then integrate later. (See the Bank of America example to follow.) Very few rehost only. (See illustration 18.)

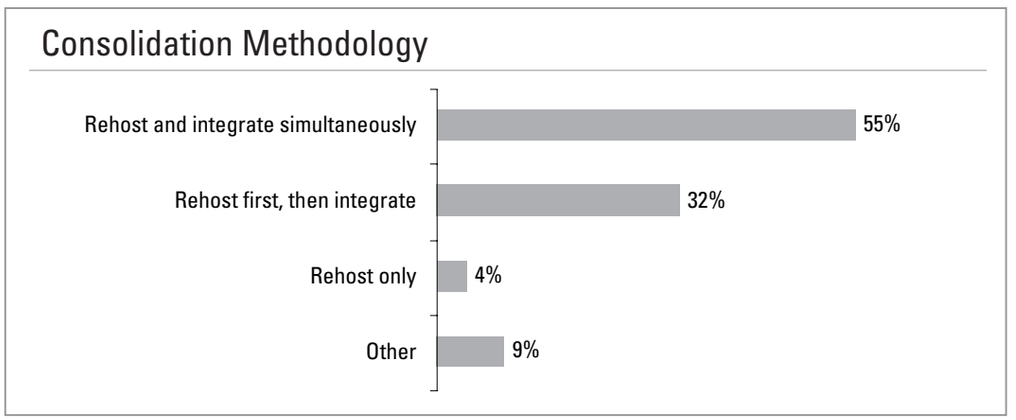


Illustration 18. Most organizations rehost as part of a strategy to integrate analytic silos, not as an endpoint in itself. Based on 615 respondents.

There are seven integration strategies, four of which deliver centralized analytic structures, and three of which result in federated or distributed structures. Most organizations prefer to centralize analytic data rather than to distribute it. Organizations that pursue distributed consolidation strategies usually have a decentralized organizational structure which gives considerable power and autonomy to its operating groups.

The eight consolidation strategies in a nutshell are:

Physical Strategies

1. **Rehost.** Move existing analytic structures onto a single platform.

Centralized Strategies

2. **Start from Scratch.** Build a new data warehouse instead of designating or merging existing ones.
3. **Designate and Evolve.** Designate an existing data warehouse or mart as the corporate standard and migrate other structures to it, either immediately or over time.
4. **Backfill.** Implement a staging area/warehouse behind existing data marts to consolidate extracts and data for marts to pull from.

5. **Synchronize.** Synchronize remote operational data stores from a central reference repository.

Distributed Strategies

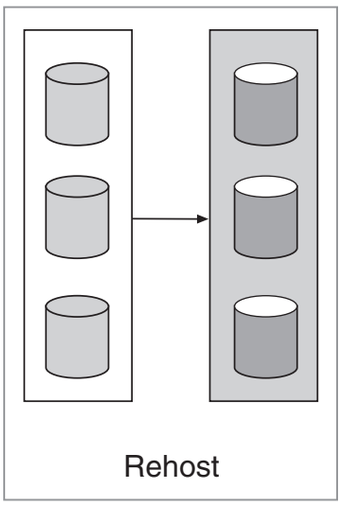
6. **Conformed Data Marts.** Conform the data models of existing data marts by standardizing shared dimensions.
7. **Mart of Marts.** Create an enterprise view across data marts by extracting data from them to create a new superset data mart.
8. **Distributed Query.** Create an enterprise view by querying multiple marts and reconciling results on the fly.

While in reality there are more than eight options, these seem to predominate. Let's now take a look at each of the migration options in more depth.

1. Rehost

Rehosting Doesn't Change the Application in Any Way

Organizations focused exclusively on reducing costs may simply opt to rehost existing analytic structures onto a single operating platform. This "forklift" option enables firms to eliminate multiple servers and the staff required to maintain them. Rehosting does not change the application in any way and does nothing to integrate data or deliver a single version of the truth. Its data model, metadata, transformation logic, and reports stay the same. Only client interfaces are redirected



Sometimes, organizations rehost to replace proprietary technology or when a vendor withdraws support for a product, such as a database management system, which is the case for many Red Brick and Informix customers, for example. Others rehost as a first step in a broader consolidation strategy.

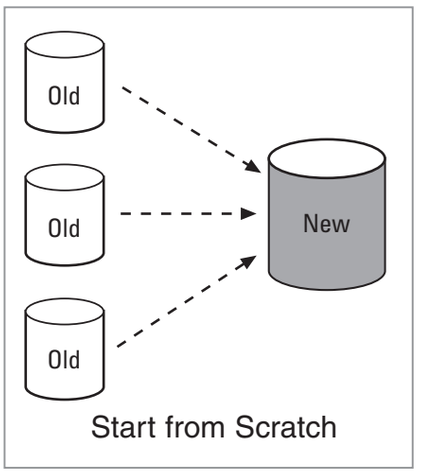
Bank of America Did a "Smart" Rehost

Two Step Process. Bank of America, like many other financial institutions, is the product of numerous mergers in the 1980s and 1990s. Following the 1998 merger of NationsBank and the former BankAmerica to form the new Bank of America, the executives of the combined organization decided to focus the bank on organic growth and customer retention. To support the information requirements of this new strategy, the bank recognized the need to integrate two massive data warehouses.

Because of the volume of data in the two warehouses (approximately 22TB), the bank decided it needed to work incrementally, but fast. So, the first thing it did was to rehost both data warehouses to a single platform to save costs. Once the two structures were rehosted, it then began building a new data warehouse to replace the rehosted structures. (See the following section "Start From Scratch.")

"Initially, we performed a 'smart rehost,'" says Bank of America's Vaughn. "We did minimal integration, adding some common keys and indexes. This made it possible for some users to look across both environments for the first time, but it was our first step towards a deeper integration, which included creating a new logical model and warehouse, which took another year to accomplish."

2. Start from Scratch



Ask any homebuilder and he will tell you it's easier to build a new home than renovate an existing one. The same concept holds true in data warehousing. Organizations faced with trying to reconcile a disintegrated analytic environment often decide that the easiest and most cost-effective option is to start anew.

"Our analytic environment was chaotic and there was little documented metadata. In the end, the best way to address our situation was to migrate everything to a new domain, essentially rebuild our environment from scratch," says the lead architect at an Australian bank who wished to remain anonymous.

The precipitating event for the bank was the Basel Accord, which motivated executives to get a better handle on risk exposure. The new risk data warehouse will consolidate three major data marts, hold nearly 4TB of data, and support hourly data feeds. The new data warehouse will be modeled broadly so that eventually it can subsume the bank's other data warehouses and data marts and provide a true, single source of information for the enterprise, says the bank's lead architect.

Borrowing from the Old. In most cases, the architects of the new environment borrow heavily from the old "legacy" environments. "We wanted to create a true enterprise data warehouse that was more comprehensive and up-to-date than the existing data warehouses," says Bank of America's Vaughn. "We borrowed some concepts from the existing data warehouses but we spent a year interviewing more than 60 business users to create our new logical model."

Evolving Legacy Environments. One problem when starting from scratch is trying to figure out what to do with the existing or "legacy" analytic silos. In some cases, the decision is easy. If the silos aren't being used much or users are unhappy with their performance, data quality, user support or something else, then it's a no-brainer to pull the plug on these silos as soon as possible.

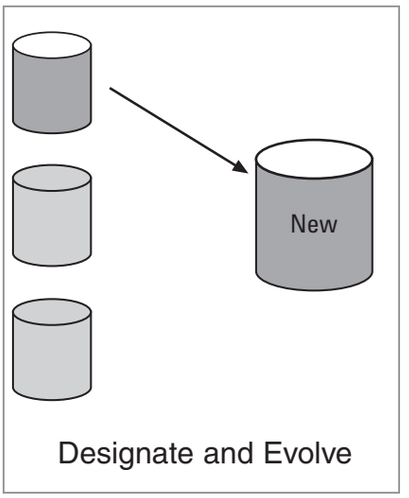
Coercion Rarely Works When Consolidating Analytic Silos

However, the cardinal rule in consolidation projects is never to prevent users from accessing their existing reports, according to William McKnight, president of McKnight Associates, a data warehousing consultancy based in Plano, TX. This may mean (1) keeping existing structures in place until they can be replaced by the new environment; (2) redirecting existing applications to run against the new environment; or (3) rewriting applications in a new tool within the new environment.

If the users are extremely attached to their application or tool, then sometimes the best option is to wait until the group is ready to migrate on its own, rather than trying to coerce them in any way.

Bank of America, for example, decided to take this type of evolutionary approach with its legacy data marts. It "grandfathered" these applications rather than try to rewrite them. But it decreed that all new or enhanced analytic applications had to run against the new data warehouse. "We chip away at the old applications," says Vaughn. "If [the business owners] want enhancements, we reengineer the applications to run against our new [Teradata] data warehouse."

3. Designate and Evolve



The "designate and evolve" approach involves designating one of the existing analytic structures to become the corporate standard. All other analytic systems and applications are then migrated into the designated environment.

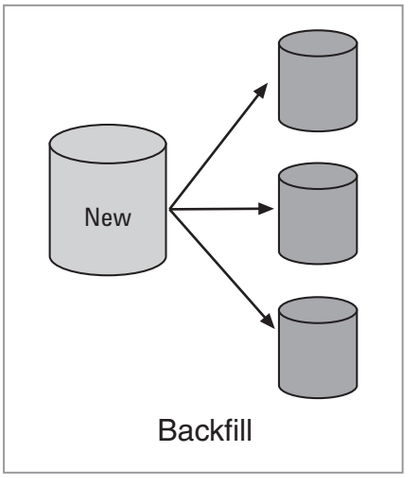
This frequently occurs when a larger company acquires a smaller one, and the data warehousing environments usually follow suit: the acquirer's data warehouse is appointed the corporate standard and the newly acquired data warehouse is folded into it. This approach also makes sense when a company makes a strategic commitment to implement products from a specific vendor, which is designated as the corporate "standard."

Simplifying an IT Portfolio Leads to Designating Winners and Losers

For example, after a series of acquisitions, the large forest products company mentioned earlier in this report decided to standardize and centralize its IT infrastructure in order to reduce costs, eliminate redundancy, and minimize data chaos. It started on the operational side, signing a contract with a major application vendor for a critical supply chain application.

Once the supply chain initiative got off the ground, the firm decided to consolidate its four data warehouses. Since the application vendor also provided a data warehousing solution, the firm decided to designate it as a standard. The firm is now migrating the other three data warehouses to the application vendor's data warehousing solution.

4. Backfill



Tug of War. When local groups hold too much power to allow corporate IT to build a new EDW from scratch or select "winners" and "losers" from among existing analytic structures, a politically acceptable approach is to backfill a data warehouse behind the existing structures.

Here, the data warehouse serves as a staging area for downstream marts. It consolidates all extracts and data feeds from source systems and logically integrates this data via keys and shared dimensions. Although this approach doesn't reduce the number of data marts, it does reduce the number of extract programs running against source systems.

Backfill a Data Warehouse Behind Existing Data Marts

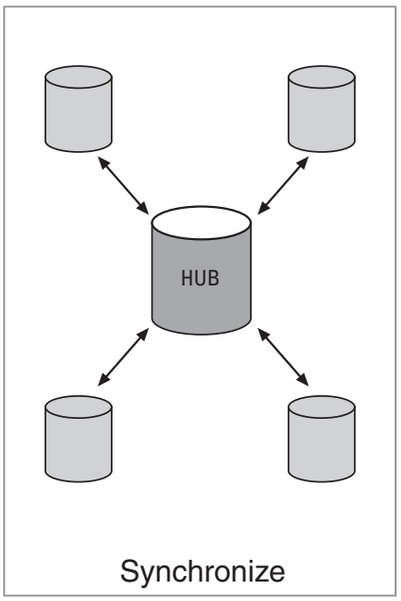
DaimlerChrysler. DaimlerChrysler Corp. used the backfill strategy when it built its data warehouse in the mid-1990s. Initially, it started building data marts for different functional areas, such as Warranty Analysis and Financial Services. However, as a traditional mainframe shop, it soon realized that building independent data marts could lead to data chaos, so it quickly altered its strategy. It gathered and modeled data requirements for all major business units and then built an enterprise data warehouse behind the two existing data marts. The enterprise data warehouse became the single source of corporate information used by more than one group.

Each group now downloads data from the data warehouse to their data marts across a high-speed switch. Although the data warehouse and data mart servers run in the same data center and are managed by the same team, each business unit owns its own servers.

"We needed an architecture that would preserve data consistency and allow business units to control their own data and servers, which was important to them," says George Mortis, manager of data resource management at DaimlerChrysler. "Now, we create shared data only once, and the groups can download it across a high-speed switch to create their own data marts."

5. Synchronize

The final centralized approach is synchronization. This approach makes sense when an organization needs to standardize reference data for core data entities, such as customers, products, suppliers, and so on, across a large number of operational applications. Often, data for these entities are captured by and maintained in multiple operational systems in different formats. This creates multiple instances of identical records, wreaking havoc with data consistency and standardization.



Use an ODS and Matching Engine to Synchronize Reference Data

To solve this master data management problem, many organizations create a centralized operational data store (ODS) to store a consolidated set of reference data for one or more entities. Then, whenever an operational system adds or updates an entity record in its own system, it submits a copy to the ODS, which checks the record against its reference set. If the record is new or updated, the ODS sends a copy to all other operational systems that subscribe to the ODS with a message to update their database with the new record. In this way, the ODS synchronizes reference data across multiple operational systems.[1]

This type of ODS is called a "data integration hub." It usually consists of several components:

- **Matching engine**. A matching engine uses rules or fuzzy logic to identify and match duplicate records. The engine also assigns a unique key to matching records.
- **Translation Tables.** Translation tables map duplicate records from different source systems via the unique keys. This helps optimize the matching process and provides a clean, consolidated view of entities across systems.
- **Standardization engine.** This converts data from different source systems into a standard format for each field in a reference set. (e.g., "St." becomes "Street" and "Rd." becomes "Road.") Often, the hub stores a small subset of entity data (i.e., name, address, key) in the standard format. This creates the core reference set that applications or users can query or download.
- **Rules Software.** The hub lets designers create rules about how to match and edit duplicate data. For example, survivorship rules designate which sources take precedence when populating or updating fields in the reference data.

[1] When the volume or volatility of reference data is not too large, firms commonly appoint a dominant operational system to be the "system of record" for each type of reference data.

- **Master Data Management software.** The hub usually provides a mechanism for authorized users to update and maintain reference data, view and modify rejected records, and so on.
- **Data extraction, movement, and transformation engine.** This engine moves data between source systems and the hub in batch or real time depending on the requirements of the source system. If needed, the engine can transform data into a format required by the hub or source system.

Customer Data. To consolidate and standardize its voluminous data about corporate customers, a leading high technology company created a customer data integration hub, which consists of 1.8 TB of customer data culled from 60 different source systems.

The hub, as described above, uses a matching engine to identify duplicate data in the source systems and assigns a unique key to each. It then standardizes a few fields across all the source systems (i.e., organization name and address) which it stores, along with unique fields from each source system. Authorized groups within the high technology firm can download this data from the hub to carry out marketing campaigns or other activities.

In addition, there are three systems at the firm that synchronize their data with the customer data integration hub. For example, a sales application sends new or updated customer records to the hub each night. The hub then matches and standardizes the records against its reference set and sends a cleansed version back to the sales application the next day, along with other attribute data in the hub that the sales application has subscribed to.

"Our [customer data integration hub] keeps customer data in sync for our key operating groups, such as sales, service, and CRM," says Kevin Mackey, IT manager at the firm. "We expect more groups to become active subscribers to [the hub] in the future, and eventually I expect our system to work in near real time using Web Services."

6. Conformed Marts

The Kimball Approach Can Be Used to Conform Existing Marts

One way to consolidate data marts without physically integrating them is to restructure the dimensions in each mart so they "conform" with each other. This approach adheres to Ralph Kimball's design methodology. But rather than starting from scratch, an organization redesigns the data models in existing data marts so that they conform. They also create a single, non-persistent staging area with which to populate the conformed marts with a single set of source data. The new staging area reduces costs and complexity by consolidating multiple, redundant data feeds.

Although conforming data marts is an industry standard design concept, it creates challenges when applied to existing analytic structures. Redesigning data mart models and changing extract feeds can cause downstream reports and applications to break. The redesign process can get unwieldy if there are a half-dozen or more data marts that need to be conformed.

The conformed mart approach works well for organizations that are committed to the Kimball approach and/or who want to preserve as much of the existing physical and logical structures as possible for political, financial, or other reasons.

Orange County. For example, Orange County Transportation Authority (OCTA) used this approach to consolidate four independent data marts that were running on Informix databases. Several factors converged to convince OCTA to pursue a Kimball approach to consolidating its data marts.

First, there was an organization-wide push to migrate from Informix to another database as a standard. This forced the data warehousing group to rethink its infrastructure. Second, a new "data-oriented" CEO began asking for more data and a single version of truth. This forced the group to rethink its architecture. Last, the data warehousing group was schooled in the Kimball method and decided to use the approach it was most familiar with to deliver a new integrated architecture on a new database infrastructure.

"It Cost Us 50 Percent More To Re-architect Our Marts During A Rehosting Project."

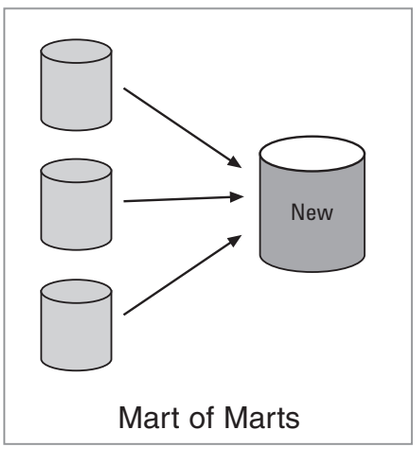
—Ray Riggins

"During our database conversion, we decided to spend the extra time and money to lay down a solid architectural foundation so that we could build data marts faster later on," says Ray Riggins, director of data management at OCTA. "We spent 50 percent more time on the database migration project because we re-architected the data marts with conformed dimensions."

OCTA's data marts run on a single server and within a single instance of the database. The environment supports a new non-persistent staging area as well as an operational data store for operational reporting.

7. Mart of Marts

If your organization is highly decentralized and only a corporate group requires consolidated information, an easy way out of this quandary is to create a data mart from all the existing analytic structures. This "mart of marts" approach pulls relevant enterprise information out of the existing structures using standard ETL practices and tools.



The benefit of this approach is that you don't change the existing environment at all, which is attractive politically. You may also be able to reuse existing staff and tools to build the new data mart, also saving substantial sums.

British Telecommunications. British Telecommunications (BT) created a financial data mart from 11 other data marts, each of which was associated with a different business unit in the firm. Although each unit was running Oracle Financials, each had structured its general ledger differently. The corporate group decided to create a "mart of marts" to reconcile and consolidate these different general ledgers for regulatory and external financial reporting.

But not long after BT implemented this approach, a new CFO came on board and decided to consolidate the general ledgers of all the business units into a single instance of Oracle Financials. This eliminated the multiple operational marts and the need for the "mart of marts." BT created a new financial data mart for the consolidated general ledger data. So, in BT's case, the "mart of mart" approach was an interim step in the evolution of its analytic architecture.

8. Distributed Query

Like the "mart of marts" approach, the distributed query approach leaves existing non-integrated analytic structures in place. Instead of using ETL tools and batch processing to create a consolidated view, the distributed query approach builds the view on the fly using SQL.

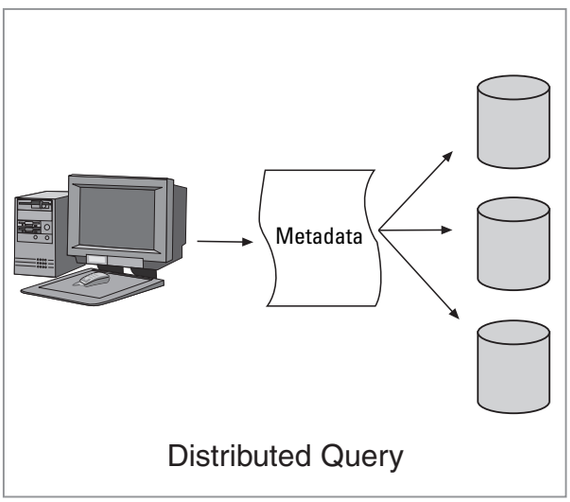
Distributed Queries Integrate Data from Distributed Systems on the Fly

Distributed query tools have been around for decades. In the 1990s, proponents advocated using these tools to create "virtual data warehouses." Today, they are called enterprise information integration (EII) tools. Whatever name, the tools provide a global metadata view of distributed data sources, including relational, legacy, and even Web, XML, or text-based data. The tools shield users from the complexity of accessing distributed, heterogeneous systems. In the background, they generate complex SQL to join disparate data and optimize query performance. They also perform lightweight transformations to convert data into a common format.

Distributed queries work well when data volumes are small, data is relatively clean, and disparate sources share common keys. The technology is useful for supplementing an existing data warehouse with up-to-the-minute real-time data or external data that can't be extracted in batch. Operational dashboards often use virtual queries to populate its dials and gauges.

Advocates admit the technology is no substitute for a general purpose data mart or warehouse. However, they say the tools help identify which data to load into a data warehouse prior to spending the time and money to create a physical model and all the extract mappings.

IBM Global Services. IBM Global Services uses a distributed query tool from Meta5, Inc., to populate and support queries against a metrics dashboard for executives at its headquarters. The tool pulls data from more than 30 other data marts and reporting systems within IBM Global Services. It populates the dashboard initially and then, when users request additional data, it queries the relevant data marts on the fly and joins the results on the screen. Users can leverage their existing business intelligence tools to access the dashboard.



Distributed Query

"[The Distributed Query Approach] Was the Quickest of All the Alternatives."

—Douglas Jones

"This was the quickest of all the alternatives," says Douglas Jones, IT Business Solutions program manager at IBM Global Services. "We're a small rapid application development team and we are paid to develop things quickly. We don't want to move a lot of data; we would rather get it where it's at. However, sometimes we replicate data for performance reasons."

Summary. The previous case studies illustrate that some consolidation strategies are best suited for different organizational structures or strategies. (See Table 1.)

Strategies Map to Organizational Structures

For instance, the "start from scratch" strategy is best used when two equally sized companies merge. The "designate and evolve" strategy is best used when a bigger company acquires a smaller one. Rehosting is used by companies looking for quick cost savings, while synchronization is good for large companies with lots of operational applications that need to share the same reference data.

STRATEGY & Strategy	Organizational Structure	Speed to Deploy	Metadata Integration Savings	Hardware
1. Rehost	Centralized – Fast Focus on quick cost savings	None	High	High
2. Start from Scratch	Centralized – Merger of equals	Slow	High	High – Eventually
3. Designate & Evolve	Centralized – Merger of unequals	Medium	Moderate	High – Eventually
4. Backfill	Decentralized – But central IT sells shared infrastructure	Medium	Moderate	Extra costs for DW
5. Synchronize	Centralized – Master data management	Slow	Moderate	Extra costs for hub
6. Conformed Data Marts	Decentralized – But with enterprise view	Medium	Moderate	None
7. Create a Mart of Marts	Decentralized – But top execs need single view	Fast	Minimal	Extra costs for data mart
8. Distributed Query	Decentralized – Quick fix	Fast	Minimal	None

Table 1. Summary of characteristics of the eight consolidation strategies.

The distributed strategies are often appropriate for decentralized organizations in which business units hold a lot of clout. Often, the distributed architectures are good interim solutions while the organization implements a centralized strategy. However, the conformed mart strategy can also be deployed in a centralized fashion if all the marts are logical schema within a single instance of a central database.

Integrating Silos Requires More Time Than Rehosting

We also discovered a correlation between the level of metadata integration required and the speed at which the strategy can be deployed. It's no surprise that the strategies that are quickest to deploy—rehosting, mart of marts, and distributed query—involve the lowest level of metadata integration. Conversely, those strategies requiring the highest levels of metadata integration—especially "start from scratch"—are the most time consuming to deploy, often taking several years to complete.

Some Approaches Increase Hardware Costs

In addition, strategies varied on the degree of hardware savings each affords. Rehosting offers quick savings, while "starting from scratch" and "designate and evolve" can pay off in the long run. Most of the other approaches actually increase hardware costs, because they require an additional server.

Ultimately, the approach an organization uses to migrate to an environment that delivers standardized information depends on many factors, including its culture, organizational structure, technical expertise, funding, and available time to accomplish the migration.

Managing a Consolidation Project: Lessons Learned

The best of migration plans can go awry unless they are well executed. Like any IT project, there are a number of steps that organizations need to follow when consolidating analytic silos.

1. Sell the Project. We've seen that strategic concerns (i.e., single version of the truth) trump tactical concerns (i.e., reduce costs) when it comes to getting executive support for consolidation projects. Although you may be distraught by the chaos and waste within your analytic architecture, selling business people on a clean technical architecture won't work, says Claudia Imhoff, president of Intelligent Solutions, Inc., a data warehousing consultancy in Boulder, CO. You need to couch the project in terms that get them excited, like delivering performance metrics, generating millions in new revenue, or getting a consolidated view of customers.

Sell the Business Value of the Project

And, since data consolidation projects can stir up political animosity, it's critical to sell the project at the top echelons of the firm. You will need to leverage executive support to convince powerful business groups to sacrifice their own preferences for the good of the organization, Imhoff says.

But don't stop with the executives. You need to sell the project to everyone, or at least set expectations properly. You need to let users understand why you are modifying their system, what impact it will have on them, and how it will make them more effective or productive. Call for their assistance and help.

If users aren't coming to meetings or aren't fully engaged while there, you may need to call upon top executives to verbalize their commitment to the project and the commitment they expect from everyone else.

2. Map Out Your Current Landscape. First, it's critical that you assess your current landscape, understand what analytic structures exist, what data they contain, where they source data from, and how they define rules and data elements.

Know Where You Are So You Know How to Get to Where You're Going

It also is important to understand the financial drivers in the environment, at least at a high level. Which data marts have hardware or software leases that are about to expire? Which systems are using non-standard hardware, databases, storage, or BI tools? Which systems have loyal users and support staff who don't want to see the system disappear? Which data marts are logical, which are physically separate?

Mapping out your existing environment gives you ammunition with which to pitch your project to top management. It also helps you prioritize which data marts or other structures are best to tackle first and which should be left until later or ignored.

3. Design Target Environments. Second, you need to design your target environment, selecting from one of the options listed in the previous section. Do you want a centralized or decentralized architecture? What is the model for the new system?

If you are going to do more than just rehost systems, you will need to create a new data model for your consolidated environment. Unfortunately, many consolidation projects operate under significant time constraints.

"Our management wanted us to hurry up and get off [our Informix] databases and tools because they didn't want to pay maintenance on them any more. As a result, we touched base with our users quickly and only added a few requested enhancements to the new environment," says Ray Riggins of Orange County Transportation Authority (OCTA).

Only 45 Percent of Firms Gather Fresh Requirements in a Consolidation Project

Our research shows that almost half (45 percent) of organizations gather fresh requirements in a consolidation project, while 42 percent gather partial requirements, like OCTA did. (See illustration 19.)

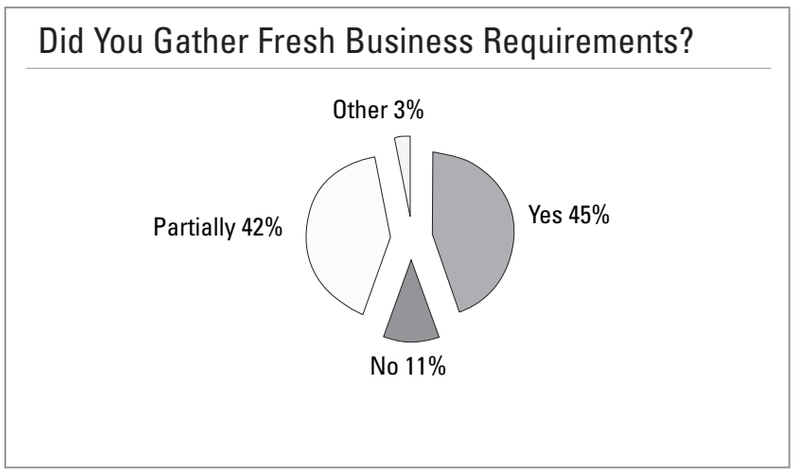


Illustration 19. Most consolidation initiatives involve gathering new requirements. Based on 615 respondents.

4. Standardize Definitions and Rules. One of the biggest challenges in consolidating environments is getting representatives of business groups to agree on terms and definitions governing data in the new consolidated environment. (See illustration 20.)

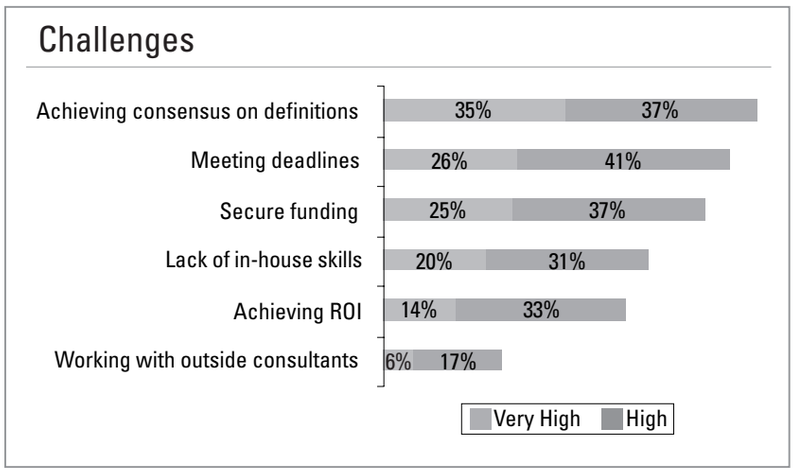


Illustration 20. Getting people to agree on terms and definitions is very challenging. Based on 573 respondents.

The politics of standardizing metadata can be horrendous. "Because we had so many sources of customer information, I finally had to stop the press, get some people in the room and ask, 'What is a customer?' It took me about a year to come up with a concise, comprehensive definition that everyone agreed with," says Wanda Black, director of information resource management at a privately held manufacturing firm.

The only solution to this problem is to leverage the clout of the top executive and driver to get all vested parties in the same room to hash out terms and definitions. However, mid-level managers often resist projects because they think that their group will lose their unique views of the data, says Laura Reeves, principal of StarSoft Solutions, a data warehousing consultancy in Naperville, IL.

An Enterprise View Doesn't Mean Managers Have to Give Up Their Local Views

Reeves suggests that you tell managers that a new enterprise view doesn't mean they have to change the way they view the world. Data marts that use conformed dimensions can roll unique views into an enterprise view using a mapping table and metadata. This way, each group gets a data mart that meets their needs without undermining an enterprise view of the data, says Reeves.

5. Decide Whether to Consolidate Tools. Organizations also need to decide whether they are going to standardize on a platform and tools as part of the consolidation project. Our research shows that about three-quarters of all companies standardize on core analytic technologies—ETL tools, BI tools, databases, hardware, and operating systems—when consolidating data warehouses and data marts. (See illustration 21.)

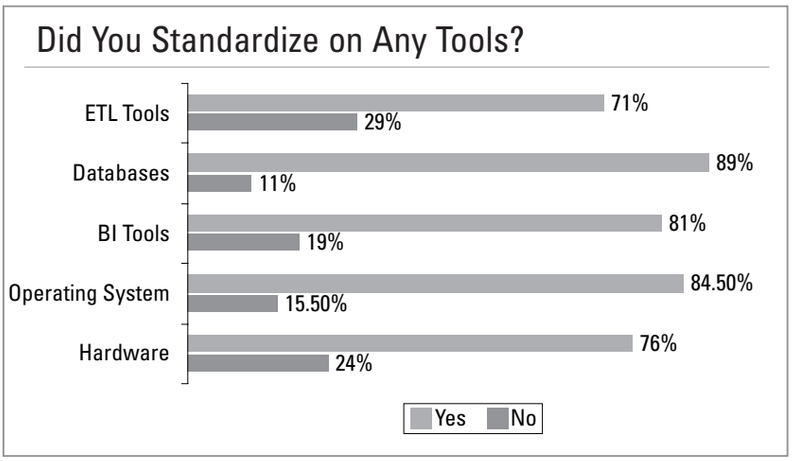


Illustration 21. Most companies standardize on tools and platforms when consolidating data warehouses and data marts. Based on 600 respondents.

When it comes to query and reporting tools, many companies think the decision is a no-brainer. "We know that 70 to 80 percent of the effort in consolidating our data warehouses is tied up in the data and sources, more specifically rewriting ETL mappings and data models. So, we might as well make an extra 20 percent effort to implement a standard toolset as well," says a data warehousing manager at the forest products firm.

However, politics often preclude companies from standardizing on BI tools. That's because users who have used a tool for a long period of time get proficient with it and don't want to give it up. Plus, the technical people who support the tool have developed expertise and clout within the organization, and they don't want that to disappear.

"It has actually been harder to shut down one [legacy BI tool] than consolidate our independent data marts," says Dirk de Wilde of Canadian National Railway Company. "First, [the tool owners] don't want us to touch the system, and second, the tool drives hundreds of complex reports, not all of which are being used, but which would take us an immense amount of time to convert."

"It has been Harder to Shut Down One [Legacy BI Tool] Than Consolidate Our Data Marts."

—Dirk de Wilde

However, de Wilde has not given up. He knows he has to provide "extra value" to get the group to use the firm's new data warehouse and standard BI tools. "We are talking with them about how the new infrastructure provides better quality data and many reports and applications that they would find valuable."

6. Create the Team and Execute the Project. Organizations need to put a team in place that can execute the consolidation while keeping the existing environments up and running. Often, this requires companies to hire outside help.

Our research shows that 60 percent of companies use consultants to help with project implementation. Slightly more than half (52 percent) also use consultants for up-front planning and strategy and just under half (44 percent) use them for project planning. (See illustration 22.)

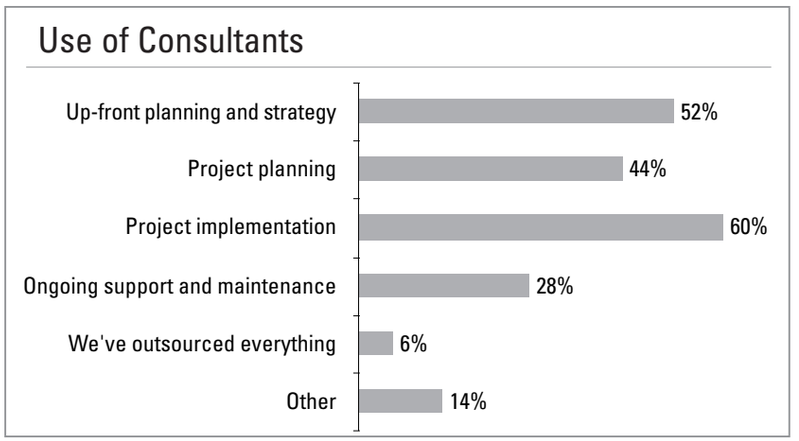


Illustration 22. Most companies use consultants in consolidation projects. Based on 531 respondents.

Tools Can Substitute for Extra Staff

Besides using consultants, about 40 percent of companies leverage tools to assist in consolidation projects. Tools can make existing developers more proficient, saving labor costs in the long run. The most commonly used tools are ETL tools, which can be used to transform and move data between analytic structures with different data models. ETL tools were selected by 56 percent of respondents, followed by data modeling tools (23 percent). (See illustration 23.)

Other companies use profiling tools to analyze the nature of source data prior to migrating it. The information derived from profiling tools can be used to build the target data model as well as one-time and ongoing data cleansing routines that get inserted into the ETL programs.

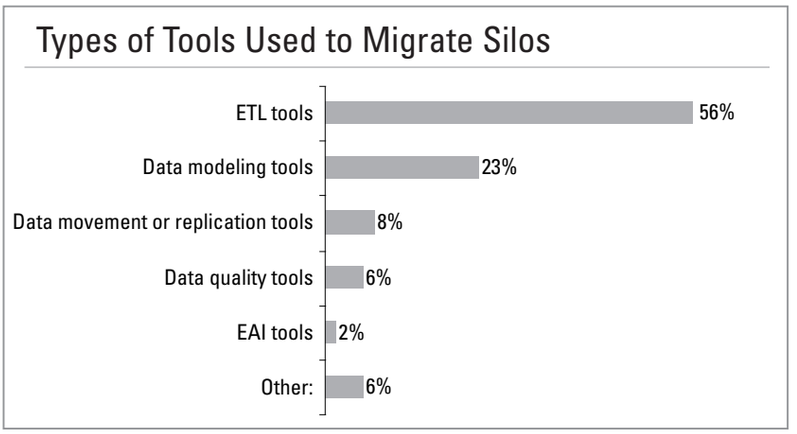


Illustration 23. A majority of firms use ETL tools to migrate analytic structures followed by data modeling tools. Based on 226 respondents.

Conclusion

Today, organizations are plagued by the proliferation of analytic silos, which make it difficult to deliver consistent information across the enterprise. Organizations are now starting to consolidate these silos into an enterprise data warehouse (EDW) to deliver a single version of truth and reduce overhead costs.

A Long Way to Go. But organizations still have a long way to go. They've consolidated one-third of their analytic structures on average and have dozens more to go. Since it takes between four and 10 months to consolidate analytic structures, these consolidation projects can last years.

The good news is that the ROI for these projects is impressive. Organizations that consolidate multiple analytic structures make $3.34 million on their investments in about two years. However, most organizations don't bother calculating the ROI because executives view the project as strategic, not tactical. The purpose of consolidating analytic silos is not to save money—although that helps—it's to deliver a single version of truth.

The Good News Is That the ROI for These Projects Is Impressive

Architectural Choices and Migration Strategies. Once a commitment to consolidate is made, organizations usually select one of eight approaches to migrate from analytic silos to an integrated environment that delivers consistent data. Four of the approaches use a centralized architecture, where all data and systems are brought together in a single data center, usually on a single machine and database management system. These four centralized approaches are generally preferred over distributed approaches, which leave existing analytic structures in place.

Ultimately, the approach an organization uses to migrate to an environment that delivers standardized information depends on many factors, including its culture, organizational structure, technical expertise, funding, and available time to accomplish the migration.

Critical Success Factors. Besides selecting the right consolidation architecture and strategy, the key to successful consolidation projects is to gain top management commitment and then put together a project plan, team, and tools that help you migrate incrementally to your target environment. It's critical to assess your current environment so you can prioritize efforts and gain momentum for the project. It's also important to define your target analytic architecture but not get caught up in rushing there too fast.

Get Top Management Support

"I try to keep it simple," says Russ Vaughn of Bank of America. "You need to know what steps one, two, and three are but you can't focus on step two until you finish step one. I stay on a 60-day commitment list."

Ascential Software Corporation
50 Washington Street
Westboro, MA 01581
800.966.9875 or 508.366.3888
Fax: 508.366.3669
info@ascentialsoftware.com
www.ascential.com

Ascential Software is the leader in enterprise data integration. Customers and partners worldwide use the Ascential Enterprise Integration Suite™ to confidently transform data into accurate, reliable, and complete business information to improve operational performance and decision-making across every critical business dimension. Our comprehensive end-to-end solutions provide on demand data integration complemented by our professional services, industry expertise, and methodologies.

Ascential Software is headquartered in Westboro, MA, and has more than 3,000 customers and partners globally across such industries as financial services and banking, insurance, healthcare, retail, manufacturing, consumer packaged goods, telecommunications and government. Visit the Ascential Software Web site at www.ascential.com.

Hummingbird Ltd.
1 Sparks Avenue
Toronto, Ontario M2H 2W1
Canada
Toll-free:
1.877.FLY.HUMM (359.4866)
1.416.496.2200
Fax: 1.416.496.2207
www.hummingbird.com

Business Intelligence and Data Integration solutions from Hummingbird provide flexible and user-friendly tools that enable companies to improve their decision-making effectiveness at the lowest cost of ownership.

Hummingbird Ltd. is a leading global provider of enterprise software solutions. Hummingbird Enterprise, our enterprise content management (ECM) solution, allows customers to manage the entire lifecycle of enterprise content from creation to disposition. Hummingbird Enterprise™, our enterprise content management (ECM) solution, provides an integrated platform for ECM and business analytics that enable organizations to address critical business needs, such as information management, business continuity, compliance, and risk mitigation.

Informatica Corporation
2100 Seaport Boulevard
Redwood City, CA 94063
650.385.5000
Fax: 650.385.5500
www.informatica.com

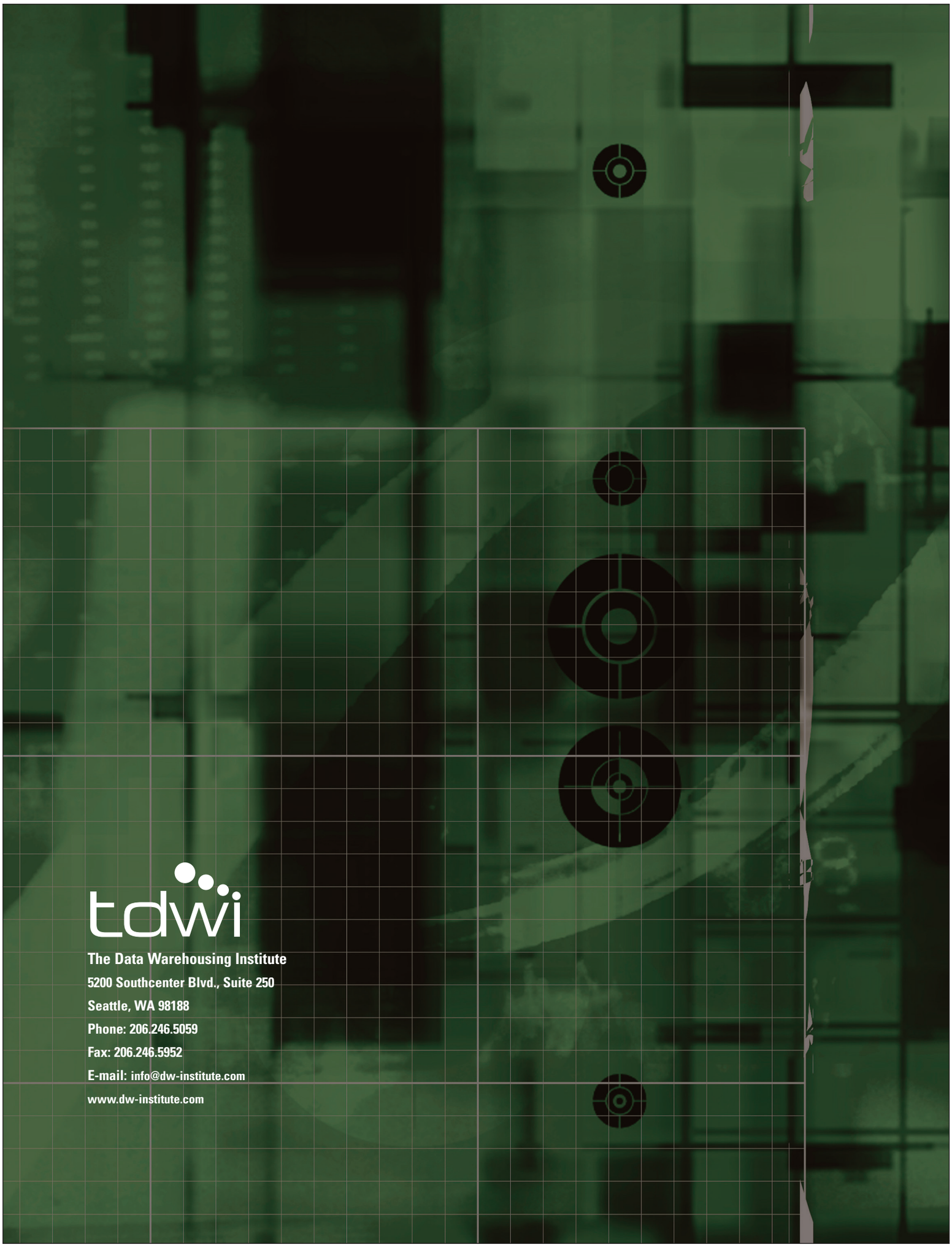
Informatica Corporation is a leading provider of data integration software. Using Informatica products, companies can access, integrate, visualize, and audit their enterprise information assets to help improve business performance, increase customer profitability, streamline supply chain operations, and proactively manage regulatory compliance. More than 2,000 companies worldwide rely on Informatica for their end-to-end enterprise data integration needs. For more information, call 650.385.5000 (800.970.1179 in the U.S.), or visit www.informatica.com/.

Teradata, a division of NCR
1700 S. Patterson Boulevard
Dayton, OH 45479
937.445.5993
www.teradata.com

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

Unisys Corporation
One Unisys Way
Blue Bell, PA 19422
215.986.4011
www.Unisys.com/bi

Unisys is a worldwide information technology services and solutions company helping clients capitalize on technology to achieve strategic advantage by turning their data into actionable information. Using our advanced Enterprise-class ES7000 family of servers based on Intel and Microsoft technology, we are now delivering a new economic model for business intelligence enabling Unisys to provide a complete business intelligence solution, including BI platform, infrastructure, and infrastructure services partnered with strategic software and integration partners. The ES7000 Business Intelligence platform enables organizations to utilize the power of the Microsoft Business Intelligence solution, including its advanced OLAP and data mining capabilities on Terabytes worth of information. For more information on Unisys's infrastructure and service offerings, access the Unisys BI home page www.unisys.com/bi.



The Data Warehousing Institute
5200 Southcenter Blvd., Suite 250
Seattle, WA 98188
Phone: 206.246.5059
Fax: 206.246.5952
E-mail: info@dw-institute.com
www.dw-institute.com